

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

K-optimal design via semidefinite programming and entropy optimization

Pierre Maréchal

Institut de Mathématiques de Toulouse, Université de Toulouse, France pr.marechal@gmail.com,
<http://www.math.univ-toulouse.fr/~marechal/>

Jane J. Ye and Julie Zhou

Department of Mathematics and Statistics, University of Victoria, Victoria, B.C. Canada V8W 2Y2,
janeje@uvic.ca, jzhou@uvic.ca, <http://www.math.uvic.ca/faculty/janeje/>, <http://www.math.uvic.ca/~jzhou/>

In this paper, we consider the problem of optimal design of experiments. A two step inference strategy is proposed. The first step consists in minimizing the condition number of the so-called information matrix. This step can be turned into a semidefinite programming (SDP) problem. The second step is more classical, and it entails the minimization of a convex integral functional under linear constraints. This step is formulated in some infinite-dimensional space and is solved by means of a dual approach. Numerical simulations will show the relevance of our approach.

Key words: Optimal design of experiments, condition numbers, semidefinite programming, entropy optimization, Fenchel duality, Chebyshev polynomials.

MSC2000 subject classification: Primary: 90C26, 90C30, 62K05; secondary: 90C22

OR/MS subject classification: Primary: design of experiments, entropy; secondary: nondifferentiable, programming, algorithms

History: Date submitted: 7/29/2012; Dates revised: 7/23/2013, 3/23/2014

1. Introduction. We consider the parametric regression model

$$y = \sum_{j=0}^p \theta_j f_j(x) + \varepsilon, \quad x \in S,$$

in which $y \in \mathbb{R}$ is the *response variable*, $x \in S \subset \mathbb{R}^k$ is the *design variable*, ε is a random error with mean 0 and variance σ^2 , f_j , $j = 0, \dots, p$ are given *basis functions* supported in S and the θ_j , $j = 0, \dots, p$ are parameters to be estimated.

Example 1.1. If $S = [-1, 1]$ and $f_j(x) = x^j$ as in [28], one speaks of *p-th order polynomial regression* on $[-1, 1]$.

We shall use the notation $f(x) = (f_0(x), \dots, f_p(x))^\top$ and $\theta = (\theta_0, \dots, \theta_p)^\top$.

A frequent problem in statistical sciences is that of defining a probability measure ξ on S that will allow for *optimal* estimation of the function

$$\varphi(x) := \sum_{j=0}^p \theta_j f_j(x) = \theta^\top f(x).$$

By *optimal*, we mean several possible criteria, all of them having to do with the so-called *information matrix*

$$A(\xi) := \int_S f(x)f(x)^\top d\xi(x). \quad (1)$$

This matrix stems from the least squares estimation of the parameter θ . As a matter of fact, in the optimization problem

$$\left| \begin{array}{l} \text{Min } h(\theta) := \int_S (y(x) - \theta^\top f(x))^2 d\xi(x) \\ \text{s.t. } \theta \in \mathbb{R}^{p+1}, \end{array} \right.$$

the objective function can be written as

$$\begin{aligned} h(\theta) &= \int_S (y(x)^2 + (f(x)^\top \theta)^2 - 2y(x)f(x)^\top \theta) d\xi(x) \\ &= \int_S y(x)^2 d\xi(x) + \theta^\top \int_S f(x)f(x)^\top d\xi(x) \theta - 2 \int_S y(x)f(x)^\top d\xi(x) \theta \\ &= c(\xi) + \langle \theta, A(\xi)\theta \rangle - 2\langle \psi(\xi), \theta \rangle, \end{aligned}$$

in which we have defined

$$c(\xi) := \int_S y(x)^2 d\xi(x) \quad \text{and} \quad \psi(\xi) := \int_S y(x)f(x) d\xi(x).$$

The standard first order necessary optimality condition (which is also sufficient by convexity of h) then reads:

$$A(\xi)\theta = \psi(\xi). \quad (2)$$

Therefore, estimating θ entails solving the above linear system. As is well known, the condition number of the matrix $A(\xi)$ measures the maximum amount by which a perturbation in an experiment measurement $y(x)$ will be transmitted to the unknown regression parameter θ . In order to minimize the error sensitivity, it is desirable to find a design ξ^* which minimizes the condition number of the information matrix $A(\xi)$, denoted throughout by $\kappa(A(\xi))$. This suggests to select ξ^* by solving the optimization problem:

$$(\mathcal{M}) \left| \begin{array}{l} \text{Min } \kappa(A(\xi)) \\ \text{s.t. } \xi \in \Pi, \end{array} \right.$$

in which Π denotes the (convex) set of all probability measures on S and κ denotes the condition number. In [28], a solution to (\mathcal{M}) was investigated under the form of a finite combination of Dirac measures $\sum_i p_i \delta_{x_i}$. Such a parametrization (by both locations x_i and probabilities p_i) introduces a rather high level of nonlinearity in $A(\xi)$. It is well-known that the maximum and the minimum eigenvalues for a positive definite matrix are in general nonsmooth and consequently

so is the condition number. Since the condition number is only quasiconvex and not convex, the condition number minimization problem is a nonsmooth and nonconvex optimization problem and nonsmooth optimization techniques are usually needed to solve such a problem (see [16, 6]). In [28], it was shown that for the special case where the design space is $[-1, 1]$ and the regression is the polynomial, the condition number is a smooth function and hence the standard smooth optimization techniques can be used.

In the present paper, we propose to divide the problem into two subproblems: first choose among all matrices of the form (1) a matrix \bar{A} with minimum condition number, and then infer a probability measure that is compatible with the previous solution. Notice that, by minimizing the condition number of the real symmetric matrix $A(\xi)$, we are likely to obtain a full rank matrix.

The first step of our inference process consists in solving the problem

$$(\mathcal{Q}) \left| \begin{array}{l} \text{Min } \kappa(A) \\ \text{s.t. } A \in \mathcal{S} := \{A(\xi) | \xi \in \Pi\}. \end{array} \right.$$

Following [16], we define the condition number of a real symmetric matrix A as

$$\kappa(A) := \begin{cases} \lambda_{\max}(A)/\lambda_{\min}(A) & \text{if } \lambda_{\min}(A) > 0, \\ \infty & \text{if } \lambda_{\min}(A) = 0 \text{ and } \lambda_{\max}(A) > 0, \\ 0 & \text{if } A = 0, \end{cases}$$

in which $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are respectively the largest and smallest eigenvalues. Note that the set \mathcal{S} is convex, as the image of the convex set Π by the linear mapping $\xi \mapsto A(\xi)$. Mild assumptions on f will also ensure compactness of \mathcal{S} , so that Problem (Q) falls into the class of problems studied in [16, 13]. We may also make the following

Standing assumption: The set \mathcal{S} contains nonsingular matrices.

Under this assumption, our minimization problem will produce full rank optimal solutions.

Remark 1.1. If we are given a family of functions $\mathcal{F} := \{f_0, \dots, f_p\}$ and a measure ξ such that \mathcal{F} is orthonormal with respect to the scalar product

$$\langle g_2, g_1 \rangle := \int_S g_1(x)g_2(x) d\xi(x),$$

then ξ is automatically K -optimal. As a matter of fact, it is readily seen that, in this case, the information matrix is the identity. An occurrence of this will be dealt with in Proposition 4.1 below. In general however, in statistics, the set of *basis functions* is given, and it may not exist a measure making this set orthonormal.

The second stage of our inference process will then consist in inferring a representing probability measure for the optimal moment matrix \bar{A} . We suggest to use the maximum entropy principle, that is, to select a probability measure $\bar{\xi}$ by solving the optimization problem

$$(\mathcal{E}_0) \left| \begin{array}{l} \text{Min } \mathcal{H}(\xi) \\ \text{s.t. } \xi \in \Pi, A(\xi) = \bar{A}. \end{array} \right.$$

Here, the entropy $\mathcal{H}(\xi)$ is defined as usual by

$$\mathcal{H}(\xi) = \begin{cases} \int_S \ln \frac{d\xi}{d\lambda}(x) d\xi(x) & \text{if } \xi \ll \lambda \\ \infty & \text{otherwise,} \end{cases}$$

in which $d\xi/d\lambda$ denotes the Radon-Nikodym derivative of ξ with respect to the Lebesgue measure λ and $\xi \ll \lambda$ means that the measure ξ is absolutely continuous with respect to the measure λ .

We emphasize that, in the above form, our entropy problem may not be solvable. The reason for this is twofold: first, nothing guarantees that \bar{A} has a representing measure that is absolutely continuous with respect to the Lebesgue measure; second, the matrix \bar{A} will result from numerical computations, and therefore may not satisfy rigorously the constraint $\bar{A} \in \mathcal{S}$. Consequently, we must relax Problem (\mathcal{E}_0) and consider instead a problem of the form

$$(\mathcal{E}) \quad \begin{cases} \text{Min } \mathcal{K}(\xi) + \frac{1}{2\alpha} \|A(\xi) - \bar{A}\|^2 \\ \text{s.t. } \xi \in \Pi. \end{cases}$$

In Remark 3.1 below, we shall further justify this relaxation by showing that the relaxed problems (which always have solutions) *approximate* the unrelaxed one, in a sense to be specified.

In this paper, we concentrate on the case where $S = [a, b]$ and the components of $f(x)$ form a family of polynomials that is graduated in degree. In this case, the information matrix depends linearly on the moments $y = (y_0, \dots, y_p)$, i.e.,

$$A(\xi) = A[y] = \sum_{k=0}^{2p} y_k A^{(k)}$$

for some matrices $A^{(k)}$. For example, if $f(x) = (1, x, \dots, x^p)^\top$, then the information matrix is given by

$$A(\xi) = H(y_0, \dots, y_{2p}) := \begin{bmatrix} y_0 & y_1 & \dots & y_p \\ y_1 & y_2 & \dots & y_{p+1} \\ \vdots & \vdots & & \vdots \\ y_p & y_{p+1} & \dots & y_{2p} \end{bmatrix}, \quad \text{with } y_k := \int_a^b x^k d\xi(x).$$

Throughout, $H(y_0, \dots, y_{2p})$ will denote the Hankel matrix associated with the sequence (y_0, \dots, y_{2p}) . The Chebyshev polynomial model, considered in Section 4 below, provides another such example.

2. K-optimal moments. In this section, we show that, since the information matrix $A(\xi)$ depends linearly on the moments, (\mathcal{Q}) can be written as an SDP problem, so that K -optimal moments can be efficiently computed. This relies in part on well-known results from the theory of truncated moment problems. Our first task is to obtain a more tractable form of the set \mathcal{S} .

Theorem 2.1. *A finite sequence $y = (y_0, \dots, y_{2p})$ is the sequence of moments of a measure ξ on $[a, b]$ if and only if the following two conditions hold:*

- (a) $H(y_0, \dots, y_{2p}) \succeq 0$;
- (b) $-H(y_2, \dots, y_{2p}) + (a+b)H(y_1, \dots, y_{2p-1}) - abH(y_0, \dots, y_{2p-2}) \succeq 0$.

PROOF. See, for example, [12], Theorem 5.39. ■

We are therefore able to describe the set \mathcal{S} in terms of *Linear Matrix Inequalities* (LMI). It is readily seen that Conditions (a) and (b) in Theorem 2.1 can be respectively written as

$$\sum_{k=0}^{2p} y_k F^{(k)} \succeq 0 \quad \text{and} \quad \sum_{k=0}^{2p} y_k G^{(k)} \succeq 0. \quad (3)$$

Here, $F^{(k)} \in \mathbb{R}^{(p+1) \times (p+1)}$ is the k -th *elementary Hankel matrix*, that is, $F_{i,j}^{(k)} = 1$ if $i + j = k$ and $F_{i,j}^{(k)} = 0$ elsewhere, and $G^{(k)} \in \mathbb{R}^{p \times p}$ is the sequence of matrices defined as follows:

$$G^{(0)} := \begin{bmatrix} -ab & 0 \\ 0 \end{bmatrix}, \quad G^{(1)} := \begin{bmatrix} a+b & -ab & 0 \\ -ab & 0 \\ 0 \end{bmatrix},$$

$$G^{(2)} := \begin{bmatrix} -1 & a+b & -ab & 0 \\ a+b & -ab & 0 \\ -ab & 0 \\ 0 \end{bmatrix}, \quad G^{(3)} := \begin{bmatrix} 0 & -1 & a+b & -ab & 0 \\ -1 & a+b & -ab & 0 \\ a+b & -ab & 0 \\ -ab & 0 \\ 0 \end{bmatrix},$$

and so on. In the above matrices, every non written entry is zero by convention. We emphasize that the feasible set \mathcal{S} is a convex cone.

Next, we rewrite Problem (\mathcal{Q}) . Since the null matrix is infeasible in Problem (\mathcal{Q}) , the latter may be rewritten as

$$\begin{cases} \text{Min } \kappa := t/s \\ \text{s.t. } t = \lambda_1(A[y]), \quad s = \lambda_{p+1}(A[y]), \\ A[y] \in \mathcal{S}. \end{cases}$$

Here, t/s is understood to be ∞ if $s = 0$. Next, we may replace equalities, in the first line of the constraint, by inequalities. We obtain the following equivalent optimization problem:

$$\begin{cases} \text{Min } \kappa := t/s \\ \text{s.t. } t \geq \lambda_1(A[y]), \quad s \leq \lambda_{p+1}(A[y]), \\ A[y] \in \mathcal{S}. \end{cases}$$

The first line of the constraint is equivalent to

$$sI \preceq A[y] \preceq tI, \quad (4)$$

in which I denotes the identity matrix of appropriate size. Now, under our standing assumption above, we may discard from \mathcal{S} all singular matrices. This amounts to replacing \mathcal{S} by $\mathcal{S}' := \mathcal{S} \cap S_{++}^{p+1}$ which, incidentally, is again a convex cone. Dividing (4) by s and making the change of variable $A' = A/s$ yields the following equivalent form of (\mathcal{Q}) :

$$\begin{cases} \text{Min } \kappa \\ \text{s.t. } I \preceq A'[y] \preceq \kappa I, \\ A'[y] \in \mathcal{S}'. \end{cases}$$

Notice that $A'[y] = A[y']$ with $y' := y/s$. We finally obtain the SDP problem in standard form

$$\left| \begin{array}{l} \text{Min } \langle (1, 0, \dots, 0), (\kappa, y'_0, \dots, y'_{2p}) \rangle \\ \text{s.t. } I - \sum_{k=0}^{2p} y'_k A^{(k)} \preceq 0, \\ \quad -\kappa I + \sum_{k=0}^{2p} y'_k A^{(k)} \preceq 0, \\ \quad -\sum_{k=0}^{2p} y'_k F^{(k)} \preceq 0, \\ \quad -\sum_{k=0}^{2p} y'_k G^{(k)} \preceq 0. \end{array} \right.$$

3. Entropy optimization. In this section, we study the entropy part of our inference process. We shall use duality techniques, in order to cope with both the constraints and the fact that the workspace is infinite dimensional. Before proceeding, we give a customary generalization of Problem (\mathcal{E}) , which will not induce much extra effort in the resolution.

3.1. Extension and reformulation. We may assume that some prior information on the measure to be inferred is available, under the form of a prior probability measure y . It is customary, in this case, to replace the entropy by Kullback-Leibler's relative entropy [11]. For $p \in [1, \infty]$, we denote by $L^p_\mu([a, b])$ the L^p -space on the measure space $([a, b], \mathcal{T}, \mu)$. Recall that the entropy of a probability measure ξ relative to a probability measure μ is given by

$$\mathcal{H}(\xi \parallel \mu) := \begin{cases} \int u(x) \ln u(x) d\mu(x) & \text{if } \xi \ll \mu, \\ \infty & \text{otherwise,} \end{cases}$$

where u denotes the Radon-Nikodym derivative of ξ with respect to μ . Clearly, the choice μ being equal to the Lebesgue measure λ gives rise to Problem (\mathcal{E}) stated in the introduction section. Moreover, minimizing $\mathcal{H}(\cdot \parallel \mu)$ will enforce absolute continuity of the solution with respect to μ , so we are led to considering the following convex (but infinite dimensional) optimization problem

$$(\mathcal{P}) \quad \left| \begin{array}{l} \text{Min } K_\mu(u) + \frac{1}{2\alpha} \|\mathbb{A}u - \bar{A}\|^2 \\ \text{s.t. } u \in L^1_\mu([a, b]), 1 = \mathbb{I}u, \end{array} \right.$$

in which $\mathbb{I}u := \int_a^b u(x) d\mu(x)$, $\mathbb{A}u := \int_a^b f(x)f(x)^\top u(x) d\mu(x)$ and

$$K_\mu(u) := \int_a^b k_\circ(u(x)) d\mu(x), \quad \text{with } k_\circ(t) := \begin{cases} t \ln t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t < 0. \end{cases}$$

Since the functions f_j are continuous on $[a, b]$, the operator \mathbb{A} is well-defined on $L^1_\mu([a, b])$. Moreover, one should keep in mind that the variable u in the above problem is a Radon-Nikodym derivative with respect to the *reference measure* μ . Note that the Kullback-Leibler relative entropy is also used in Bayesian experimental design in a different manner; see [5, Section 2.2].

3.2. Reminder on partially finite convex programming. We shall use notions from convex analysis. Our reference books for convex analysis are [10, 21, 30]. In order to solve Problem (\mathcal{P}) , we use a dual strategy. The latter consists in the following steps:

- (1) write the dual problem of (\mathcal{P}) ;
- (2) study the qualification conditions;
- (3) establish the *primal-dual* relationship.

Let X be a real vector space with X^* as its algebraic dual space (i.e. the space of all linear functions on X). Recall that for a function $f : X \rightarrow [-\infty, \infty]$, the functions $f^* : X^* \rightarrow [-\infty, \infty]$ and $f_* : X^* \rightarrow [-\infty, \infty]$ defined by

$$f^*(x^*) := \sup\{\langle x, x^* \rangle - f(x) \mid x \in X\}$$

$$\text{and } f_*(x^*) := \inf\{\langle x, x^* \rangle - f(x) \mid x \in X\},$$

where $\langle x, x^* \rangle := x^*(x)$, are called the upper and lower conjugate or upper and lower Fenchel conjugate of f , respectively. The functions $f^{**}, f_{**} : X \rightarrow [-\infty, \infty]$ defined by

$$f^{**}(x) := \sup\{\langle x, x^* \rangle - f^*(x^*) \mid x^* \in X^*\}$$

$$\text{and } f_{**}(x) := \inf\{\langle x, x^* \rangle - f_*(x^*) \mid x^* \in X^*\}$$

are called the upper and lower biconjugate of f , respectively. Note that the upper conjugate is always a convex function and the lower conjugate is always a concave function.

The following theorem [15] is a partially finite version of Fenchel's duality theorem. The proof, given below for completeness, follows in essence the argument of [4].

Theorem 3.1. *Let U and V be real vector spaces and let $\langle \cdot, \cdot \rangle$ be a bilinear form on $U \times V$. Let $\mathbb{T} : U \rightarrow \mathbb{R}^d$ be a linear mapping with (formal) adjoint $\mathbb{T}^* : \mathbb{R}^d \rightarrow V$. Let $F : U \rightarrow (-\infty, \infty]$ be a convex proper function and let $g : \mathbb{R}^d \rightarrow [-\infty, \infty)$ be a concave proper function. If the condition*

$$(QC) \quad \text{ri}(\mathbb{T} \text{ dom } F) \cap \text{ri}(\text{dom } g) \neq \emptyset,$$

is satisfied, then

$$\eta := \inf\{F(u) - g(\mathbb{T}u) \mid u \in U\} = \max\{g_*(\boldsymbol{\lambda}) - F^*(\mathbb{T}^*\boldsymbol{\lambda}) \mid \boldsymbol{\lambda} \in \mathbb{R}^d\}.$$

PROOF. Throughout this proof, $\langle \cdot, \cdot \rangle$ denotes not only the bilinear form on $U \times V$ but also the standard scalar product on \mathbb{R}^d . For every $u \in U$, $\boldsymbol{\xi} \in \mathbb{R}^d$ and $\boldsymbol{\lambda} \in \mathbb{R}^d$, one has

$$\langle \boldsymbol{\xi}, \boldsymbol{\lambda} \rangle - g(\boldsymbol{\xi}) \geq g_*(\boldsymbol{\lambda}) \quad \text{and} \quad \langle u, \mathbb{T}^*\boldsymbol{\lambda} \rangle - F(u) \leq F^*(\mathbb{T}^*\boldsymbol{\lambda})$$

by definition of conjugacy. Letting $\boldsymbol{\xi} = \mathbb{T}u$ and subtracting the second inequality to the first one, we get $F(u) - g(\mathbb{T}u) \geq g_*(\boldsymbol{\lambda}) - F^*(\mathbb{T}^*\boldsymbol{\lambda})$ for every u and $\boldsymbol{\lambda}$. Taking the *infimum* with respect to u yields

$$\eta \geq \sup\{g_*(\boldsymbol{\lambda}) - F^*(\mathbb{T}^*\boldsymbol{\lambda}) \mid \boldsymbol{\lambda} \in \mathbb{R}^d\}. \tag{5}$$

If $\eta = -\infty$, the reverse inequality is trivial. Let us then assume that $\eta > -\infty$ and define on \mathbb{R}^d the *perturbation function* π by

$$\pi(\mathbf{y}) := \inf\{F(u) \mid \mathbf{y} = \mathbb{T}u\}. \tag{6}$$

We shall prove that

- (a) $\inf\{F(u) - g(\mathbb{T}u) \mid u \in U\} = \inf\{\pi(\mathbf{y}) - g(\mathbf{y}) \mid \mathbf{y} \in \mathbb{R}^d\}$;
- (b) π is proper convex with domain $\mathbb{T} \text{ dom } F$;

(c) $F^*(\mathbb{T}^*\boldsymbol{\lambda}) = \pi^*(\boldsymbol{\lambda})$ for every $\boldsymbol{\lambda} \in \mathbb{R}^d$.

The result will then follow from Fenchel's duality theorem (see [21, Theorem 31.1 page 327]) applied to the functions π and g .

One has:

$$\begin{aligned} & \inf\{F(u) - g(\mathbb{T}u) \mid u \in U\} \\ &= \inf\{F(u) - g(\mathbf{y}) \mid u \in U, \mathbf{y} \in \mathbb{R}^d, \mathbf{y} = \mathbb{T}u\} \\ &= \inf\{\inf\{F(u) - g(\mathbf{y}) \mid u \in U, \mathbf{y} = \mathbb{T}u\} \mid \mathbf{y} \in \mathbb{R}^d\} \\ &= \inf\{\pi(\mathbf{y}) - g(\mathbf{y}) \mid \mathbf{y} \in \mathbb{R}^d\}, \end{aligned}$$

and (a) is clear. Suppose next that one can find $\mathbf{y}_1, \mathbf{y}_2$ in \mathbb{R}^d and β_1, β_2 in \mathbb{R} such that $\beta_1 > \pi(\mathbf{y}_1)$ and $\beta_2 > \pi(\mathbf{y}_2)$. Then, there must exist u_1, u_2 in U such that $\mathbf{y}_1 = \mathbb{T}u_1$, $\mathbf{y}_2 = \mathbb{T}u_2$ and $\beta_1 > F(u_1)$, $\beta_2 > F(u_2)$. Hence, for every $\alpha \in (0, 1]$, we have

$$\begin{aligned} (1 - \alpha)\beta_1 + \alpha\beta_2 &> (1 - \alpha)F(u_1) + \alpha F(u_2) \\ &\geq F((1 - \alpha)u_1 + \alpha u_2) \\ &\geq \inf\{F(u) \mid (1 - \alpha)\mathbf{y}_1 + \alpha\mathbf{y}_2 = \mathbb{T}u\} \\ &= \pi((1 - \alpha)\mathbf{y}_1 + \alpha\mathbf{y}_2). \end{aligned}$$

The convexity of π then results from [21, Theorem 4.2 page 25]. Next, if $u \in \text{dom } F$, then $\mathbb{T}u \in \text{dom } \pi$, so that $\mathbb{T} \text{dom } F \subset \text{dom } \pi$. Conversely, if \mathbf{y} belongs to $\text{dom } \pi$, there exists $u \in U$ such that $\mathbf{y} = \mathbb{T}u$ and $F(u) < \infty$, we shows that $\text{dom } \pi \subset \mathbb{T} \text{dom } F$. Let us show that π is proper. From (QC), there exists $\hat{\mathbf{y}}$ in both $\text{ri } \mathbb{T} \text{dom } F = \text{ri dom } \pi$ and $\text{ri dom } g$. Then, (a) shows that $\eta < \infty$. Moreover, $\pi(\hat{\mathbf{y}}) > -\infty$, for otherwise (a) would contradict the working assumption that $\eta > -\infty$. Consequently, $\pi(\hat{\mathbf{y}})$ is finite. From [21, Theorem 7.2 page 53], π is proper, and (b) is established. Finally, one has:

$$\begin{aligned} \pi^*(\boldsymbol{\lambda}) &= \sup\{\langle \mathbf{y}, \boldsymbol{\lambda} \rangle - \pi(\mathbf{y}) \mid \mathbf{y} \in \mathbb{R}^d\} \\ &= \sup\{\langle \mathbf{y}, \boldsymbol{\lambda} \rangle - F(u) \mid \mathbf{y} \in \mathbb{R}^d, \mathbf{y} = \mathbb{T}u\}, \end{aligned}$$

$$\begin{aligned} \text{and } F^*(\mathbb{T}^*\boldsymbol{\lambda}) &= \sup\{\langle \mathbb{T}u, \boldsymbol{\lambda} \rangle - F(u) \mid u \in U\} \\ &= \sup\{\langle \mathbf{y}, \boldsymbol{\lambda} \rangle - F(u) \mid u \in U, \mathbf{y} = \mathbb{T}u\}, \end{aligned}$$

which shows that $\pi^*(\boldsymbol{\lambda}) = F^*(\mathbb{T}^*\boldsymbol{\lambda})$ and finishes the proof of the theorem. ■

The condition (QC) is called a *qualification condition*. The theorem establishes the equality between the optimal values of two optimization problems: the minimization of $F - g \circ \mathbb{T}$, called *primal problem*, and the maximization of $g_* - F^* \circ \mathbb{T}^*$, called *dual problem*. In the next theorem, we establish (for the favorable cases), a connection between the solutions of these problems, respectively called *primal* and *dual* solutions.

Theorem 3.2. *With the notation and assumptions of Theorem 3.1, assume that the condition*

$$(QC^*) \quad \text{ri dom } g_* \cap \text{ri dom}(F^* \circ \mathbb{T}^*) \neq \emptyset.$$

is satisfied, and that

- (a) $F^{**} = F$ and $g_{**} = g$;
- (b) *there exists a dual solution $\bar{\boldsymbol{\lambda}}$ as well as $\bar{u} \in \partial F^*(\mathbb{T}^*\bar{\boldsymbol{\lambda}})$ such that $\nabla(F^* \circ \mathbb{T}^*)(\bar{\boldsymbol{\lambda}}) = \mathbb{T}\bar{u}$.*

Then \bar{u} is a primal solution.

PROOF. Since $\bar{u} \in \partial F^*(\mathbb{T}^*\bar{\lambda})$ and $F^{**} = F$, we have:

$$F^*(\mathbb{T}^*\bar{\lambda}) + F(\bar{u}) = \langle \mathbb{T}^*\bar{\lambda}, \bar{u} \rangle. \quad (7)$$

The optimality of $\bar{\lambda}$ implies that $\mathbf{0}$ belongs to $\partial(g_* - F^* \circ \mathbb{T}^*)(\bar{\lambda})$. The condition (QC^*) implies, via a standard subdifferential calculus rule, that

$$0 \in \partial g_*(\bar{\lambda}) - \partial(F^* \circ \mathbb{T}^*)(\bar{\lambda}).$$

But $\partial(F^* \circ \mathbb{T}^*)(\bar{\lambda}) = \{\mathbb{T}\bar{u}\}$ by assumption, so that the optimality of $\bar{\lambda}$ reads

$$\mathbb{T}\bar{u} \in \partial g_*(\bar{\lambda}).$$

since $g_{**} = g$, we deduce that

$$g_*(\bar{\lambda}) + g(\mathbb{T}\bar{u}) = \langle \bar{\lambda}, \mathbb{T}\bar{u} \rangle. \quad (8)$$

Combining Equations (7) and (8), we obtain the equality

$$F(\bar{u}) - g(\mathbb{T}\bar{u}) = g_*(\bar{\lambda}) - F^*(\mathbb{T}^*\bar{\lambda}),$$

and the optimality of \bar{u} then results from Theorem 3.1. ■

The above theorems are powerful tools for partially finite convex programming whenever it is possible to compute the conjugate functions F^* and g_* . In general, the main difficulty is the computation of F^* , since such a computation entails solving infinite dimensional optimization problems. In the context of this paper, the conjugacy is easily performed, thanks to results by Rockafellar [20, 22, 23, 24] on the conjugacy of integral functionals, which we recall in Appendix B.

3.3. Primal-dual relationship. We now use results in Appendix B in order to deal with dual pairs of problems such as those of Theorem 3.1 in which $(U, V) = (L^1_\mu([a, b]), L^\infty_\mu([a, b]))$, F is an integral functional, with integrand $k: \mathbb{R} \times S \rightarrow \bar{\mathbb{R}}$ defined by $F(u) := \int k(u(x), x) d\mu(x)$ and \mathbb{T} is a well defined integral operator acting on L^1 -functions:

$$\mathbb{T}u = \int \gamma(x)u(x) d\mu(x).$$

Here, each component of $\gamma = (\gamma_1, \dots, \gamma_d): [a, b] \rightarrow \mathbb{R}^d$ is assumed to be an L^∞ -function, and an easy calculation shows that, in this case, $\mathbb{T}^*: \mathbb{R}^d \rightarrow L^\infty_\mu([a, b])$ is given by

$$\mathbb{T}^*\lambda(x) = \langle \lambda, \gamma(x) \rangle \quad \forall \lambda \in \mathbb{R}^d.$$

Fenchel duality will work peacefully under the following qualification assumptions:

$$\begin{aligned} (QC) \quad & \text{ri}(\mathbb{T} \text{dom } F) \cap \text{ri dom } g \neq \emptyset \\ \text{and } (QC^*) \quad & \text{ri dom } g_* \cap \text{ri dom } (F^* \circ \mathbb{T}^*) \neq \emptyset. \end{aligned}$$

The effective domain of the dual function $D = g_* - F^* \circ \mathbb{T}^*$ satisfies

$$\text{dom } D = \text{dom } g_* \cap \text{dom } (F^* \circ \mathbb{T}^*).$$

The next theorem will provide us with an explicit relationship allowing for the computation of a primal solution from the knowledge of a dual solution, in the case which we just specified, with suitable assumptions.

Theorem 3.3. *With the notation and assumptions of Theorem 3.1, suppose that (QC) is satisfied, that $\text{dom } D$ has nonempty interior, $F = K$ is the integral functional of integrand k , that conjugacy through the integral sign is permitted, that is,*

$$K(u) = \int k(u(x), x) d\mu(x), \quad K^*(v) = \int k^*(v(x), x) d\mu(x),$$

*and that $K^{**} = K$ and $g_{**} = g$. Suppose in addition that $k^*(\cdot, x)$ is differentiable on \mathbb{R} , with derivative $k^{*\prime}(\cdot, x)$. Let $\bar{\lambda}$ be a dual solution such that $\bar{\lambda} \in \text{int dom } D$. If*

$$\bar{u}(x) := k^{*\prime}(\langle \bar{\lambda}, \gamma(x) \rangle, x) \in U,$$

then \bar{u} is a primal solution.

PROOF. It is readily seen that, since $\text{int dom } D \neq \emptyset$, $\text{int dom } g_*$ and $\text{int dom } (K^* \circ \mathbb{T}^*)$ have a nonempty intersection. The condition (QC*) is therefore trivially satisfied. From the definition of \bar{u} , it is clear that $\bar{u}(x) \in \partial k^*(\langle \mathbb{T}^* \bar{\lambda} \rangle(x), x)$ for every x . Consequently, \bar{u} belongs to $\partial K^*(\mathbb{T}^* \bar{\lambda})$. Moreover,

$$(K^* \circ \mathbb{T}^*)(\lambda) = \int k^*(\mathbb{T}^* \lambda(x), x) d\mu(x) = \int k^*(\langle \lambda, \gamma(x) \rangle, x) d\mu(x).$$

Since the function $\lambda \mapsto k^*(\langle \lambda, \gamma(x) \rangle, x)$ is convex and differentiable, the results in Appendix C show that integration and the differentiation can be interchanged. We therefore have

$$\nabla (K^* \circ \mathbb{T}^*)(\bar{\lambda}) = \int \gamma(x) k^{*\prime}(\langle \bar{\lambda}, \gamma(x) \rangle, x) d\mu(x) = \mathbb{T} \bar{u},$$

and the result then stems from Theorem 3.2. ■

3.4. Application to Problem (\mathcal{P}). We now use the results from the above subsections to produce a dual strategy for the resolution of problem (\mathcal{P}). For any $y \in \mathbb{R}$ and $Y \in \mathbb{R}^{(p+1)^2}$, we define $g(y, Y) := -\delta(y|\{1\}) - \|Y - \bar{A}\|^2 / (2\alpha)$. Here, $y \mapsto \delta(y|\{1\})$ denotes the *indicator function* of the singleton $\{1\}$. Recall that the indicator function of a set $S \subset \mathbb{R}^d$ is defined (on \mathbb{R}^d) by

$$\delta(\mathbf{x}|S) = \begin{cases} 0 & \text{if } \mathbf{x} \in S, \\ \infty & \text{otherwise.} \end{cases}$$

Then problem (\mathcal{P}) can be rewritten as

$$(\mathcal{P}) \quad \begin{cases} \text{Min } K_\mu(u) - g(\mathbb{T}u) \\ \text{s.t. } u \in L_\mu^1([a, b]). \end{cases}$$

It is in the form of the primal problem in Theorem 3.1 with $F = K_\mu$. In order to write the dual problem of (\mathcal{P}), we must compute the dual function $D := g_* - K_\mu^* \circ \mathbb{T}^*$. An easy computation shows that the conjugate of g is given by

$$g_*(z, Z) := \inf \{ zy + \langle Z, Y \rangle - g(y, Y) \mid (y, Y) \in \mathbb{R} \times \mathbb{R}^{(p+1)^2} \} = z + \langle Z, \bar{A} \rangle - \frac{\alpha}{2} \|Z\|^2.$$

Since g is, as can be easily verified, closed proper concave, the biconjugate of g is nothing but g itself. We now turn to the conjugacy of K_μ . We apply the results of Appendix B, with

(S, \mathcal{F}, μ) the (completion of the) measure space $([a, b], \mathcal{B}([a, b]), \mu)$,

K_μ the integral functional with integrand $k(u, s) = k_\circ(u)$.

It is readily verified that both $U = L_\mu^1([a, b])$ and $V = L_\mu^\infty([a, b])$ are decomposable, and that the integrand k_\circ is Borel measurable and closed proper convex. Moreover, the convex conjugate k_\circ^* of k_\circ is given, via an easy computation, by $k_\circ^*(\tau) = \exp(\tau - 1)$. Since k_\circ is closed proper convex, we also have that $k_\circ^{**} = k_\circ$. It then follows from the results in Theorem B.2 that

$$K_\mu^*(v) = \int \exp(v(x) - 1) d\mu(x) \quad \text{for every } v \in L_\mu^\infty([a, b]), \quad \text{and} \quad K_\mu^{**} = K_\mu.$$

The adjoint $\mathbb{T}^*: \mathbb{R}^{1+(p+1)^2} \rightarrow L_\mu^\infty([a, b])$ of \mathbb{T} is given by

$$\mathbb{T}^*(z, Z)(x) = z + \langle Z, f(x)f(x)^\top \rangle = z + \langle f(x), Zf(x) \rangle,$$

so that the dual of Problem (\mathcal{P}) reads:

$$(\mathcal{D}) \quad \left\{ \begin{array}{l} \text{Max } D(z, Z) = z + \langle Z, \bar{A} \rangle - \frac{\alpha}{2} \|Z\|^2 - \exp(z - 1) \int \exp \langle f(x), Zf(x) \rangle d\mu(x) \\ \text{s.t. } (z, Z) \in \mathbb{R}^{1+(p+1)^2}. \end{array} \right.$$

By the strong duality in Theorem 3.1, the dual solution exists. Notice that the effective domain of the (concave) function D is $\mathbb{R}^{1+(p+1)^2}$, so that the dual problem (\mathcal{D}) is actually unconstrained (and finite dimensional). Clearly, D is also continuously differentiable. The first order optimality condition takes the form of the following system:

$$(\Sigma) \quad \left\{ \begin{array}{l} 0 = 1 - \exp(z - 1) \int \exp \langle f(x), Zf(x) \rangle d\mu(x), \\ 0 = \bar{A} - \alpha Z - \exp(z - 1) \int f(x)f(x)^\top \exp \langle f(x), Zf(x) \rangle d\mu(x). \end{array} \right.$$

Solving for $\exp(z - 1)$ in the above system yields the following:

$$0 = \bar{A} - \alpha Z - \left(\int \exp \langle f(x), Zf(x) \rangle d\mu(x) \right)^{-1} \int f(x)f(x)^\top \exp \langle f(x), Zf(x) \rangle d\mu(x).$$

Observe that the latter system is also the optimality system for the optimization problem:

$$(\tilde{\mathcal{D}}) \quad \left\{ \begin{array}{l} \text{Max } \tilde{D}(Z) := \langle Z, \bar{A} \rangle - \frac{\alpha}{2} \|Z\|^2 - \ln \int \exp \langle f(x), Zf(x) \rangle d\mu(x) \\ \text{s.t. } Z \in \mathbb{R}^{(p+1)^2}. \end{array} \right.$$

Proposition 3.1. The above defined function \tilde{D} is concave.

PROOF. Clearly, $\tilde{D}(Z) \in \mathbb{R}$ for every $Z \in \mathbb{R}^{(p+1)^2}$. In order to check that \tilde{D} is concave, it suffices to check the convexity of the function

$$F: Z \mapsto \ln \int \exp \langle Z, f(x)f(x)^\top \rangle d\mu(x).$$

Let $Z_1, Z_2 \in \mathbb{R}^{(p+1)^2}$ and $\alpha \in (0, 1)$. Then, $p := 1/(1 - \alpha)$ and $q := 1/\alpha$ are Hölder conjugate numbers, and Hölder's inequality yields:

$$\begin{aligned} & \int \exp \langle (1 - \alpha)Z_1 + \alpha Z_2, f(x)f(x)^\top \rangle d\mu(x) \\ &= \int (\exp \langle Z_1, f(x)f(x)^\top \rangle)^{1/p} (\exp \langle Z_2, f(x)f(x)^\top \rangle)^{1/q} d\mu(x) \\ &\leq \left(\int \exp \langle Z_1, f(x)f(x)^\top \rangle d\mu(x) \right)^{1/p} \left(\int \exp \langle Z_2, f(x)f(x)^\top \rangle d\mu(x) \right)^{1/q} \\ &= \left(\int \exp \langle Z_1, f(x)f(x)^\top \rangle d\mu(x) \right)^{1-\alpha} \left(\int \exp \langle Z_2, f(x)f(x)^\top \rangle d\mu(x) \right)^\alpha. \end{aligned}$$

It follows that

$$F((1 - \alpha)Z_1 + \alpha Z_2) \leq (1 - \alpha)F(Z_1) + \alpha F(Z_2). \blacksquare$$

Now, since $\text{dom } g = \{1\} \times \mathbb{R}^{(p+1)^2}$, $\text{dom } g_* = \mathbb{R} \times \mathbb{R}^{(p+1)^2}$ and $\mathbb{I} \text{ dom } F = [0, \infty)$ it is clear that (QC) and (QC^*) always hold. A dual solution may be found by maximizing \tilde{D} (and this is an unconstrained smooth concave maximization problem). Finally, $k^*(\tau, x) = \exp(\tau - 1)$ obviously meets the requirements of Theorem 3.3 and, provided we can solve the dual problem, the primal solution satisfies,

$$\begin{aligned} \bar{u}(x) &= k^{*'}(\mathbb{T}^*(\bar{z}, \bar{Z})(x), x) \\ &= \exp(\bar{z} - 1 + \langle \bar{Z}, f(x)f(x)^\top \rangle) \\ &= \left(\int \exp \langle f(x), \bar{Z}f(x) \rangle d\mu(x) \right)^{-1} \exp \langle f(x), \bar{Z}f(x) \rangle. \end{aligned}$$

Remark 3.1. The entropy problem which we have considered throughout this section is a relaxed version of the initial one. We stress here that this relaxation is legitimate since, in essence, letting the relaxation parameter go to zero produces a sequence of solutions which approximate the solution to the unrelaxed problem, when the latter has a solution. Assume \bar{A} has a representing measure which is absolutely continuous with respect to μ , so that the problem

$$(\mathcal{P}_0) \left| \begin{array}{l} \text{Min } K_\mu(u) \\ \text{s.t. } u \in L_\mu^1([a, b]), 1 = \mathbb{I}u, \bar{A} = \mathbb{A}u \end{array} \right.$$

has a (unique) solution. Letting $\bar{u}_\alpha := \bar{u}$ denote the solution to the relaxed problem (\mathcal{P}) obtained above, given any sequence (α_k) converging to 0, any cluster point

$$\bar{u} \in C := \{u \in L_\mu^1([a, b]) \mid 1 = \mathbb{I}u\}$$

of (u_{α_k}) is an optimal solution to Problem (\mathcal{P}_0) . As a matter of fact, using Appendix B, we may easily show that K_μ satisfies $K_\mu^{**} = K_\mu$, so that our functional is actually $\sigma(L_\mu^1([a, b]), L_\mu^\infty([a, b])$ -lower semicontinuous. The existence of a solution to each relaxed problem being clear by the preceding developments, we may then safely apply theorem A.1. \blacksquare

Remark 3.2. Although in theory the reference measure μ can be chosen arbitrarily, in the numerical section we will only demonstrate how we can infer a measure by using the Lebesgue measure. Observe that using the Lebesgue measure rules out atomic measures, which constitute an important class of measures. For example for a p -th order polynomial regression model, it has been shown in [28] that a K -optimal design can be chosen as an atomic measure with $p + 1$ support points. For a polynomial regression model the entropy optimization approach will necessarily produce an approximate measure which is absolutely continuous. ■

4. Chebyshev polynomials. In [28], it was shown that for the p -th order polynomial regression model on the design space $S = [-1, 1]$ as in Example 1.1, the condition number is a smooth function of the moments and hence the K -optimal design can be solved as a standard nonlinear programming problem. In this section we consider the K -optimal design problem with Chebyshev polynomials as basis functions on the design space $S = [-1, 1]$. We show that for the Chebyshev polynomial model, the condition number is no longer a smooth function of the moments. It is interesting to note that for the Chebyshev polynomial model, the optimal condition number is 1 and the analytic solution for the K -optimal design can be obtained. The result will also be compared with the numerical result from the two stage approach in the next section.

Chebyshev polynomials of the first kind are defined as follows: $T_0(x) = 1$, $T_1(x) = x$, and

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n = 1, 2, \dots$$

It is obvious that $T_n(x) = T_n(-x)$ for even n and that $T_n(x) = -T_n(-x)$ for odd n . It is also well-known that

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & \text{if } n \neq m, \\ \pi & \text{if } n = m = 0, \\ \pi/2 & \text{if } n = m \neq 0. \end{cases} \quad (9)$$

Define $f_0(x) = T_0(x)$ and $f_j(x) = \sqrt{2}T_j(x)$, $j \geq 1$. From (9), we have:

$$\int_{-1}^1 \frac{f_i(x)f_j(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & \text{if } i \neq j, \\ \pi & \text{if } i = j. \end{cases} \quad (10)$$

If we fit the parametric regression model

$$y = \sum_{j=0}^p \theta_j f_j(x) + \varepsilon, \quad x \in [-1, 1], \quad (11)$$

the information matrix $A(\xi) = \int_{-1}^1 f(x)f(x)^\top d\xi(x)$ has several nice properties which we discuss now.

Define $y_j = \int_{-1}^1 x^j d\xi(x)$ as the j -th moment of the measure ξ , $j = 0, 1, \dots$. The elements of the information matrix can be written as linear functions of the moments. For example, when $p = 2$,

$$A(\xi) = \begin{bmatrix} y_0 & \sqrt{2}y_1 & \sqrt{2}(2y_2 - y_0) \\ \sqrt{2}y_1 & 2y_2 & 2(2y_3 - y_1) \\ \sqrt{2}(2y_2 - y_0) & 2(2y_3 - y_1) & 2(4y_4 - 4y_2 + y_0) \end{bmatrix} \quad (12)$$

Similarly as in [28, Theorem 2.1], we can show that K -optimal designs for this model may be chosen to be symmetric about 0.

As observed in [28], in Example 1.1, the condition number is a smooth function of the moments. However, in the general case, the condition number $\kappa(A(\xi))$ is not a smooth function of the moments y_j . We now give an example that illustrates this nonsmoothness.

For $p = 2$, consider symmetric designs $\xi(x)$ with $y_2 = y_4$ (such a design can be found for example by taking the support points as $-1, 0, 1$). Since $y_0 = 1$ and $y_1 = y_3 = 0$ for symmetric designs, the information matrix in (12) becomes

$$A(\xi) = \begin{bmatrix} 1 & 0 & \sqrt{2}(2y_2 - 1) \\ 0 & 2y_2 & 0 \\ \sqrt{2}(2y_2 - 1) & 0 & 2 \end{bmatrix}.$$

The three eigenvalues of $A(\xi)$ are

$$\lambda_1 := \frac{3 + \sqrt{8(2y_2 - 1)^2 + 1}}{2}, \quad \lambda_2 := 2y_2, \quad \lambda_3 := \frac{3 - \sqrt{8(2y_2 - 1)^2 + 1}}{2}.$$

The largest eigenvalue λ_{\max} is always λ_1 , and the smallest eigenvalue is

$$\lambda_{\min} = \begin{cases} \lambda_2 & \text{if } y_2 \leq 1/2, \\ \lambda_3 & \text{otherwise.} \end{cases}$$

The condition number is given by

$$\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}} = \begin{cases} \frac{3 + \sqrt{8(2y_2 - 1)^2 + 1}}{4y_2} & \text{if } y_2 \leq 1/2, \\ \frac{3 + \sqrt{8(2y_2 - 1)^2 + 1}}{3 - \sqrt{8(2y_2 - 1)^2 + 1}} & \text{otherwise.} \end{cases}$$

It is clear from the above formula that the derivative of κ with respect to y_2 does not exist at $y_2 = 1/2$.

All in all, we have proved the following result.

Proposition 4.1. *For the Chebyshev polynomial model in Equation (11), the minimum of $\kappa(A)$ is equal to 1, and the K -optimal design $\xi^*(x)$ has density*

$$p^*(x) = \frac{1}{\pi\sqrt{1-x^2}}, \quad x \in [-1, 1].$$

5. Numerical results. We now examine computational issues. Recall that our inference process is divided into two stages:

1. compute the moments of K -optimal designs;
2. from the moments obtained in the first stage, compute a K -optimal design density.

Both stages entail optimization problems of well-known form. In subsection 5.1, the approximate K -optimal design densities for Chebyshev polynomial model (11) are computed for $p = 2, 3$ and 4, and they are compared with the theoretical density in Proposition 4.1. In Subsection 5.2, we give two approximate K -optimal design densities for the model in Example 1.1. The results show that our two stage approach works very well.

In the computation, we select a sequence for α in the following manner: $\alpha_1 = 0.1, \alpha_k = 0.1\alpha_{k-1}$ for $k \geq 2$ and $\delta = 10^{-6}$. Let $\bar{Z}^{(1)}$ be an approximate minimizer for $\alpha = \alpha_1$ up to the default

relative tolerance 10^{-10} of the solver. For $k = 1, 2, \dots$, we use $\bar{Z}^{(k)}$ as the starting value to find $\bar{Z}^{(k+1)}$ for $\alpha = \alpha_{k+1}$, and the iteration stops when

$$\frac{\|\bar{Z}^{(k+1)} - \bar{Z}^{(k)}\|_1}{\|\bar{Z}^{(k)}\|_1} < \delta.$$

5.1. K -optimal density functions for Chebyshev polynomial model. Stage 1 Let $t := \lambda_{\max}(A(\xi))$, $s := \lambda_{\min}(A(\xi))$ and $A[y_0, \dots, y_{2p}] = A(\xi)$, with $y_j = \int_{-1}^1 x^j d\xi(x)$. Let $u_i := y_i/s$, $i = 0, 1, \dots, 2p$ and

$$B[u_0, \dots, u_{2p}] := \frac{A[y_0, \dots, y_{2p}]}{s}.$$

Then the optimization problem under consideration takes the form:

$$\left| \begin{array}{l} \text{Min } v \\ \text{s.t. } I \preceq B \preceq vI, \\ H(u_0, \dots, u_{2p}) \succeq 0, \\ H(u_0, \dots, u_{2p-2}) - H(u_2, \dots, u_{2p}) \succeq 0. \end{array} \right.$$

Using the `SeDuMi` code [26], we can get the scaled moments $\hat{u}_0, \dots, \hat{u}_{2p}$. Since $y_0 = 1$ necessarily, the optimal moments are given by

$$\hat{y}_i = \frac{\hat{u}_i}{\hat{u}_0}, \quad i = 1, 2, \dots, 2p.$$

Since K -optimal designs may be chosen to be symmetric, we can set $y_j = 0 = u_j$ for odd j in the SDP problem so as to reduce the number of variables in the computation.

Stage 2 Suppose $\hat{y}_1, \dots, \hat{y}_{2p}$ are obtained in Stage 1. In the optimization problem $\tilde{\cdot}$, let

$$f_0(x) = T_0(x), f_j(x) = \sqrt{2}T_j(x), j = 1, \dots, p, \quad d\mu(x) = d\lambda(x),$$

and $\bar{A} = A(1, \hat{y}_1, \dots, \hat{y}_{2p})$. Using the routine `nlinb` in the statistical software R to minimize $-\tilde{D}(Z)$, we can find the minimizer \bar{Z} . Then the estimated density function is given by

$$\bar{u}(x) = c \cdot \exp [f^\top(x) \bar{Z} f(x)], \quad x \in [-1, 1],$$

where the constant c is determined to have $\int_{-1}^1 \bar{u}(x) dx = 1$.

Table 1 presents some representative results. As $\alpha \downarrow 0$, the numerical results for \bar{Z} converge. Figure 1 shows the theoretical density function of the K -optimal design and the estimated density functions for $p = 2, 3, 4$. It is clear that the estimated density functions are very close to the theoretical one. The approximation gets better as p goes larger. For $p = 4$, the moments of the estimated density function are: $\hat{y}_1 = 0.000$, $\hat{y}_2 = 0.5000$, $\hat{y}_3 = 0.000$, $\hat{y}_4 = 0.3750$, $\hat{y}_5 = 0.000$, $\hat{y}_6 = 0.3125$, $\hat{y}_7 = 0.000$, $\hat{y}_8 = 0.2734$, which are almost the same values as the estimated moments in Table 1.

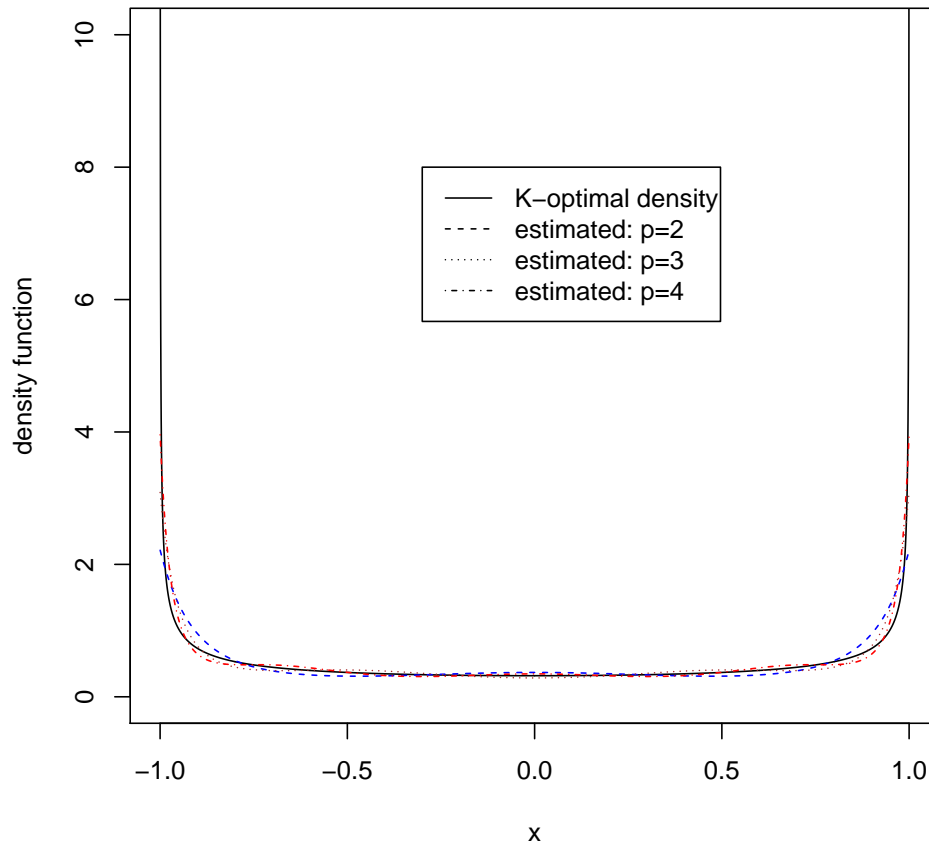


FIGURE 1. The theoretical K -optimal density and the estimated density functions for the Chebyshev polynomial model.

TABLE 1. Results for model (11)

p	Estimated moments (stage 1)	Matrix \bar{Z} (stage 2)
2	$\hat{y}_1 = 0 \quad \hat{y}_2 = 0.5$ $\hat{y}_3 = 0 \quad \hat{y}_4 = 3/8$	$\begin{bmatrix} 0.3261 & 0.0000 & 0.4045 \\ 0.0000 & -0.246 & 0.0000 \\ 0.4045 & 0.0000 & 0.4090 \end{bmatrix}$
3	$\hat{y}_1 = 0 \quad \hat{y}_2 = 0.5$ $\hat{y}_3 = 0 \quad \hat{y}_4 = 3/8$ $\hat{y}_5 = 0 \quad \hat{y}_6 = 5/16$	$\begin{bmatrix} 0.2923 & 0.0000 & -0.1754 & 0.0000 \\ 0.0000 & 0.5951 & 0.0000 & 0.4125 \\ -0.1754 & 0.0000 & -0.3942 & 0.0000 \\ 0.0000 & 0.4125 & 0.0000 & 0.2689 \end{bmatrix}$
4	$\hat{y}_1 = 0 \quad \hat{y}_2 = 0.5$ $\hat{y}_3 = 0 \quad \hat{y}_4 = 3/8$ $\hat{y}_5 = 0 \quad \hat{y}_6 = 5/16$ $\hat{y}_7 = 0 \quad \hat{y}_8 = 35/128$	$\begin{bmatrix} 1.7327 & 0.0000 & 1.0270 & 0.0000 & 0.6173 \\ 0.0000 & -1.1505 & 0.0000 & -0.7680 & 0.0000 \\ 1.0270 & 0.0000 & 0.2337 & 0.0000 & 0.3605 \\ 0.0000 & -0.7680 & 0.0001 & -0.4401 & 0.0000 \\ 0.6173 & 0.0000 & 0.3605 & 0.0000 & 0.2001 \end{bmatrix}$

5.2. K-optimal density functions for Example 1.1. Using the two stage approach, we compute the K -optimal design densities for the model in Example 1.1.

Stage 1 Let $t := \lambda_{\max}(A(\xi))$, $s := \lambda_{\min}(A(\xi))$ and $A(\xi) = H(y_0, \dots, y_{2p})$, with $y_j = \int_{-1}^1 x^j d\xi(x)$. Let $u_i := y_i/s$, $i = 0, 1, \dots, 2p$. Since the information matrix is a Hankel matrix $H(y_0, \dots, y_{2p})$, the optimization problem under consideration takes the form:

$$\begin{cases} \text{Min } v \\ \text{s.t. } I \preceq H(u_0, \dots, u_{2p}) \preceq vI, \\ H(u_0, \dots, u_{2p-2}) - H(u_2, \dots, u_{2p}) \succeq 0. \end{cases}$$

The rest is similar to that in Section 5.1 to get $\hat{y}_1, \dots, \hat{y}_{2p}$.

Stage 2 In the optimization problem $\tilde{\cdot}$, let

$$Z = (Z_{ij}) \in \mathbb{R}^{(p+1) \times (p+1)}, \quad f^\top(x) = (1, x, \dots, x^p), \quad d\mu(x) = d\lambda(x),$$

and $\bar{A} = H(1, \hat{y}_1, \dots, \hat{y}_{2p})$. Similar to Section 5.1, we can get the estimated density function $\bar{u}(x)$.

Some representative results are reported here for $p = 3$ and $p = 4$. In Stage 1, we get

$$\hat{y}_2 = 0.3626, \hat{y}_4 = 0.2287, \hat{y}_6 = 0.2006$$

for $p = 3$, and

$$\hat{y}_2 = 0.3257, \hat{y}_4 = 0.2072, \hat{y}_6 = 0.1552, \hat{y}_8 = 0.1324$$

for $p = 4$ (and moments with odd orders are zero). In Stage 2, the approximate densities are computed and plotted in Figure 2, where the dots represent the $p + 1$ support points of the discrete K -optimal designs obtained in [28]. It is clear that the density functions approximate the discrete K -optimal designs well.

6. Concluding Remarks. In this paper, we construct K -optimal designs by finding the probability distribution of the design variable. The distribution can be discrete or continuous, supported on the design space $S = [a, b]$, a compact interval. Solving the SDP problem in the first step gives the K -optimal moments, and the second step finds the probability density that approximates the true K -optimal design. Note that if the design space S is discretized so that S includes a set of finitely many points, there are various effective algorithms investigated in the literature to compute optimal weights for approximate A -optimal, D -optimal and E -optimal designs. They include the multiplicative algorithm, for example in [18, 7, 9, 27, 29], the interior point method in [14], the SDP method in [17], and the simulated annealing method in [31]. Some of their algorithms can be applied to compute optimal weights for approximate K -optimal designs as well. However, this paper does not focus on discrete design spaces.

There are other optimal design criteria studied in the literature, and they include A -optimal, D -optimal and E -optimal design criteria: for example, see [19, 8]. The two step inference strategy can be easily applied to find A -optimal and E -optimal designs. In [28], A -optimal and D -optimal and K -optimal designs are compared for polynomial regression models. They are similar to each other in terms of the number and the symmetry of support points. However the shape of the distribution of K -optimal designs is more similar to that of A -optimal designs than D -optimal designs.

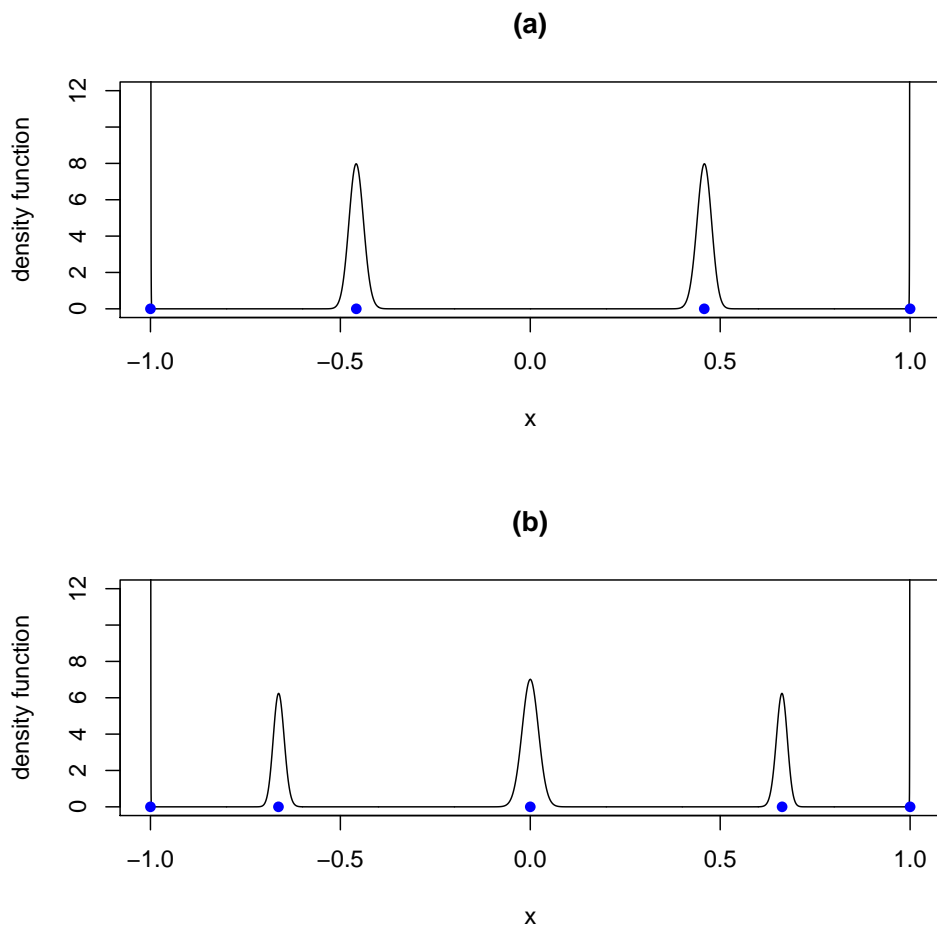


FIGURE 2. K -optimal density functions: (a) $p=3$, (b) $p=4$. The dots represent the $p+1$ support points of the discrete K -optimal designs obtained in [28].

Acknowledgements The first author was supported by UMI 3069 PIMS EUROPE and the CNRS. The second and third authors were partially supported by NSERC. All three authors wish to thank Jean-Bernard Lasserre and Didier Henrion for useful comments.

Appendix A: Exterior penalty function method This appendix gives a possibly infinite dimensional version of results which are stated and proved in [1] (Section 9.2) in the finite dimensional case.

Let U be any topological vector space, let C be a subset of U , let X be any norm space, let $f: U \rightarrow \bar{\mathbb{R}}$ be lower semicontinuous, and let $h: U \rightarrow X$ be continuous.

Consider the constrained optimization problem

$$(P) \quad \left| \begin{array}{l} \text{Min } f(u) \\ \text{s.t. } u \in C, h(u) = 0, \end{array} \right.$$

as well as the corresponding penalty problem

$$(P_\alpha) \quad \begin{cases} \text{Min } f(u) + \alpha^{-1} \|h(u)\|^2 \\ \text{s.t. } u \in C. \end{cases}$$

Let $\theta(\alpha) = \mathcal{V}(P_\alpha)$ be the optimal value of (P_α) , that is,

$$\theta(\alpha) := \inf\{f(u) + \alpha^{-1} \|h(u)\|^2 \mid u \in C\}.$$

Lemma A.1. *Suppose that, for each $\alpha > 0$, (P_α) has a solution u_α (i.e. $u_\alpha \in C$ and $f(u_\alpha) + \alpha^{-1} \|h(u_\alpha)\|^2 = \theta(\alpha)$). Then the following statements hold:*

(1) *For every u that is feasible for (P) and every $\alpha > 0$, $f(u) \geq \theta(\alpha)$. Consequently,*

$$\mathcal{V}(P) := \inf\{f(u) \mid u \in C, h(u) = 0\} \geq \sup\{\theta(\alpha) \mid \alpha > 0\}.$$

(2) *The function $(\alpha \mapsto \|h(u_\alpha)\|)$ is wide-sense increasing, and both $(\alpha \mapsto f(u_\alpha))$ and $(\alpha \mapsto \theta(\alpha))$ are wide-sense decreasing.*

PROOF. Let u be feasible for (P) and let $\alpha > 0$. Then

$$f(u) = f(u) + \alpha^{-1} \|h(u)\|^2 \geq \inf\{f(v) + \alpha^{-1} \|h(v)\|^2 \mid v \in C\} = \theta(\alpha).$$

Statement (1) follows. To show Statement (2), let $\beta > \alpha > 0$. Since u_α and u_β are solutions to (P_α) and (P_β) , respectively, we have the following two inequalities:

$$f(u_\alpha) + \beta^{-1} \|h(u_\alpha)\|^2 \geq f(u_\beta) + \beta^{-1} \|h(u_\beta)\|^2, \quad (13)$$

$$f(u_\beta) + \alpha^{-1} \|h(u_\beta)\|^2 \geq f(u_\alpha) + \alpha^{-1} \|h(u_\alpha)\|^2, \quad (14)$$

from which we deduce that

$$(\beta^{-1} - \alpha^{-1})(\|h(u_\alpha)\|^2 - \|h(u_\beta)\|^2) \geq 0.$$

Since $\beta > \alpha$, we obtain $\|h(u_\alpha)\| \leq \|h(u_\beta)\|$. Now, combining the latter inequality with (13) yields $f(u_\alpha) \geq f(u_\beta)$. Finally, adding and subtracting $\alpha^{-1} \|h(u_\alpha)\|^2$ to the left hand side of (13) gives rise to the

$$\theta(\alpha) + (\beta^{-1} - \alpha^{-1}) \|h(u_\alpha)\|^2 \geq \theta(\beta).$$

The inequality $\theta(\alpha) \geq \theta(\beta)$ follows. ■

Theorem A.1. *Suppose that, for each $\alpha > 0$, (P_α) has a solution u_α . Given any sequence (α_k) converging to 0, suppose that $\bar{u} \in C$ is a cluster point of (u_{α_k}) . Then \bar{u} is an optimal solution to Problem (P) and*

$$f(\bar{u}) = \mathcal{V}(P) = \inf\{f(u) \mid u \in C, h(u) = 0\} = \sup\{\theta(\alpha) \mid \alpha > 0\} = \lim_{\alpha \downarrow 0} \theta(\alpha).$$

Moreover $\alpha_k^{-1} \|h(u_{\alpha_k})\|^2$ goes to 0 as $k \rightarrow \infty$.

PROOF. By Lemma A.1(2), $\theta(\alpha)$ is wide-sense decreasing, so that

$$\sup\{\theta(\alpha)|\alpha > 0\} = \lim_{\alpha \downarrow 0} \theta(\alpha).$$

We first show that $\|h(u_\alpha)\|$ goes to zero as $\alpha \downarrow 0$. Fix $\varepsilon > 0$. We shall prove that, for α sufficiently small, $\|h(u_\alpha)\| \leq \sqrt{\varepsilon}$. Let u be feasible for (P) and let u_1 be an optimal solution to (P_1) . Assume $\alpha^{-1} \geq \varepsilon^{-1}|f(u) - f(u_1)| + 2$. By Lemma A.1(2), we must have $f(u_\alpha) \geq f(u_1)$. Suppose, in order to obtain a contradiction, that $\|h(u_\alpha)\|^2 > \varepsilon$. By Lemma A.1(1), we have:

$$\begin{aligned} \inf\{f(v)|v \in C, h(v) = 0\} &\geq \theta(\alpha) = f(u_\alpha) + \alpha^{-1}\|h(u_\alpha)\|^2 \\ &\geq f(u_1) + \alpha^{-1}\|h(u_\alpha)\|^2 \\ &> f(u_1) + \alpha^{-1}\varepsilon \\ &\geq f(u_1) + |f(u) - f(u_1)| + 2\varepsilon \\ &> f(u), \end{aligned}$$

in contradiction with the fact that u is feasible for (P) . Now suppose that $u_{\alpha_k} \rightarrow \bar{u}$ as $k \rightarrow \infty$. Then for all k ,

$$\sup\{\theta(\alpha)|\alpha > 0\} \geq \theta(\alpha_k) = f(u_{\alpha_k}) + \frac{1}{\alpha_k}\|h(u_{\alpha_k})\|^2 \geq f(u_{\alpha_k}).$$

Since f is lower semicontinuous, we deduce that

$$\sup\{\theta(\alpha)|\alpha > 0\} \geq \liminf_{k \rightarrow \infty} f(u_{\alpha_k}) \geq f(\bar{u}). \quad (15)$$

Since $\|h(u_{\alpha_k})\|^2 \rightarrow 0$ as $k \rightarrow \infty$ by the first part of the proof, the continuity of h shows that we must have $h(\bar{u}) = 0$. Thus, \bar{u} is feasible for (P) . By (15) and Lemma A.1(1), $f(\bar{u}) \leq \mathcal{V}(P)$, so equality holds in fact. Thus \bar{u} is an optimal solution to (P) and

$$\sup\{\theta(\alpha)|\alpha > 0\} = f(\bar{u}).$$

Finally, $\alpha^{-1}\|h(u_\alpha)\|^2 = \theta(\alpha) - f(u_\alpha)$, so that

$$\begin{aligned} 0 &\leq \limsup_{k \rightarrow \infty} \alpha^{-1}\|h(u_{\alpha_k})\|^2 \\ &= \limsup_{k \rightarrow \infty} (\theta(\alpha_k) - f(u_k)) \\ &= \lim_{k \rightarrow \infty} \theta(\alpha_k) - \liminf_{k \rightarrow \infty} f(u_k) \\ &\leq \mathcal{V}(P) - f(\bar{u}) \\ &= 0, \end{aligned}$$

in which the last inequality stems from the lower semicontinuity of f . The result follows. ■

The penalty method in Theorem A.1 is only applicable if the iteration sequence (u_{α_k}) has a cluster point. This can be guaranteed if for example C is compact. In practice if an iteration sequence does not have a cluster point, we change (α_k) to another sequence and try to obtain a sequence (u_{α_k}) that has a cluster point.

Corollary A.1. *If $h(u_\alpha) = 0$ for some α , then u_α is an optimal solution to (P) .*

PROOF. If $h(u_\alpha) = 0$ for some α , then u_α is a feasible solution to (P) . Furthermore since

$$\inf\{f(u) : h(u) = 0, u \in C\} \geq \theta(\alpha) = f(u_\alpha) + \alpha^{-1}\|h(u_\alpha)\|^2 = f(u_\alpha)$$

it immediately follows that u_α is an optimal solution of (P) .

Appendix B: Conjugacy through the integral sign Let (S, \mathcal{T}, μ) be a measure space, where μ is nonnegative and σ -finite. Let U and V be two spaces of measurable functions on S such that, for every $u \in U$ and every $v \in V$, the product uv is integrable. We can then define on $U \times V$ the bilinear form

$$(u, v) \longmapsto \langle u, v \rangle := \int u(s)v(s) \, d\mu(s). \quad (16)$$

A function $k: \mathbb{R} \times S \rightarrow \bar{\mathbb{R}}$ is called an integrand. Its *epigraph multifunction* is defined on S by

$$E_k(s) := \text{epi } k(\cdot, s) = \{(u, \alpha) \in \mathbb{R} \times \mathbb{R} \mid \alpha \geq k(u, s)\}.$$

The integrand k is said to be:

1. *lower semicontinuous*, if $k(\cdot, s)$ is lower semicontinuous for all s ;
2. *normal*, if k is lower semicontinuous and E_k is a measurable multifunction;
3. *proper*, if $k(\cdot, s)$ is proper for every s (a function is proper if it doesn't take the value $-\infty$ and is not identically equal to ∞);
4. *convex*, if $k(\cdot, s)$ is convex for every s .

For every measurable function u from (S, \mathcal{T}) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (where $\mathcal{B}(\mathbb{R})$ denote the Borel σ -algebra on \mathbb{R}), the function $s \mapsto k(u(s), s)$ is then measurable from (S, \mathcal{T}) to $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$. See [24], Corollary 2B.

Furthermore, one can define on U the *integral functional*

$$u \longmapsto K(u) := \int k(u(s), s) \, d\mu(s), \quad (17)$$

where k is a normal integrand. The above integral is well-defined with the convention $\infty - \infty = \infty$. If $k(\cdot, s)$ is convex for every $s \in S$, then K is also convex. See [24], Section 3.

Kullback-Leibler's relative entropy [11] is a standard example of proper convex integral functional.

For every measurable set T , let $\mathbb{1}_T$ denote the characteristic function of T . A space U of measurable functions is said to be *decomposable* if it contains all functions of the form

$$u_0 \mathbb{1}_T + u \mathbb{1}_{T^c}$$

with u_0 measurable, $\mu(T)$ finite, $u_0(T)$ bounded and $u \in U$. Here, T^c denotes the complement of T .

Theorem B.1 ([24], Proposition 2S). *If k is a normal integrand, then so are its conjugate and biconjugate integrands.*

For every measurable function $v \in V$, the integral

$$\int k^*(v(s), s) \, d\mu(s)$$

is well-defined (with the convention $\infty - \infty = \infty$).

Theorem B.2 ([24], Theorem 3C). *Let k be a normal integrand and let K be the corresponding integral functional. Let U and V be spaces of measurable functions on which one can define the bilinear form (16). If U is decomposable and if the functional K defined in (17) has a nonempty domain, then*

$$K^*(v) = \int k^*(v(s), s) \, d\mu(s)$$

for every $v \in V$. Moreover, the functional K^* is convex. If in turn V is decomposable and K^* has nonempty domain, then, for every $u \in U$,

$$K^{**}(u) = \int k^{**}(u(s), s) d\mu(s).$$

Appendix C: Differentiation through the integral sign in the convex case Here, we state and prove a few results on the differentiation under the integral sign in the case of convex integrands. These results, which can be found in [15], are given here for the sake of completeness.

Lemma C.1. Let I be an open interval in \mathbb{R} and let $f: I \times \mathbb{R} \rightarrow \mathbb{R}$ be a mapping. Assume that

(a) for μ -almost every s , $f(\cdot, s)$ is convex and differentiable on I , with derivative denoted by $f'(\lambda, s)$;

(b) for every $\lambda \in I$, $f(\lambda, \cdot)$ is integrable.

Then $f'(\lambda, \cdot)$ is integrable for μ -almost every s . Moreover, the function F defined on I by $F(\lambda) = \int f(\lambda, s) d\mu(s)$ is (convex and) differentiable on I , and one has

$$\forall \lambda \in I, \quad F'(\lambda) = \int f'(\lambda, s) d\mu(s). \quad (18)$$

Let $\lambda \in I$ and let $\lambda_1, \lambda_2, \dots$ be a sequence in I which converges to λ . then there exists λ^- and λ^+ in I such that λ and the λ_k 's belong to the open interval (λ^-, λ^+) . Let us define

$$h^-(s) := \frac{f(\lambda^-, s) - f(\lambda, s)}{\lambda^- - \lambda}, \quad h^+(s) := \frac{f(\lambda^+, s) - f(\lambda, s)}{\lambda^+ - \lambda},$$

$$\text{and } h_k(s) := \frac{f(\lambda_k, s) - f(\lambda, s)}{\lambda_k - \lambda}.$$

It is clear that $h_k \rightarrow h := f'(\lambda, \cdot)$ μ -almost everywhere. By convexity, one also has, for every k , $h^- \leq h_k \leq h^+$ μ -almost everywhere. Therefore, $|h_k| \leq \max\{|h^-|, |h^+|\}$ μ -almost everywhere, and the right hand side is integrable since h^- et h^+ are integrable. The Dominated Convergence Theorem then yields

$$\int h(s) d\mu(s) = \lim_{k \rightarrow \infty} \int h_k(s) d\mu(s) = \lim_{k \rightarrow \infty} \frac{F(\lambda_k) - F(\lambda)}{\lambda_k - \lambda}.$$

Since the sequence $\lambda_1, \lambda_2, \dots$ was arbitrary, we have shown that F is differentiable, and that its derivative is given by (18). ■

Theorem C.1. Let Ω be an open ball in \mathbb{R}^d , and let $f: \Omega \times S \rightarrow \mathbb{R}$. Assume that

(a) $f(\cdot, s)$ is convex and differentiable (for μ -almost every s) at every $\lambda \in \Omega$;

(b) $f(\lambda, \cdot)$ is integrable for every $\lambda \in \Omega$.

Then the partial derivatives of $f(\cdot, s)$ are integrable for μ -almost every s . Moreover, the function defined on Ω by $F(\lambda) = \int f(\lambda, s) d\mu(s)$ is (convex and) differentiable on Ω , and one has

$$\nabla F(\lambda) = \int \nabla f(\lambda, s) d\mu(s). \quad (19)$$

L et $\lambda = (\lambda_1, \dots, \lambda_d) \in \Omega$. Denote $f'_1(\lambda, s), \dots, f'_d(\lambda, s)$ the partial derivatives of $f(\cdot, s)$ at λ . From Lemma C.1, F admits at λ partial derivatives F'_j such that

$$F'_j(\lambda) = \int f'_j(\lambda, s) d\mu(s), \quad j = 1, \dots, d.$$

Since F is convex, Theorem 25.2 in [21] shows that F is differentiable, and that the gradient of F is then given by (19). ■

Acknowledgments. The first author was supported by UMI 3069 PIMS EUROPE and the CNRS. The second and third authors were partially supported by NSERC. All three authors wish to thank Jean-Bernard Lasserre and Didier Henrion for useful comments.

References

- [1] M.S. BAZARAA, H.D. SHERALI and C.M. SHETTY, *Nonlinear Programming, Theory and Algorithms*, second edition, Wiley, 1993.
- [2] J.M. BORWEIN and A.S. LEWIS, *Duality relationships for entropy-like minimization problems*, SIAM Journal on Control and Optimization, **29**, pp. 325-338, 1991.
- [3] J.M. BORWEIN and A.S. LEWIS, *On the convergence of moment problems*, Transactions of the American Mathematical Society, **325**, pp. 249-271, 1991.
- [4] J.M. BORWEIN and A.S. LEWIS, *Partially finite convex programming, Part I: quasi relative interiors and duality theory*, Mathematical Programming, **57**, pp. 15-48, 1992.
- [5] K. CHALONER AND I. VERDINELLI, *Bayesian experimental design: a review*, Statistical Science, **10**, pp. 273-304, 1995.
- [6] X. CHEN, R.S. WOMERSLEY and J.J. YE, *Minimizing the condition number of a gram matrix*, SIAM Journal on Optimization, **21**, pp. 127-148, 2011.
- [7] H. DETTE, A. PEPELYSHEV and A. ZHIGLJAVSKY, *Improving updating rules in multiplicative algorithms for computing D-optimal designs*, Computational Statistics and Data Analysis, **53**, pp. 312-320, 2008.
- [8] V. V. FEDOROV and J. LEE, *Design of experiments in statistics*, In H. Wolkowicz, R. Saigal and L. Vandenbergh (Eds.), Handbook of Semidefinite Programming: Theory, Algorithms and Applications. Kluwer Academic Publishers, 2000.
- [9] R. HARMAN and M. TRNOVSKÁ, *Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices*, Mathematica Slovaca, **59**, pp. 693-704, 2009.
- [10] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms, I and II*, Springer-Verlag, Berlin, 1993.
- [11] S. KULLBACK and R.A. LEIBLER, *On information and sufficiency*, Annals of Mathematical Statistics, **22**, pp. 79-86, 1951.
- [12] M. LAURENT, Updated version of *Sums of squares, Moment Matrices and Optimization over Polynomials*, <http://homepages.cwi.nl/~monique/files/moment-ima-update-new.pdf>, (February 6, 2010).
- [13] Z. LU and T.K. PONG, *Minimizing condition number via convex programming*, SIAM Journal on Matrix Analysis and Applications, **32**, pp. 1193-1211, 2011.
- [14] Z. LU and T.K. PONG, *Computing optimal experimental designs via interior point method*, To appear in SIAM Journal on Analysis and Applications.
- [15] P. MARÉCHAL, *Sur la régularisation des problèmes mal posés*, PhD. thesis, Université Paul Sabatier (Toulouse III), 1997.

-
- [16] P. MARÉCHAL and J.J. YE, *Optimizing condition numbers*, SIAM Journal on Optimization, **20**, pp. 935-947, 2009.
- [17] D. PAPP, *Optimal designs for rational function regression*, Journal of the American Statistical Association, **107**, pp. 400-411, 2012.
- [18] A. PÁZMAN, *Foundations of Optimum Experimental Design*, Reidel, Dordrecht, 1986.
- [19] F. PUKELSHEIM, *Optimal Design of Experiments*, John Wiley and Sons Inc., New York, 1993.
- [20] R.T. ROCKAFELLAR, *Integrals which are convex functionals*, Pacific Journal of Mathematics, **24**, pp. 525-539, 1968.
- [21] R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, N. J., 1970.
- [22] R.T. ROCKAFELLAR, *Convex integral functionals and duality*, In E. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, Academic Press, pp. 215-236, 1971.
- [23] R.T. ROCKAFELLAR, *Integrals which are convex functionals, II*, Pacific Journal of Mathematics, **39**, pp. 439-469, 1971.
- [24] R.T. ROCKAFELLAR, *Integral functionals, normal integrands and measurable selections*, in Lucien Waelbroeck, editor, *Nonlinear Operators and the Calculus of Variations*, Springer-Verlag series Lecture Notes in Math., **543**, pp. 157-207, 1976.
- [25] R.T. ROCKAFELLAR AND R. J-B WETS, *Variational Analysis*, Springer, 2nd-printing, 2004.
- [26] SeDuMi: <http://sedumi.ie.lehigh.edu/>
- [27] B. TORSNEY and R. MARTÍN-MARTÍN, *Multiplicative algorithms for computing optimum designs*, Journal of Statistical Planning and Inference, **139**, pp. 3947-3961, 2009.
- [28] J.J. YE and J. ZHOU, *Minimizing the condition number to construct design points for polynomial regression models*, SIAM Journal on Optimization, **23**, pp. 666-686, 2013.
- [29] Y. YU, *Monotonic convergence of a general algorithm for computing optimal designs*, Annals of Statistics, **38**, pp. 1593-1606, 2010.
- [30] C. ZALINESCU, *Convex Analysis in General Vector Spaces*, World Scientific, 2001.
- [31] J. ZHOU, *D-optimal minimax regression designs on discrete design space*, Journal of Statistical Planning and Inference, **138**, pp. 4081-4092, 2008.