arXiv: 1706.04059

# Approximate Optimal Designs for Multivariate Polynomial Regression

**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*
*Université Paul Sabatier, CNRS UMR 5219, 118 route de Narbonne, 31062 Toulouse, France*

[•]*Laboratoire d'Analyse et d'Architecture des Systèmes CNRS*
*Université de Toulouse, LAAS, 7 avenue du colonel Roche, F-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 and approximately the 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 exactly in polynomial time thanks to our method. Finite convergence of the hierarchy can be standardly certified numerically as well. 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.

**MSC 2010 subject classifications:** Primary 62K05, 90C25; secondary 41A10, 49M29, 90C90, 15A15.

**Keywords and phrases:** Experimental Design, Semidefinite Programming, Christoffel Polynomial, Linear Model, Equivalence theorem.

## 1. Introduction

### *1.1. Convex design theory*

The optimum experimental designs are computational and theoretical objects that aim to minimize the uncertainty contained in 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* $\mathfrak{f}_1,\dots,\mathfrak{f}_p$, namely

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

where $\theta_1,\dots,\theta_p$ are unknown parameters that the experimenter wants to estimate, $\varepsilon_i$ is some noise and the inputs $\xi_i$ are chosen by the experimenter in a *design space* $\mathscr{X} \subseteq \mathbb{R}^n$. Assume that the inputs $\xi_i$, for $i = 1,\dots,N$, are chosen within a set of distinct points $x_1,\dots,x_\ell$ with $\ell \le N$, 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

$$\zeta := \begin{pmatrix} x_1 & \cdots & x_\ell \\ \frac{n_1}{N} & \cdots & \frac{n_\ell}{N} \end{pmatrix}, \tag{1}$$

whose first row gives distinct points in the design space $\mathscr{X}$ 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 $\zeta$ the standard analysis of the Gaussian linear model


imsart-generic ver. 2014/10/16 file: ArXivV1.1.tex date: June 23, 2017

shows that the minimal covariance matrix (with respect to Loewner ordering) of unbiased estimators can be expressed in terms of the Moore-Penrose pseudo-inverse of the *information matrix* which is defined by

$$\mathfrak{I}(\zeta) := \sum_{i=1}^{\ell} w_i \mathfrak{F}(x_i)\,\mathfrak{F}^{\top}(x_i), \tag{2}$$

where $\mathfrak{F} := (\mathfrak{f}_1, \ldots, \mathfrak{f}_p)$ is the column vector of regression functions and $w_i := n_i/N$ is the weight corresponding to the point $x_i$. One major aspect of design of experiment theory seeks to maximize the information matrix over the set of all possible $\zeta$. Notice that the Loewner ordering is partial and, in general, there is no greatest element among all possible information matrices $\mathfrak{I}(\zeta)$. The standard approach is then to consider some statistical criteria, namely *Kiefer's $\phi_q$-criteria* [13], 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 $p$. 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 $\mathrm{trace}(M)$, $\det(M)$ and $\lambda_{\min}(M)$ denote respectively the trace, determinant and least eigenvalue of the symmetric nonnegative definite matrix $M$. 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.

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

$$\max \phi_q(\mathfrak{I}(\zeta)) \tag{3}$$

where the maximum is taken over all $\zeta$ of the form (1). Standard criteria are given by the parameters $q = 0, -1, 1, -\infty$ and are referred to $D$-, $A$-, $T$- or $E$-optimum designs respectively.

### *1.2. State of the art*

Optimal design is at the heart of statistical planning for inference in the linear model, see for example [3]. While the case of discrete input factors is generally tackled by algebraic and combinatoric arguments (*e.g.*, [2]), the one of continuous input factors often leads to an optimization problem. In general, the continuous factors are generated by a vector $\mathfrak{F}$ of linearly independent regular functions on the *design space* $\mathscr{X}$.

One way to handle the problem is to focus only on $\mathscr{X}$ ignoring the function $\mathfrak{F}$ and to try to draw the design points *filling* the set $\mathscr{X}$ in the best way. This is generally done by optimizing a cost function on $\mathscr{X}^N$ 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 [29, 37]) and the minimum discrepancy designs, see for example [23].

Another point of view—which is the one developed here—relies on the *maximization* of the information matrix. Of course, as explained before, the set of information matrices is a partially ordered set with respect to the Loewner ordering, and so the optimization cannot be performed directly on this matrix but on a real function on it. A pioneer paper adopting this point of view is the one of Elfving [7]. In the early 60's, in a series of papers, Kiefer and Wolwofitz throw new light on this kind of methods for experimental design by introducing the equivalence principle and proposing algorithms to solve the optimization problem for some cases; see [13] and references therein. Following the early works of Karlin and Studden [11, 12], the case of polynomial regression on a compact interval on $\mathbb{R}$ has been widely studied. In this frame, the theory is almost complete and many things can be said about the optimal solutions for the design problem, see for instance [5]. Roughly speaking, the optimal design points are related to the zeros of orthogonal polynomials built on an equilibrium measure. We refer to the inspiring book of Dette and Studden [6] and references therein for a complete overview on the subject.

In the one dimensional frame, other systems of functions $\mathfrak{F}$—trigonometric functions or $T$-systems, see [14] for a definition—are studied in the same way in [6], [20] and [10] (see also the recent paper [15] for another perspective on the subject). In the multidimensional case, even for polynomial systems, very few cases of explicit solutions are known. Using tensoring arguments the case of a rectangle is treated in [6, 35]. Particular models of degree two are studied in [4, 30].

Apart from these particular cases, the construction of the optimal design relies on numerical optimization procedures. The case of the determinant—which corresponds to the choice $q = 0$, i.e., the $D$-optimality—is studied for example in [38] and [36]. Another criterion based on matrix conditioning—referred to as $G$-optimality—is developed in [25]. In the latter paper, the construction of an optimal design is performed in two steps. In the first step one only deals with an optimization problem on the set of all possible information matrices, while in the second step, one wishes to identify a possible probability distribution associated with the optimal information matrix. General optimization algorithms are discussed in [8] and [1]. A general optimization frame on measure sets including gradient descent methods is considered in [26]. In the frame of fixed given support points, efficient SDP based algorithms are proposed and studied in [32] and [33]. Let us mention the paper [36] which is one of the original motivations 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 approximate optimal designs—in the sense of Kiefer's $\phi_q$-criteria—on a large variety of design spaces that we refer to as semi-algebraic sets, see [16] or Section 2 for a definition. These can be understood as sets given by intersections and complements of superlevel sets of multivariate polynomials.

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. 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-like polynomials involved in the optimality conditions. The term "Christoffel-like polynomials" will become clear in the sequel

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. This contrasts with the classical use of ellipsoids for linear regressions where points are obtained on the boundary.

### 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 short conclusion.

## 2. Polynomial optimal design and moments

This section collects preliminary material on semi-algebraic sets, moments and moment matrices, using the notation of [16]. 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 $p$.

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

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

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 $\mathscr{X}$ 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 [6, Chapter 5]. In particular, observe that the design space $\mathscr{X}$ defined by (4) is not necessarily convex and note that the polynomial regressors $\mathfrak{F}$ can handle incomplete $m$-way $d$th 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

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

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

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

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

### 2.2. Moments, the moment cone and the moment matrix

Given $\mu \in \mathscr{M}_+(\mathscr{X})$ and $\alpha \in \mathbb{N}^n$, we call

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

the moment of order $\alpha$ of $\mu$. Accordingly, we call the sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ the moment sequence of $\mu$. 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 $\mu$ such that $\mathbf{y}$ is its moment sequence.

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

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

Let $\mathscr{P}_d(\mathscr{X})$ denotes the convex cone of all polynomials of degree at most $d$ that are nonnegative on $\mathscr{X}$.

When $\mathscr{X}$ is a compact set, then $\mathscr{M}_d(\mathscr{X}) = \mathscr{P}_d(\mathscr{X})^\star$ and $\mathscr{P}_d(\mathscr{X}) = \mathscr{M}_d(\mathscr{X})^\star$, see *e.g.,* [17, Lemma 2.5] or [14].

When the design space is given by the univariate interval $\mathscr{X} = [a, b]$, *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.,* [36]. Unfortunately, in the general case, there is no efficient representation of this cone. It has actually been shown in [34] 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 $\mu$ supported on $\mathscr{X}$ if and only if $L_\mathbf{y}(f) \geqslant 0$ for all polynomials $f \in \mathbb{R}[x]$ nonnegative on $\mathscr{X}$ [16, 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 ($M_d(\mathbf{y})(\alpha, \beta) = M_d(\mathbf{y})(\beta, \alpha)$), 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 $f = \sum_{|\alpha|\le r} f_\alpha x^\alpha \in \mathbb{R}[x]_r$ 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 $\alpha, \beta \in \mathbb{N}^n, |\alpha|, |\beta| \le d$ 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 $\mu$, then $M_d(f\mathbf{y}) \succcurlyeq 0$ for $f \in \mathbb{R}[x]_d$ whenever the support of $\mu$ is contained in the set $\{x \in \mathbb{R}^n : f(x) \geqslant 0\}$.

Since $\mathscr{X}$ is basic semi-algebraic with a certificate of compactness, by Putinar's theorem—see for instance the book [16, 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 $\mu \in \mathscr{M}_+(\mathscr{X})$ 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 $\mathscr{M}_{2d}(\mathscr{X})$ by the following semidefinite representable cones for $\delta \in \mathbb{N}$:

$$\mathscr{M}^{\mathrm{SDP}}_{2(d+\delta)}(\mathscr{X}) := \Big\{ \mathbf{y}_{d,\delta} \in \mathbb{R}^{\binom{n+2d}{n}} \ :\ \exists \mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}} \text{ such that } \mathbf{y}_{d,\delta} = (y_{\delta,\alpha})_{|\alpha|\le 2d} \text{ and } M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0,\ M_{d+\delta-v_j}(g_j\mathbf{y}_\delta) \succcurlyeq 0,\ j = 1, \ldots, m \Big\}. \tag{7}$$

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

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

This hierarchy actually converges, meaning $\mathscr{M}_{2d}(\mathscr{X}) = \overline{\bigcap_{\delta=0}^{\infty} \mathscr{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathscr{X})}$, 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.*, $\Sigma[x]_{2d} = \{\sigma \in \mathbb{R}[x] : \sigma(x) = \sum_{i=1}^{k} h_i(x)^2 \text{ for some } h_i \in \mathbb{R}[x]_d\}$. The topological dual of $\mathscr{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathscr{X})$ is a quadratic module, which we denote by $\mathscr{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathscr{X})$. It is given by

$$\mathscr{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathscr{X}) = \Big\{h = \sigma_0 + \sum_{j=1}^{m} g_j \sigma_j : \deg(h) \leq 2d,\ \sigma_0 \in \Sigma[x]_{2(d+\delta)},\ \sigma_j \in \Sigma[x]_{2(d+\delta-\nu_j)},\ j = 1, \ldots, m\Big\}. \tag{9}$$

Equivalently, see for instance [16, Proposition 2.1], $h \in \mathscr{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathscr{X})$ 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 $\mathscr{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathscr{X})$ are polynomials of degree at most $2d$ which are non-negative on $\mathscr{X}$. Hence, it is a subset of $\mathscr{P}_{2d}(\mathscr{X})$.

## 3. Approximate Optimal Design

### 3.1. *Problem reformulation in the multivariate polynomial case*

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

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

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

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

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

Consider the optimization problem

$$\begin{aligned} &\max\ \phi_q(M) \\ &\text{s.t. } M = \sum_{|\gamma|\leq 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^{\ell} \frac{n_i}{N} x_i^\gamma, \quad \sum_{i=1}^{\ell} n_i = N, \\ &\qquad x_i \in \mathscr{X},\ n_i \in \mathbb{N},\ i = 1, \ldots, \ell, \end{aligned} \tag{11}$$

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

$$\zeta := \begin{pmatrix} x_1 & \cdots & x_\ell \\ w_1 & \cdots & w_\ell \end{pmatrix}, \tag{12}$$

where the frequencies $w_i$ belong to the unit simplex $\mathscr{W} := \{w \in \mathbb{R}^\ell : 0 \leq w_i \leq 1,\ \sum_{i=1}^\ell w_i = 1\}$. 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)

$$\begin{aligned} &\max\ \phi_q(M) \\ &\text{s.t. } M = \sum_{|\gamma|\leq 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^\ell w_i x_i^\gamma, \\ &\quad x_i \in \mathscr{X},\ w \in \mathscr{W}, \end{aligned} \tag{13}$$

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 $\mathscr{M}_{2d}(\mathscr{X})$ defined in (6) and such that $y_0 = 1$, is exactly the set:

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

where $\ell \leq \binom{n+2d}{n}$, see the so-called Tchakaloff theorem [16, Theorem B12].

Hence, Problem (13) is equivalent to

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

where the maximization is now with respect to the sequence $\mathbf{y}$. Moment problem (14) is finite-dimensional and convex, yet the constraint $\mathbf{y} \in \mathscr{M}_{2d}(\mathscr{X})$ is difficult to handle. We will show that by approximating the truncated moment cone $\mathscr{M}_{2d}(\mathscr{X})$ 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 (12).

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 $\mathfrak{F} = (\mathfrak{f}_1, \ldots, \mathfrak{f}_p) := (x^\alpha)_{|\alpha| \leq d}$ with $p = \binom{n+d}{n}$. This case corresponds to $\mathfrak{A} = \mathrm{Id}$ in (5). Note that this is not a restriction, since one can get the results for other choices of $\mathfrak{F}$ 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 [6, 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 $\phi_q$ criteria in Table 1.

| Name<br>q | $D$-opt.<br>0 | $A$-opt<br>$-1$ | $T$-opt.<br>1 | $E$-opt.<br>$-\infty$ | generic case<br>$q \neq 0, -\infty$ |
|---|---|---|---|---|---|
| $\phi_q(M)$ | $\det(M)^{1/p}$ | $p(\mathrm{trace}(M^{-1}))^{-1}$ | $\mathrm{trace}(M)/p$ | $\lambda_{\min}(M)$ | $\left[\frac{\mathrm{trace}(M^q)}{p}\right]^{\frac{1}{q}}$ |
| $\nabla\phi_q(M)$ | $\det(M)^{1/p} M^{-1/p}$ | $p(\mathrm{trace}(M^{-1})M)^{-2}$ | $\mathrm{Id}/p$ | $\Pi_{\min}(M)$ | $\left[\frac{\mathrm{trace}(M^q)}{p}\right]^{\frac{1}{q}-1} \frac{M^{q-1}}{p}$ |

TABLE 1

*Gradients of the Kiefer's $\phi_q$ 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 [22]. Notice further that $\phi_q$ 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

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

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*

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

*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*

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

*In particular, the following statements are equivalent:*

- $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ *is the unique solution to Problem* (15);
- $\mathbf{y}^\star \in \left\{ \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1 \right\}$ *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 [13, Theorem 1]. The case of $\phi_q$-criteria is tackled in [31] and [6, 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 Appendix A. □

**Remark 1** (On the optimal dual polynomial)**.** *The polynomial $p_d^\star$ contains all the information concerning the optimal design. Indeed, its level set $\Omega$ 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-like polynomial. Notice further that*

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

*Hence, the optimal design problem related to $\phi_q$ 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 [17] for a discussion and generalizations of Löwner-John's ellipsoid theorem for general homogenous polynomials on non convex domains.*

**Remark 2** (Equivalence theorem for $T$-optimality)**.** *Theorem 1 holds also for $q = 1$. This is the $T$-optimal design case for which the objective function is linear. Hence, in this case, $\mathbf{y}^\star$ is not unique. Further, note that the polynomial $p_d^\star$ can be explicitly written, as it does not depend on $\mathbf{y}^\star$. It yields*

$$x \mapsto p_d^\star(x) = ||\mathbf{v}_d(x)||_2^2.$$

*Thus, the support of any solution is included in the level set*

$$\Omega = \arg\max_{\mathcal{X}} p_d^\star\,.$$

*It follows that the set of solutions is exactly the set of probability measures supported by $\Omega$.*

**Remark 3** (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*

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

*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*

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

*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 $\mu$ whose moments up to order $2d$ coincide with $\mathbf{y}^\star$.

**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*

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

*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

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

when $\mathbf{y}$ has a representing measure $\mu$—when $\mathbf{y}$ does not have a representing measure $\mu$ 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 [19] 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 [21].

**Remark 4** (Equivalence theorem for $D$-optimality)**.** *In the case of D-optimal designs, observe that*

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

*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 $\mathscr{X} \subseteq \mathbb{R}^n$ 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 $\mathscr{X}$.

Since the ideal moment Problem (15) involves the moment cone $\mathscr{M}_{2d}(\mathscr{X})$ 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

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

Since Problem (17) is a relaxation of the ideal Problem (15), necessarily $\rho_\delta \geq \rho$ for all $\delta$. 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 $\mathscr{X} \subseteq \mathbb{R}^n$ be a compact semi-algebraic set as defined in (4) and 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 $p^\star := \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star$ is non-negative on $\mathscr{X}$ and $L_{\mathbf{y}^\star}(p^\star) = 0$.*

*In particular, the following statements are equivalent:*

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

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

*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 $\mathscr{X}$, scaled to be a probability measure.

Next, let $\mathbf{y}\in\binom{n+d}{n}$ 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 $\mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X})$ given in (7). As $g_1(x)=1-\|x\|^2$ and $M_{d+\delta-1}(g_1\,\mathbf{y}_\delta)\succcurlyeq 0$, we have

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

and so by [18],

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

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 $\phi_q$ is upper semi-continuous, the supremum in (17) is attained at some optimal solution $\mathbf{y}^\star\in\mathbb{R}^{2d}$. It is unique due to convexity of the feasible set and strict concavity of the objective function $\phi_q$, e.g., see [31, 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

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

First, we notice that there exists a strictly feasible solution to (17) because the cone $\mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X})$ has nonempty interior as a supercone of $\mathscr{M}_{2d}(\mathscr{X})$, which has nonempty interior by [17, Lemma 2.6]. Hence, Slater's condition[2] holds for (17). Further, by an argument in [31, Chapter 7.13]) the matrix $M_d(\mathbf{y}^\star)$ is non-singular. Therefore, $\phi_q$ is differentiable at $\mathbf{y}^\star$. Since additionally Slater's condition is fulfilled and $\phi_q$ 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

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

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

$$1_{\alpha=0}\,\lambda^\star-\langle\nabla\phi_q(M_d(\mathbf{y}^\star)),\mathbf{B}_\alpha\rangle\,=\,\langle\Lambda_0,\tilde{\mathbf{B}}_\alpha\rangle+\sum_{j=1}^m\langle\Lambda_j,\mathbf{C}_\alpha^j\rangle,\quad|\alpha|\leq 2(d+\delta), \tag{19}$$

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

$$\langle M_{d+\delta}(\mathbf{y}^\star_\delta),\Lambda_0\rangle=0;\quad\langle M_{d+\delta-v_j}(\mathbf{y}^\star_\delta\,g_j),\Lambda_j\rangle=0,\quad j=1,\ldots,m. \tag{20}$$

[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

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

We deduce that

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

by the Euler formula for homogeneous functions.

Similarly, multiplying by $x^\alpha$ and summing up yields

$$\begin{aligned}
&\lambda^\star - \mathbf{v}_d(x)^\top \nabla\phi_q(M_d(\mathbf{y}^\star))\mathbf{v}_d(x)\\
&= \Big\langle \Lambda_0, \sum_{|\alpha|\le 2(d+\delta)} \tilde{\mathbf{B}}_\alpha\, x^\alpha \Big\rangle + \sum_{j=1}^m \Big\langle \Lambda_j, \sum_{|\alpha|\le 2(d+\delta-v_j)} \mathbf{C}^j_\alpha\, x^\alpha \Big\rangle\\
&= \underbrace{\Big\langle \Lambda_0, \mathbf{v}(x)_{d+\delta}\,\mathbf{v}_{d+\delta}(x)^\top \Big\rangle}_{\sigma_0(x)} + \sum_{j=1}^m g_j(x) \underbrace{\Big\langle \Lambda_j, \mathbf{v}_{d+\delta-v_j}(x)\,\mathbf{v}_{d+\delta-v_j}(x)^\top \Big\rangle}_{\sigma_j(x)}\\
&= \sigma_0(x) + \sum_{j=1}^n \sigma_j(x)\, g_j(x)\\
&= \hat{p}^\star(x) \in \mathscr{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathscr{X}).
\end{aligned} \tag{23}$$

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

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

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

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

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

The equivalence follows from the argumentation in b). □

**Remark 5** (Finite convergence). *If the optimal solution $\mathbf{y}^\star$ of Problem (17) is coming from a measure $\mu^\star$ on $\mathscr{X}$, that is $\mathbf{y}^\star \in \mathscr{M}_{2d}(\mathscr{X})$, then $\rho_\delta = \rho$ and $\mathbf{y}^\star$ is the unique optimal solution of Problem (15). In addition, by 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 $\Omega := \{x \in \mathscr{X} :\ \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p^\star_d(x) = 0\}$.*

**Remark 6** (SDP relaxation for $T$-optimality). *In this case, recall that $\mathbf{y}^\star$ is not unique and recall that the polynomial $p^\star_d$ can be explicitly written as $x \mapsto p^\star_d(x) = ||\mathbf{v}_d(x)||_2^2$. The above proof can be extended to the case $q = 1$ to derive that any solution $\mathbf{y}^\star$ satisfies $\mathrm{trace}(M_d(\mathbf{y}^\star)) - ||\mathbf{v}_d(x)||_2^2 \geqslant 0$ for all $x \in \mathscr{X}$ and $L_{\mathbf{y}^\star}(p^\star_d) = \mathrm{trace}(M_d(\mathbf{y}^\star))$. In particular, the following statements are equivalent*

- $\mathbf{y}^\star \in \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X})$ *is a solution to Problem (17) (for $q = 1$);*
- $\mathbf{y}^\star \in \{\mathbf{y} \in \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X}) : y_0 = 1\}$ *and* $p^\star(x) := \mathrm{trace}(M_d(\mathbf{y}^\star)) - ||\mathbf{v}_d(x)||_2^2 \in \mathscr{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathscr{X})$.

**Remark 7** (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 $u = (u_\alpha)_{|\alpha|\le 2d}$ 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*

- $\mathbf{y}^\star \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$ *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* $p^\star(x) = \lambda_{\min}(M_d(\mathbf{y}^\star))\|u\|_2^2 - \left(\sum_\alpha u_\alpha x^\alpha\right)^2 \in \mathcal{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathcal{X})$.

*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 $\delta$ tends to infinity.

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

*a) $\rho_\delta \to \rho$ as $\delta \to \infty$, where $\rho$ is the supremum in (15).*
*b) For every $\alpha \in \mathbb{N}^n$ with $|\alpha| \le 2d$, we have $\lim_{\delta\to\infty} y^\star_{d,\delta,\alpha} = y^\star_\alpha$, where $\mathbf{y}^\star = (y^\star_\alpha)_{|\alpha|\le 2d} \in \mathcal{M}_{2d}(\mathcal{X})$ is the unique optimal solution to (15).*
*c) $p^\star_{d,\delta} \to p^\star_d$ as $\delta \to \infty$, where $p^\star_d$ is the Christoffel-like polynomial associated with $\mathbf{y}^\star$ defined in (16).*
*d) If the dual polynomial $p^\star := \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p^\star_d$ to Problem (15) belongs to $\mathcal{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathcal{X})$ for some $\delta$, then finite convergence takes place, that is, $\mathbf{y}^\star_{d,\delta}$ is the unique optimal solution to Problem (15) and $\mathbf{y}^\star_{d,\delta}$ has a representing measure, namely the target measure $\mu^\star$.*

*Proof.* We prove the four claims consecutively.

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

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

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

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

Notice that the subvectors $\mathbf{y}^\star_{d,\delta} = (y^\star_{\delta,\alpha})_{|\alpha|\le 2d}$ with $\delta = 0, 1, 2, \ldots$ belong to a compact set. Therefore, since $\phi_q(M_d(\mathbf{y}^\star_{d,\delta})) < \infty$ for every $\delta$, we also have $\phi_q(M_d(\hat{\mathbf{y}})) < \infty$.
Next, by Putinar's theorem [16, Theorem 3.8], $\hat{\mathbf{y}}$ is the sequence of moments of some measure $\hat{\mu} \in \mathcal{M}_+(\mathcal{X})$, and so $\hat{\mathbf{y}}_d = (\hat{y}_\alpha)_{|\alpha|\le 2d}$ is a feasible solution to (15), meaning $\rho \ge \phi_q(M_d(\hat{\mathbf{y}}_d))$. On the other hand, as (17) is a relaxation of (15), we have $\rho \le \rho_{\delta_k}$ for all $\delta_k$. So the convergence (24) yields

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

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 $(\mathbf{y}^\star_{d,\delta})_{\delta\in\mathbb{N}}$ converges to $\mathbf{y}^\star$, that is, for $\alpha \in \mathbb{N}^n$ with $|\alpha| \le 2d$ fixed

$$\lim_{d,\delta\to\infty} y^\star_{\delta,\alpha} = \lim_{\delta\to\infty} y^\star_{\delta,\alpha} = \hat{y}_\alpha = y^\star_\alpha. \tag{25}$$

c) It suffices to observe that the coefficients of Christoffel-like polynomial $p^\star_{d,\delta}$ are continuous functions of the moments $(y^\star_{d,\delta,\alpha})_{|\alpha|\leq 2d} = (y^\star_{\delta,\alpha})_{|\alpha|\leq 2d}$. Therefore, by the convergence (25) one has $p^\star_{d,\delta} \to p^\star_d$ where $p^\star_d \in \mathbb{R}[x]_{2d}$ 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. □

In order to illustrate what happens for general $\phi_q$ criteria, we give a corollary describing the relation between the solutions of Program (15) and Program (17) for $T$-optimal designs ($q=1$).

**Corollary 5.** *Let $\delta \geq 1$. Denote by $\mathbf{y}^\star$ (resp. $\mathbf{y}^\star_{d,\delta}$) a solution to Program (15) (resp. Program (17)). Only one of the following two cases can occur.*

- *Either*

$$\left\{\mathrm{trace}(M_d(\mathbf{y}^\star)) - ||\mathbf{v}_d(\cdot)||_2^2 \in \mathscr{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathscr{X})\right\} \Leftrightarrow \left\{\mathrm{trace}(M_d(\mathbf{y}^\star_{d,\delta})) - ||\mathbf{v}_d(\cdot)||_2^2 \text{ has a root in } \mathscr{X}\right\}$$
$$\Leftrightarrow \left\{\mathrm{trace}(M_d(\mathbf{y}^\star_{d,\delta})) = \mathrm{trace}(M_d(\mathbf{y}^\star))\right\},$$

- *or*

$$\left\{\mathrm{trace}(M_d(\mathbf{y}^\star)) - ||\mathbf{v}_d(\cdot)||_2^2 \notin \mathscr{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathscr{X})\right\} \Leftrightarrow \left\{\mathrm{trace}(M_d(\mathbf{y}^\star_{d,\delta})) - ||\mathbf{v}_d(\cdot)||_2^2 > 0 \text{ on } \mathscr{X}\right\}$$
$$\Leftrightarrow \left\{\mathrm{trace}(M_d(\mathbf{y}^\star_{d,\delta})) > \mathrm{trace}(M_d(\mathbf{y}^\star))\right\}.$$

*Recall that* $\mathrm{trace}(M_d(\mathbf{y}^\star)) = \max_{x\in\mathscr{X}} ||\mathbf{v}_d(x)||_2^2$.

### 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 $\mathfrak{F}(x) = \mathfrak{A}\,\mathbf{v}_d(x)$ for all $x \in \mathscr{X}$. 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., $\mathbf{y} \in \mathscr{M}_{2d}(\mathscr{X}),\ y_0 = 1$ in the ideal problem and $\mathbf{y} \in \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X}),\ y_0 = 1$ in the SDP relaxation scheme. We recall the notation $\mathfrak{I}_d(\mathbf{y}) := \mathfrak{A}M_d(\mathbf{y})\mathfrak{A}^\top$ and we get that the KKT conditions are given by

$$\forall x \in \mathscr{X}, \quad \phi_q(\mathfrak{I}_d(\mathbf{y})) - \underbrace{\mathfrak{F}(x)^\top \nabla\phi_q(\mathfrak{I}_d(\mathbf{y}))\,\mathfrak{F}(x)}_{p^\star_d(x)} = p^\star(x)$$

where $p^\star \in \mathscr{M}_{2d}(\mathscr{X})^\star$ ($= \mathscr{P}_{2d}(\mathscr{X})$) in the ideal problem and $p^\star \in \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X})^\star$ ($= \mathscr{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathscr{X})$) in the SDP relaxation scheme. Our analysis leads to the following equivalence results in this case.

**Proposition 6.** *Let $q \in (-\infty, 1)$ and let $\mathscr{X} \subseteq \mathbb{R}^n$ be a compact semi-algebraic set as defined in (4) and with nonempty interior. Problem (13) is a convex optimization problem with an optimal solution $\mathbf{y}^\star \in \mathscr{M}_{2d}(\mathscr{X})$. Denote by $p^\star_d$ the polynomial*

$$x \mapsto p^\star_d(x) := \mathfrak{F}(x)^\top \mathfrak{I}_d(\mathbf{y})^{q-1}\mathfrak{F}(x) = ||\mathfrak{I}_d(\mathbf{y})^{\frac{q-1}{2}}\mathfrak{F}(x)||_2^2. \tag{26}$$

*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*

$$\Omega := \{x \in \mathscr{X} : \mathrm{trace}(\mathfrak{I}_d(\mathbf{y})^q) - p^\star_d(x) = 0\}.$$

*In particular, the following statements are equivalent:*

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

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

The SDP relaxation is given by the program

$$\begin{aligned} \rho_\delta = \max_{\mathbf{y}} &\quad \phi_q(\mathfrak{I}_d(\mathbf{y})) \\ \text{s.t.} &\quad \mathbf{y} \in \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X}),\ y_0 = 1, \end{aligned} \tag{27}$$

for which we prove the following result.

**Proposition 7.** *Let $q \in (-\infty, 1)$ and let $\mathscr{X} \subseteq \mathbb{R}^n$ 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 $p^\star_d$ be as defined in (26), associated with $\mathbf{y}^\star$. Then $p^\star := \operatorname{trace}(\mathfrak{I}_d(\mathbf{y}^\star_{d,\delta})^q) - p^\star_d(x) \geqslant 0$ on $\mathscr{X}$ and $L_{\mathbf{y}^\star_{d,\delta}}(p^\star) = 0$.*

*In particular, the following statements are equivalent:*

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

*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 $\mathbf{y}^\star \mathscr{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathscr{X})$, 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 [28], and a second one which uses properties of the Christoffel-like polynomial associated with $\mathbf{y}^\star$.

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

### *5.1. Via the Nie method*

This approach to recover a measure from its moments is based on a formulation proposed by Nie in [28].

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

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

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 $\mathscr{X}$, and again $v_j = \lceil d_j/2 \rceil$, $j = 1, \dots, m$. We check whether the optimal solution $\mathbf{y}^\star_r$ of (28) satisfies the rank condition

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

where $v := \max_j v_j$. Indeed if (29) holds then $\mathbf{y}^\star_r$ is the sequence of moments (up to order $2r$) of a measure supported on $\mathscr{X}$; see [16, Theorem 3.11, p. 66]. If the test is passed, then we stop, otherwise we increase $r$ by one and repeat the procedure. As $\mathbf{y}^\star \in \mathscr{M}_{2d}(\mathscr{X})$, the rank condition (29) is satisfied for a sufficiently large value of $r$.

We extract the support points $x_1, \dots, x_\ell \in \mathscr{X}$ of the representing atomic measure of $\mathbf{y}^\star_r$, and $\mathbf{y}^\star$ respectively, as described in [16, Section 4.3].

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

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

instead of using a random positive polynomial on $\mathscr{X}$. 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 Christoffel-like polynomials

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-like 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 $\Omega = \{x \in \mathscr{X} : \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) = 0\}$, which due to Theorem 3 is the support of the atomic representing measure.

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

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

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 $\mathscr{X}$. 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 $x_1,\dots,x_\ell \in \mathscr{X}$ of the representing atomic measure of $\mathbf{y}_r^\star$, and $\mathbf{y}^\star$ respectively, as described in [16, Section 4.3].

Alternatively, we can solve the SDP

$$\begin{array}{ll} \min\limits_{\mathbf{y}_r} & \operatorname{trace}(M_{d+r}(\mathbf{y})) \\ \text{s.t.} & L_{\mathbf{y}_r}(p^\star) = 0, \\ & M_{d+r}(\mathbf{y}_r) \succcurlyeq 0, \quad y_{r,0} = 1, \\ & M_{d+r-v_j}(g_j\,\mathbf{y}_r) \succcurlyeq 0, \quad j = 1,\dots,m, \end{array} \tag{31}$$

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 $\{x_1,\dots,x_\ell\}$ of the atomic representing measure by one of the previously presented methods, we might be interested in also computing the corresponding weights $\omega_1,\dots,\omega_\ell$. 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.*, $\int_{\mathscr{X}} x^\alpha \mu^\star(dx) = y_\alpha^\star$.

## 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. We concentrate on $D$-optimal designs ($q = 0$) and $T$-optimal designs ($q = 1$).

All examples are modeled by GloptiPoly 3 [9] and YALMIP [24] and solved by MOSEK 7 [27] 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 $\mathscr{X} = [-1, 1]$ and on it the polynomial measurements $\sum_{j=0}^{d} \theta_j x^j$ with unknown parameters $\theta \in \mathbb{R}^{d+1}$.

### 6.1.1. D-optimal design

To compute the $D$-optimal design we first solve Problem (17), in other words

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

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

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

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.,* [6, 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 $x \mapsto p^\star(x) = 6 - p^\star_5(x)$, where $p^\star_5$ is the Christoffel polynomial of degree $2d = 10$ on $\mathscr{X}$ 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.

### 6.1.2. T-optimal design

For the $T$-optimal design, instead of solving (32) we solve

$$\begin{array}{rl} \max\limits_{\mathbf{y}_\delta} & \operatorname{trace} M_d(\mathbf{y}_\delta) \\ \text{s.t.} & M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0, \\ & M_{d+\delta-1}(1-\|x\|^2)\mathbf{y}_\delta) \succcurlyeq 0, \\ & y_{\delta,0} = 1 \end{array}$$

for $\mathbf{y}_\delta \in \mathbb{R}^{s(2(d+\delta))}$ and given $d$ and $\delta$. Then we take $\mathbf{y}^\star := (y^\star_{\delta,\alpha})_{|\alpha|\leqslant 2d}$. For example, for $d=3$ and $\delta=0$ we get the sequence $\mathbf{y}^\star \approx (1, 0, 1, 0, 1, 0, 1)^\top$.

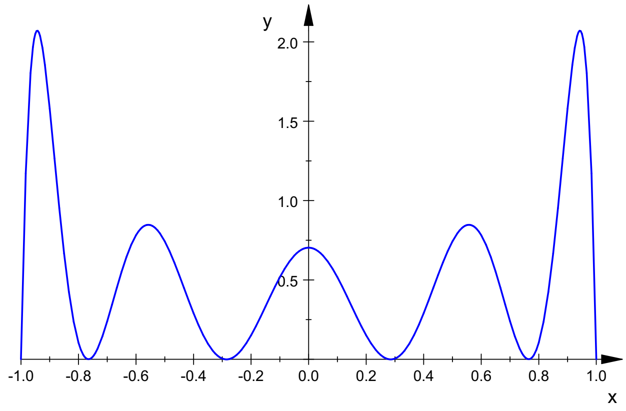


FIGURE 1. *Polynomial $p^\star$ for Example 6.1.1.*

To recover the corresponding measure we solve Problem (33) as before. In contrary to the $D$-optimal design, we get the points -1 and 1 with weights $\omega_1 = \omega_2 = 0.5$ for all $d \leq 7$, meaning our $T$-optimal design is the measure $\mu^\star = 0.5\,\delta_{-1} + 0.5\,\delta_1$ independently of the regression order $d$.

The measure $\mu^\star$ is also the $D$-optimal design for regression order $d = 1$, but for the $D$-optimal design the support gets larger when $d$ increases.

Note that we can also compute the support points via Christoffel-like polynomials. Evidently, this gives the same points. Generally, in the case of $T$-optimal design the method via Christoffel-like polynomials is numerically more stable than in the $D$-optimal design case, since $p_d^\star(x) = \mathbf{v}_d(x)^\top \mathbf{v}_d(x)$ for $q = 1$. Hence, $p_d^\star$ is independent of the moment matrix and it is not necessary to compute its inverse or the orthogonal polynomials.

### 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

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

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

#### 6.2.1. *D-optimal design*

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 $\mathscr{X}$ 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 $d$.

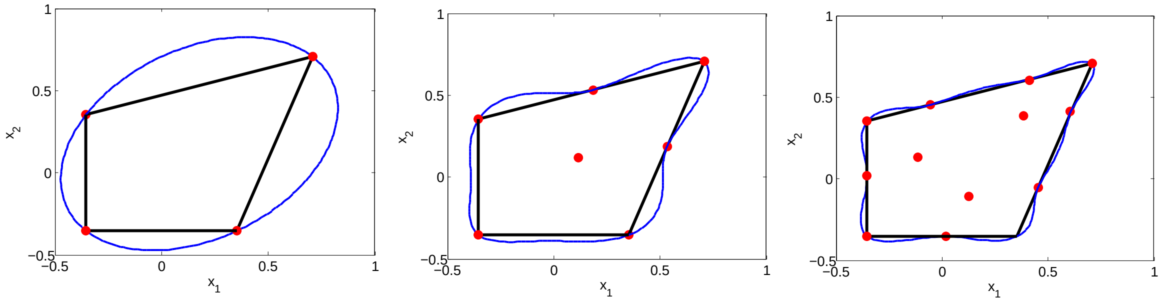


FIGURE 2. *The polygon (bold black) of Example 6.2, the support of the optimal design measure (red points) and the $\binom{2+d}{2}$-level set of the Christoffel polynomial (thin blue) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*

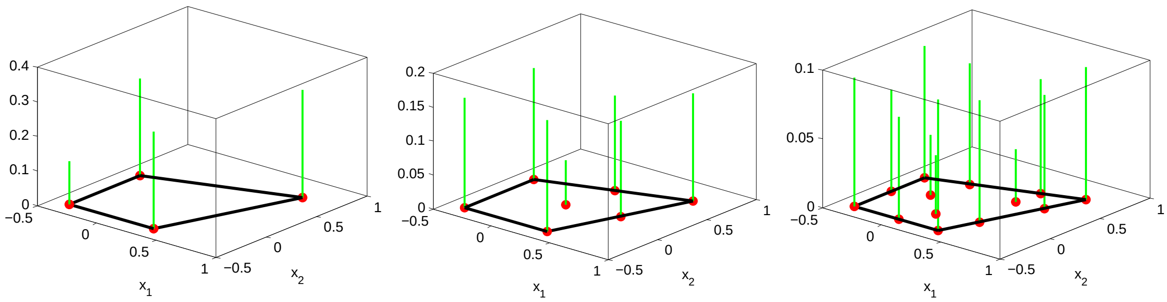


FIGURE 3. *The polygon (bold black) of Example 6.2 and the support of the optimal design measure (red points) with the corresponding weights (green bars) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*

#### 6.2.2. *T-optimal design*

For the $T$-optimal design, we obtain a moment sequence $\mathbf{y}^\star \in \mathbb{R}^{45}$ with a rank 1 moment matrix, when solving Problem (17) with the trace as objective function and with $d = 1$ and $\delta = 3$. Hence, when recovering the corresponding measure, we obtain a Dirac measure supported on one point, namely the point (0.7,0.7) which is the right most vertex of the polygon. For $d = 2$ and $d = 3$ we also receive moment matrices of rank one and recover the same support point. See Figure 5 for an illustration of the recovered point and the respective zero level set of the polynomial $p^\star$.

### 6.3. *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 $\mathscr{X} = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$.

#### 6.3.1. *D-optimal design*

Again, we first solve Problem (17). For $d = 1$ and $\delta \geq 0$ we obtain the sequence $\mathbf{y}^\star \in \mathbb{R}^{10}$ with $y^\star_{000} = 1$, $y^\star_{200} = y^\star_{020} = y^\star_{002} = 0.333$ 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 $\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\} \subseteq \mathscr{X}$ on which the optimal measure $\mu \in \mathscr{M}_+(\mathscr{X})$ 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 $\delta = 0$.

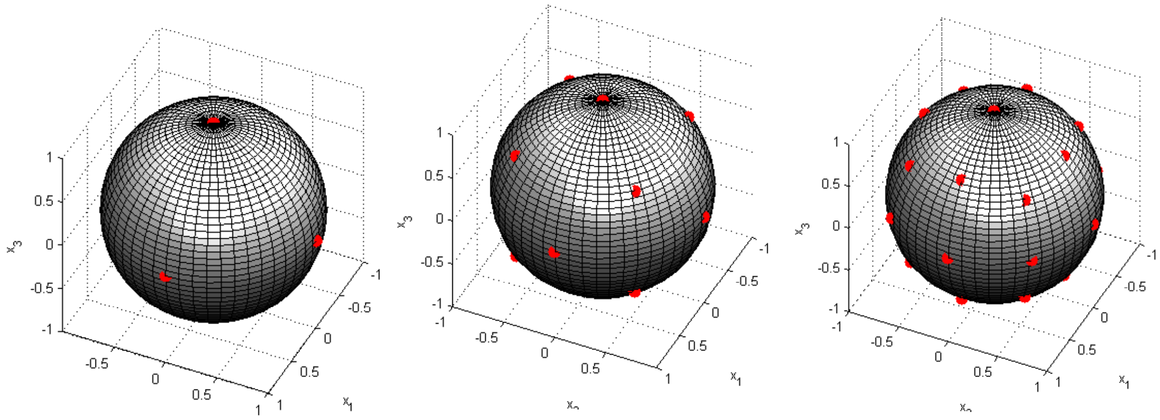


FIGURE 4. *The red points illustrate the support of the optimal design measure for* $d = 1$ *(left),* $d = 2$ *(middle),* $d = 3$ *(right) and* $\delta = 0$ *for Example 6.3.*

Using the method via Christoffel-like polynomials gives again less points. No solution is extracted when solving Problem (31) and we find only two supporting points for Problem (30).

### 6.3.2. *T-optimal design*

For this example, the solutions for the $T$-optimal design and the $D$-optimal design coincide for regression order $d = 1$. As for the other examples, here again the $T$-optimal design gives the same solution for all $d$, so we obtain the same six support points as for $d = 1$ when choosing $d = 2$ or $d = 3$.

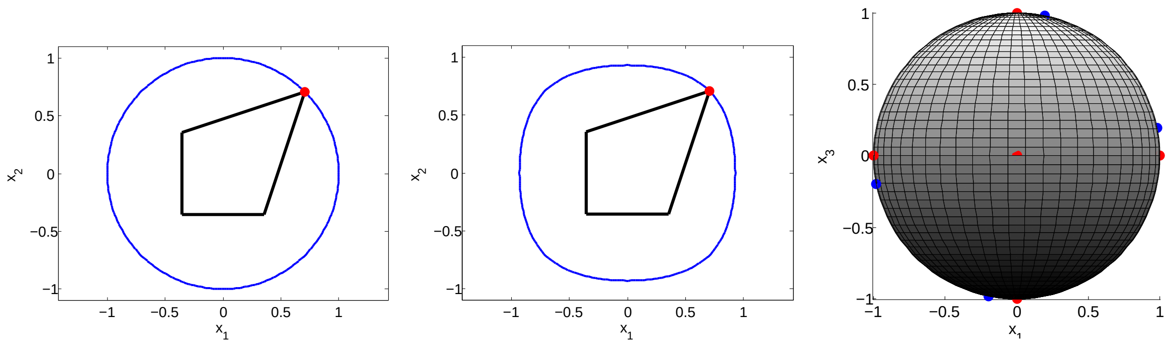


FIGURE 5. *The two pictures showing the polygon illustrate the recovered point (red) and the zero level set of the polynomial* $p^\star$ *for the* $T$*-optimal design case of Example 6.2; we have* $d = 1$ *on the left and* $d = 3$ *in the middle. The picture on the right shows the support points recovered in Example 6.3 for the D-optimal design and* $d = 1$ *(red) and the points which are recovered when additionally fixing some moments as described in Subsection 6.3.3 (blue).*

### 6.3.3. *Fixing some moments*

Our method has an additional nice feature. Indeed in Problem (17) one may easily include the additional constraint that some moments $(y_\alpha)$, $\alpha \in \Gamma \subset \mathbb{N}^n_{2d}$ 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 the picture on the right hand side of Figure 5 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).

## 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 [26]). In future works, we will extend the method to multi-response polynomial regression problems and to general smooth parametric regression models by linearization.

### *Acknowledgments*

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

## Appendix A: Proof of Theorem 1

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

In addition, the feasible set $\{\mathbf{y} \in \mathscr{M}_{2d}(\mathscr{X}) : y_0 = 1\}$ of Problem (15) is compact. Indeed there exists $M > 1$ such that it holds $\int_{\mathscr{X}} x_i^{2d}\, d\mu < M$ for every probability measure $\mu$ on $\mathscr{X}$ and every $i = 1, \ldots, n$. Hence, $\max\{y_0, \max_i\{L_y(x_i^{2d})\}\} < M$ which by [18] implies that $|y_\alpha| \le M$ for every $|\alpha| \le 2d$, which in turn implies that the feasible set of (15) is compact.

Next, as the function $\phi_q$ is upper semi-continuous, the supremum in (15) is attained at some optimal solution $\mathbf{y}^\star \in \mathscr{M}_{2d}(\mathscr{X})$. Moreover, as the feasible set is convex and $\phi_q$ is strictly concave (see, e.g., [31, 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 $\mathrm{int}(\mathscr{M}_{2d}(\mathscr{X}))$ is nonempty by Lemma 2.6 in [17]. Hence, Slater's condition[4] holds for (15). Further, by a an argument in [31, Chapter 7.13], the matrix $M_d(\mathbf{y}^\star)$ is non-singular. Therefore, $\phi_q$ is differentiable at $\mathbf{y}^\star$. Since additionally Slater's condition is fulfilled and $\phi_q$ is concave, this implies that the *Karush-Kuhn-Tucker* (KKT) optimality conditions[5] at $\mathbf{y}^\star$ are necessary (and sufficient) for $\mathbf{y}^\star$to be an optimal solution.

The KKT-optimality conditions read

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

[4]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 $C$.

[5]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 $\langle x, u^\star\rangle = 0$.

(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 reads $\langle \mathbf{y}^\star, p^\star \rangle = 0$.

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

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

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

$$\Big(1_{\alpha=0}\,\lambda^\star - \langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \mathbf{B}_\alpha \rangle\Big)_{|\alpha| \le 2d} = \hat{\mathbf{p}}, \quad \hat{p}^\star \in \mathscr{P}_{2d}(\mathscr{X}). \tag{34}$$

Multiplying (34) term-wise by $y^\star_\alpha$, summing up and invoking the complementarity condition, yields

$$\lambda^\star = \lambda^\star\, y_0^\star \overset{(34)}{=} \Big\langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \sum_{|\alpha| \le 2d} y^\star_\alpha \mathbf{B}_\alpha \Big\rangle = \Big\langle \nabla \phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \Big\rangle = \phi_q(M_d(\mathbf{y}^\star)), \tag{35}$$

where the last equality holds by Euler formula for the positively homogeneous function $\phi_q$.

Similarly, multiplying Equation (34) term-wise by $x^\alpha$ and summing up yields for all $x \in \mathscr{X}$

$$x \mapsto \hat{p}^\star(x) \overset{(34)}{=} \lambda^\star - \Big\langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \sum_{|\alpha| \le 2d} \mathbf{B}_\alpha x^\alpha \Big\rangle = \lambda^\star - \Big\langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \mathbf{v}_d(x)\mathbf{v}_d(x)^\top \Big\rangle \ge 0.$$

For $q \neq 0$ let $c^\star := \binom{n+d}{n}\Big[\binom{n+d}{n}^{-1} \operatorname{trace}(M_d(\mathbf{y}^\star)^q)\Big]^{1-\frac{1}{q}}$. 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

$$\begin{aligned} c^\star \lambda^\star &\overset{(35)}{=} c^\star \langle \nabla \phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \rangle = \operatorname{trace}(M_d(\mathbf{y}^\star)^q) \\ \text{and} \quad c^\star \langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \mathbf{v}_d(x)\mathbf{v}_d(x)^\top \rangle &\overset{(16)}{=} p_d^\star(x) \end{aligned}$$

Therefore, equation (36) is equivalent to $p^\star := c^\star\, \hat{p}^\star = c^\star\, \lambda^\star - p_d^\star \in \mathscr{P}_{2d}(\mathscr{X})$. To summarize,

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

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

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

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

The measure $\mu^\star$ is an atomic measure supported on at most $\binom{n+2d}{n}$ points. This follows from Tchakaloff's theorem [16, Theorem B.12], which states that for every finite Borel probability measure on $\mathscr{X}$ and every $s \in \mathbb{N}$, there exists an atomic measure $\mu_s$ supported on $\ell \le \binom{n+s}{n}$ points such that all moments of $\mu_s$ and $\mu^\star$ agree up to order $s$. For $s = 2d$ we get that $\ell \le \binom{n+2d}{n}$. If $\ell < \binom{n+d}{n}$, then $\operatorname{rank} M_d(\mathbf{y}^\star) < \binom{n+d}{n}$ in contradiction to $M_d(\mathbf{y}^\star)$ being non-singular. Therefore, $\binom{n+d}{n} \le \ell \le \binom{n+2d}{n}$.

**References**


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

[2] R. A. Bailey. Design of comparative experiments, volume 25. Cambridge University Press, 2008.

[3] G. E. P. Box, J. S. Hunter, and W. G. Hunter. Statistics for experimenters. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2005. Design, innovation, and discovery.

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

[5] H. Dette and W. J. Studden. Geometry of $E$-optimality. The Annals of Statistics, pages 416–433, 1993.

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

[7] G. Elfving. Optimum allocation in linear regression theory. The Annals of Mathematical Statistics, 23(2):255–262, 1952.

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

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

[10] L. A. Imhof and W. J. Studden. E-optimal designs for rational models. Annals of statistics, pages 763–783, 2001.

[11] S. Karlin and W. J. Studden. Optimal experimental designs. The Annals of Mathematical Statistics, 37(4):783–815, 1966.

[12] S. Karlin and 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.

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

[14] 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.

[15] S. Kuriki and H. P. Wynn. Optimal experimental design that minimizes the width of simultaneous confidence bands. ArXiv e-prints, Apr. 2017.

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

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

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

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

[20] T.-S. Lau and W. Studden. Optimal designs for trigonometric and polynomial regression using canonical moments. The Annals of Statistics, 13(1):383–394, 1985.

[21] 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.

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

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

[24] 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.

[25] P. Maréchal, J. J. Ye, and J. Zhou. K-optimal design via semidefinite programming and entropy optimization. Mathematics of Operations Research, 2014.

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

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

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

[29] L. Pronzato and W. G. Müller. Design of computer experiments: space filling and beyond. Statistics and Computing, 22(3):681–701, 2012.

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

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

[32] 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.

[33] 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.

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

[35] R. Schwabe. Optimum designs for multi-factor models, volume 113. Springer Science & Business Media, 2012.

[36] 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.

[37] E. Walter, L. Pronzato, and J. Norton. Identification of parametric models from experimental data, volume 1. Springer Berlin, 1997.

[38] H. P. Wynn. The sequential generation of $D$-optimum experimental designs. The Annals of Mathematical Statistics, 41(5):1655–1664, 1970.