OPTIMISATION IN SPACE OF MEASURES AND OPTIMAL DESIGN

Ilya Molchanov[1] and Sergei Zuyev[2]

Abstract. The paper develops an approach to optimal design problems based on application of abstract optimisation principles in the space of measures. Various design criteria and constraints, such as bounded density, fixed barycentre, fixed variance, etc. are treated in a unified manner providing a universal variant of the Kiefer-Wolfowitz theorem and giving a full spectrum of optimality criteria for particular cases. Incorporating the optimal design problems into conventional optimisation framework makes it possible to use the whole arsenal of descent algorithms from the general optimisation literature for finding optimal designs. The corresponding steepest descent involves adding a signed measure at every step and converges faster than the conventional sequential algorithms used to construct optimal designs. We study a new class of design problems when the observation points are distributed according to a Poisson point process arising in the situation when the total control on the placement of measurements is impossible.

1. Introduction

The basic problem in the theory of linear optimal design concerns the best choice of design (or observation) points in the following regression model:

$$y_i = \sum_{j=1}^{k} \beta_j f_j(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $x_1, \ldots, x_n$ belong to a *design space* X (assumed to be a locally compact separable metric space), $\beta = (\beta_1, \ldots, \beta_k)$ is a vector of unknown parameters, $f(x) = (f_1(x), \ldots, f_k(x))$ is a row of linearly independent functions on X and $\varepsilon_1, \ldots, \varepsilon_n$ are independent identically distributed random variables with $\mathbf{E}\,\varepsilon_i = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$. The theory of optimal experimental design addresses the problem of choosing the observation points in order to achieve better properties of the least-squares estimator $\widehat{\beta}$.

[1] Department of Mathematical Statistics and Actuarial Sciences, University of Berne, 3012 Berne, Switzerland;
e-mail: `ilya@stat.unibe.ch`

[2] Department of Statistics and Modelling Science, University of Strathclyde, Glasgow G1 1XH, UK;
e-mail: `sergei@stams.strath.ac.uk`

Since "exact" finite optimal designs are mostly difficult to compute, it is usual to describe a design in terms of a measure μ, so that $\mu(\mathrm{d}x)$ describes the frequency of taking x as an observation point. As a criterion for optimality one may take, for instance, the so-called *generalized variance*, that is the determinant of the covariance matrix $\|\mathbf{cov}(\widehat{\beta}_i, \widehat{\beta}_j)\|$. Then it can be shown that the latter covariance matrix equals $\sigma^2 M(\mu)^{-1}$, where

$$M(\mu) = \int f(x)^\top f(x)\, \mu(\mathrm{d}x) \tag{1.2}$$

is the so-called *information matrix* (see, *e.g.* [1] for details). The measure μ that minimizes $\det M^{-1}(\mu)$ over all measures μ with a given total mass (usually over all probability measures) is called *D-optimal design measure*. To create a *convex* optimisation criterion it is convenient to minimise $\Psi(M) = -\log(\det M)$, which is a convex functional on the space of matrices. Then any local minimum that is found will necessarily be global. In general, a measure μ that provides a global minimum to a given differentiable functional $\Psi(M(\mu))$ is called a Ψ-*optimal design measure*.

The above optimisation problem is usually solved by first finding an optimal information matrix from the class defined by all admissible design measures and afterwards by determining the corresponding design measure. In other words, such methods of finding the optimal design break into the following two stages:

$$\Psi(M(\mu)) \quad \overset{\text{Step I}}{\longrightarrow} \quad M(\mu) \quad \overset{\text{Step II}}{\longrightarrow} \quad \mu\,. \tag{1.3}$$

Step I concerns optimisation in the Euclidean space of dimension $k(k+1)/2$, while Step II aims to identify the measure by the obtained in the previous step information matrix.

However, the chain approach based on (1.3) cannot be easily adopted when the formulation of the problem changes, *e.g.*, when new constraints on μ appear. This usually calls for major changes in proofs, since a new family of matrices has to be analysed at Step I of (1.3). Therefore, for each new type of constraint on μ *both* steps in (1.3) have to be reworked, for example, as in [6] and [11], where various types of constraint are analysed. Specific issues concerning optimal designs on general (not necessarily symmetric) experimental domains are considered in [19].

Although the family of non-negative measures is not a linear space, many objective functionals can be extended onto the linear space of *signed* measures as the definition (1.2) of the information matrix $M(\mu)$ applies literally to signed measures. This allows us to treat the design problems described above as a special case of the optimisation of functionals formally defined on signed measures but the solution constrained on the cone of positive measures. This idea goes back to Fedorov and Hackl [11] who did not however explore the full scope of the related abstract constraint optimisation problems. We show that the abstract setting incorporates naturally many different optimal design problems and leads to new (apparently more efficient) steepest descent algorithms for numerical computation of optimal designs. It is possible to specify the necessary conditions for an optimal design measure for an arbitrary differentiable goal functional Ψ. If the latter is convex, then such necessary conditions immediately become necessary and sufficient, as typically happens in the classical D-optimality theory.

Furthermore, the suggested approach allows us to incorporate easily into the model various constraints on the optimal design measure such as bounded density, fixed barycentre, fixed variance, etc. In a sense, the general result given in Corollary 3.2 below provides a universal variant of the Kiefer-Wolfowitz theorem [16, 17] that yields a full spectrum of particular optimality criteria.

It is worth also mentioning that the general optimality conditions of Theorem 2.1 are applicable to *any* differentiable goal function rather than a function of the information matrix which is typical for a finite dimensional space of response functions as in (1.1). For instance, infinite dimensional response spaces as in [25] can also be treated in the same framework although a particular form of the optimality conditions will, of course, be different from these given in Theorem 3.2. This is, however, out of scope of the present publication.

In this paper we consider only *differentiable* functionals and constraints, see [27] for a comprehensive discussion of non-differentiable cases. A natural approach to solving optimisation problems of that kind relies on differentiation in linear spaces. Recall that the directional (or *Gateaux*) derivative of a real-valued functional f

in a Banach space $\mathbb{E}$ is defined by

$$Df(x)[v] = \lim_{t \downarrow 0} t^{-1}(f(x + tv) - f(x)), \tag{1.4}$$

where $v \in \mathbb{E}$ is a direction (see, *e.g.* [15]). The functional f is said to have a *Fréchet* (or strong) derivative at $x \in \mathbb{E}$ if

$$f(x + v) = f(x) + Df(x)[v] + o(\|v\|) \quad \text{as } \|v\| \to 0,$$

where $D(f(x))[v]$ is a bounded linear continuous functional of v. In the design context f is a functional on the space of measures, such as $\Psi(M(\mu))$, and $\|\cdot\|$ is a norm defined for signed measures, *e.g.*, the total variation norm. If its Fréchet derivative exists, it is then written as $D\Psi(\mu)[\eta]$.

Clearly, if μ is a probability measure, then $\mu + t\eta$ as in (1.4) above is not necessarily a probability measure or even a positive measure. Therefore, $\Psi(\mu + t\eta)$ may loose a direct interpretation within the design framework for an arbitrary signed measure η. To circumvent this difficulty, it is quite typical in the optimal design literature to replace (1.4) by the following definition of the directional derivative:

$$\tilde{D}\Psi(\mu)[\eta] = \lim_{t \downarrow 0} t^{-1}(\Psi((1 - t)\mu + t\eta) - \Psi(\mu)) \tag{1.5}$$

(see, *e.g.* [1, 24, 29, 37]). Now $(1 - t)\mu + t\eta$ is again a probability measure for $t \in [0, 1]$ if both μ and η are probability measures. This definition of the directional derivative is used to construct the steepest (with respect to differential operator $\tilde{D}$) descent algorithms for finding optimal designs when it is not possible to obtain analytical solutions, see [37]. However, steepest descent related to (1.5) differs from the true steepest descent as defined by D in (1.4). Indeed, it is easy to see that

$$\tilde{D}\Psi(\mu)[\eta] = D\Psi(\mu)[\eta - \mu].$$

Thus, the descent direction $\tilde{\eta}^*$ determined by minimizing $\tilde{D}\Psi(\mu)[\eta]$ over all η with the given norm is not the *true* steepest descent direction η^* obtained by minimizing $D\Psi(\mu)[\eta]$. This limitation was recognized in [31] as unavoidable in view of the need to deal with the positiveness constraint on μ. Additional work is thus needed to show that the descent in direction $\tilde{\eta}^*$ indeed converges to a minimum of the goal function. Convexity of the goal function is usually assumed for that.

The true steepest descent method described in Section 5 emerges from our theoretical results on constrained optimisation in the space of measures presented in Section 2. In contrast to the classical sequential algorithms in the optimal design literature (see, *e.g.* [33–36]), that are widely used in the existing optimal design software, we do not renormalise the obtained design measure on each step. Instead, the algorithm adds a *signed* measure chosen so to minimise the Fréchet derivative of the goal function and to maintain the constraints, in particular, the positivity and fixed total mass 1 of the measure. This extends the ideas of Atwood [2, 3], Fedorov [9], Ford [12], pp. 59–64) and Gaffke and Mathar [14] by incorporating the whole algorithm into the conventional optimisation framework. Working on the linear space of signed measures rather than on the cone of positive measures makes this algorithm just a special case of the general steepest descent algorithm known from the optimisation literature. To establish its required convergence properties it is now sufficient to refer to the general results for the steepest descent method. Theorem 3.2 provides an easily verifiable stopping rule for the descent algorithm for various classes of design problems with multiple constraints.

The proposed descent method has a direct statistical bearing by reducing the calculation time and increasing the precision of the solution, since at the final stage it reports correctly the support points of the optimal design measure. Various constrained design problems can be naturally handled and so do not require individually tailored treatments if the nature of constraints changes.

The structure of the paper is the following. In Section 2 we outline several concepts of abstract optimisation for functionals defined on the cone of positive measures. Applications to optimal design are discussed in Section 3. Section 4 introduces the so-called Poisson designs that appear if the design points are distributed

according to a non-homogeneous Poisson point process. Finally, Section 5 is devoted to steepest descent type algorithms stemming from our approach. While the corresponding algorithms were described in [21], here we provide full derivations of optimal descent directions.

2. Optimisation in the space of signed measures

In this section we state first-order necessary conditions, for an extremum in constrained optimisation problems on the cone of positive measures. Let X be a locally compact separable metric space with its Borel σ-algebra $\mathfrak{B}$. Further, $\mathbb{M}$ denotes the family of all non-negative finite measures on $\mathfrak{B}$; $\widetilde{\mathbb{M}}$ is the Banach space of all signed measures on $\mathfrak{B}$ with bounded total variation and equipped with the total variation norm (see, *e.g.* [8], Sect. IV.2.16). The Jordan decomposition of a signed measure $\mu \in \widetilde{\mathbb{M}}$ is denoted by

$$\mu = \mu^+ - \mu^- \,,$$

so that $\|\mu\| = \mu^+(X) + \mu^-(X)$ is the total variation of μ. Furthermore, $\mu|_B$ is the restriction of a measure μ onto a Borel set B, *i.e.*

$$\mu|_B(\bullet) = \mu(\bullet \cap B)\,.$$

Let Y be a Banach space with dual space Y^*, and let $u{\cdot}y$ be the canonical bilinear form defined for $y \in Y$ and $u \in Y^*$. Often Y is the Euclidean space $\mathbb{R}^m$ and $u{\cdot}y$ is the conventional scalar product in $\mathbb{R}^m$.

Consider the following optimisation problem:

$$\Psi(\mu) \to \inf, \quad \mu \in \mathbb{A}, \quad H(\mu) \in C\,, \tag{2.1}$$

where $\mathbb{A}$ and C are closed convex subsets of $\mathbb{M}$ and Y, respectively, $\Psi : \widetilde{\mathbb{M}} \to \mathbb{R}$ and $H : \widetilde{\mathbb{M}} \to Y$ are continuous functions.

The following is the definition of Robinson's regularity condition ([28], Th. 1) which, as shown in [38], guarantees the existence and boundedness of the Lagrange multipliers (see also [20] for the discussion of different forms of the regularity condition).

Definition 2.1. A measure μ is called *regular* for Problem (2.1) if

$$0 \in \mathsf{core}\big\{H(\mu) - C + DH(\mu)[\mathbb{A} - \mu]\big\}\,, \tag{2.2}$$

where $\mathsf{core}\,A$ is the set of all $y \in A \subset Y$ such that $y + ty_1 \in A$ for all $y_1 \in Y$ and all sufficiently small positive t. Here the '+' (respectively, '−') operation on sets designates all pairwise sums of (respectively, differences between) the points from the corresponding sets.

The family $\mathbb{M}$ of all non-negative measures is a cone in $\widetilde{\mathbb{M}}$ with a very rich boundary (in the topology generated by the total variation norm) that coincides with the cone itself if X is non-countable. Indeed, it suffices to subtract from μ an atom of a small mass located at a point where μ has no atom in order to obtain a signed measure arbitrarily close to μ in the total variation norm.

Optimality conditions in (2.1) involving the derivative of Ψ typically refer to tangent cones drawn to $\mathbb{A}$ near the feasible solution. Recall that the tangent cone $T_B(x)$ to a set B at a point x in a Banach space is defined as the set of limits as $t \downarrow 0$ for all convergent sequences $\{x_t, t > 0\}$ such that $x_t \in (B - x)/t$ for all $t > 0$.

The following theorem from [23] gives first-order necessary conditions for a minimum in Problem (2.1). Note that no convexity of Ψ is assumed.

Theorem 2.1. *Assume that both $\Psi : \widetilde{\mathbb{M}} \to \mathbb{R}$ and $H : \widetilde{\mathbb{M}} \to Y$ are continuous on $\mathbb{A}$ and Fréchet differentiable at a regular $\mu \in \mathbb{A}$ such that $H(\mu) \in C$. If μ is a local minimum in Problem (2.1), then for all $\eta \in T_{\mathbb{A} \cap H^{-1}(C)}(\mu)$ one has*

$$D\Psi(\mu)[\eta] \geq 0\,.$$

Moreover, there exists (a Lagrange multiplier or a Kuhn-Tucker vector) $u \in Y^*$ such that $u \cdot y \geq 0$ for every $y \in T_C\big(H(\mu)\big)$, and, for the Lagrangian $L(\mu) = \Psi(\mu) - u \cdot H(\mu)$, the following inequality holds:

$$DL(\mu)[\eta] = D\Psi(\mu)[\eta] - u \cdot DH(\mu)[\eta] \geq 0 \quad \text{for all} \ \ \eta \in T_{\mathbb{A}}(\mu) \, .$$

In order to apply Theorem 2.1 one has to characterize tangent cones in the space of measures. By [5] (Th. 3.1), for a regular μ, differential $DH(\mu)$ is surjective on Y and one can incorporate the constraints as

$$T_{\mathbb{A} \cap H^{-1}(C)}(\mu) = T_{\mathbb{A}}(\mu) \cap \big(DH(\mu)\big)^{-1}\big[T_C(H(\mu))\big] \, .$$

Tangent cones to various families of measures are characterized in [22].

Remark 2.1. Although differentiability of Ψ implies its continuity in the total variation norm, often this is not sufficient to deduce that Ψ attains its minimum at some measure μ, because many interesting families of measures are not compact in the total variation norm. However, if Ψ is continuous in some other topology in which the set of feasible measures *is* compact then Ψ does attain its minimum value. For instance, this will be the case if Ψ is continuous in the weak topology and $\mathbb{A}$ is the family of measures $\{\mu \in \mathbb{M} : \mu(X) \leq \text{Const}\}$ on a *compact X*.

3. General equivalence theorem and examples

Most differentiable functionals of measures met in practice have derivatives which can be represented in the integral form. Consider a common case of finitely many differentiable constraints of the equality and inequality types:

$$\begin{cases} H_i(\mu) = 0 \, , & i = 1, \dots, m \, ; \\ H_j(\mu) \leq 0 \, , & j = m+1, \dots, l \, . \end{cases} \tag{3.1}$$

Here $Y = Y^* = \mathbb{R}^l$, $H = (H_1, \dots, H_l) : \widetilde{\mathbb{M}} \to \mathbb{R}^l$, $C = \{0\}^m \times (-\infty, 0]^{l-m}$ and $u \cdot y$ is the scalar product of vectors u and y in $\mathbb{R}^l$. Note that all integrals are over X unless specified otherwise. The following theorem immediately follows from [22] (Th. 4.1).

Theorem 3.1. *Let μ be a regular local minimum of $\Psi(\mu)$ over $\mu \in \mathbb{M}$ subject to (3.1). Assume that Ψ and H are continuous on $\mathbb{M}$ and Fréchet differentiable at μ and there exist measurable real-valued functions $d_\Psi(x, \mu)$ and $h_i(x, \mu)$, $i = 1, \dots, l$, such that for all $\eta \in \widetilde{\mathbb{M}}$*

$$D\Psi(\mu)[\eta] = \int d_\Psi(x, \mu) \eta(\mathrm{d}x) \quad and \quad DH(\mu)[\eta] = \int h(x, \mu) \eta(\mathrm{d}x) \, , \tag{3.2}$$

where $h = (h_1, \dots, h_l)$. Then there exists $u = (u_1, \dots, u_l)$, where $u_j \leq 0$ (resp. $u_j = 0$) for those $j \in \{m+1, \dots, l\}$ satisfying $H_j(\mu) = 0$ (resp. $H_j(\mu) < 0$), such that

$$\begin{cases} d_\Psi(x, \mu) = u \cdot h(x, \mu) & \text{μ-almost everywhere,} \\ d_\Psi(x, \mu) \geq u \cdot h(x, \mu) & \text{for all } x \in X \, . \end{cases} \tag{3.3}$$

Optimal design problems can be naturally treated within the above described general framework of optimisation of functionals defined on finite measures. Theorem 3.1 directly applies to functionals of measures typical in the optimal design literature and under quite general constraints. Although we do not assume any convexity assumptions on Ψ and work exclusively with necessary optimal conditions, convexity of Ψ immediately ensures that necessary conditions become sufficient, as typically happens in [16].

The following result specialises Theorem 3.1 for functionals Ψ that effectively depend on the information matrix. Recall that, given an objective function $\Psi_0 : \mathbb{R}^{k^2} \to \mathbb{R}$, a Ψ-*optimal design measure* μ minimizes

$\Psi(\mu) = \Psi_0(M(\mu))$ over all probability measures on X (or, more generally, over all measures with the fixed total mass $\mu(X) = a$). In most cases it is possible to extend the functional Ψ onto the family $\widetilde{\mathbb{M}}$ of all signed measures, allowing for infinite values, whenever the information matrix becomes degenerated.

Let $m_{ij} = \int f_i(x)f_j(x)\,\mu(\mathrm{d}x)$ denote the (i,j)-th entry of the information matrix $M(\mu)$. Similarly to [16], by the chain rule,

$$
\begin{aligned}
D\Psi(\mu)[\eta] &= \sum_{i,j} \frac{\partial \Psi_0(M)}{\partial m_{ij}}(\mu) Dm_{ij}(\mu)[\eta] \qquad (3.4) \\
&= \sum_{i,j} \frac{\partial \Psi_0(M)}{\partial m_{ij}}(\mu) \int f_i(x)f_j(x)\eta(\mathrm{d}x) \\
&= \int f(x)D\Psi_0(M)(\mu)f^\top(x)\eta(\mathrm{d}x)\,.
\end{aligned}
$$

Theorem 3.2. *Assume that* (3.2) *holds and μ is a regular Ψ-optimal design measure under constraints* (3.1). *Then*

$$
\begin{cases}
d_\Psi(x,\mu) = u{\cdot}h(x,\mu) & \mu\text{-almost everywhere}, \\
d_\Psi(x,\mu) \geq u{\cdot}h(x,\mu) & \text{for all } x \in X\,,
\end{cases} \qquad (3.5)
$$

for some $u = (u_1,\ldots,u_k) \in \mathbb{R}^k$ with $u_j \leq 0$ for those $j \in \{m+1,\ldots,k\}$ for which $H_j(\mu) = 0$ and $u_j = 0$ if $H_j(\mu) < 0$, where $d_\Psi(x,\mu) = f(x)D\Psi_0(M)(\mu)f^\top(x)$ and

$$
D\Psi_0(M)(\mu) = \left\| \frac{\partial \Psi_0(M)}{\partial m_{ij}} \right\|_{ij}(\mu)\,.
$$

If Ψ is convex then (3.5) *is equivalent to μ being Ψ-optimal.*

Example 3.1. *(D-optimal design)* Let $\Psi_0(M) = -\log\det M$. Then $D\Psi_0(M)(\mu) = -M^{-1}(\mu)$. The only constraint $H(\mu) = \mu(X) - a$ has the derivative $h(x,\mu) \equiv 1$. Any measure is regular in this case, and $d_\Psi(x,\mu) = -f(x)M^{-1}(\mu)f^\top(x) = -\mathrm{d}(x,\mu)$, so that Theorem 3.2 becomes the Kiefer-Wolfowitz characterization of D-optimal designs (since Ψ is convex).

An advantage of our approach is that one can easily incorporate additional constraints on the design measure. The only change concerns the right-hand side of (3.5).

Example 3.2. *(D-optimal design with fixed moments)* Let $X = \mathbb{R}^d$ and assume that along with the constraint on the total mass $\mu(X) = 1$ we fix the expectation of μ, which is a vector $m = \int x\mu(\mathrm{d}x)$. These constraints can be written as $H(\mu) = (m,0)$, where $H(\mu)$ is a $(d+1)$-dimensional vector function with the components $H_i(\mu) = \int x_i\mu(\mathrm{d}x)$ for $i = 1,\ldots,d$ and $H_{d+1}(\mu) = \mu(X) - 1$. Clearly, $DH(\mu)[\eta] = \int h(x,\mu)\eta(\mathrm{d}x)$ for $h(x,\mu) = (x,1)$, $x \in \mathbb{R}^d$. By Theorem 3.1, if μ minimizes $\Psi(\mu)$ under the conditions $H(\mu) = (m,0)$, then

$$
\begin{cases}
d_\Psi(x,\mu) = v{\cdot}x + u & \mu\text{-almost everywhere}, \\
d_\Psi(x,\mu) \geq v{\cdot}x + u & \text{for all } x \in X\,,
\end{cases}
$$

for some $v \in \mathbb{R}^d$. In other words, $d(x,\mu)$ is *affine* for μ-almost all x.

Assume in addition that the generalized variance of the design measure is fixed at a value $z > 0$. This variance is defined as the determinant of the matrix

$$
G_\mu = \left\| \int x_i x_j \mu(\mathrm{d}x) - \left(\int x_i\mu(\mathrm{d}x)\right)\left(\int x_j\mu(\mathrm{d}x)\right) \right\|_{i,j}\,.
$$

This change of the problem's formulation manifests itself just in adding one more constraint $H_{d+2}(\mu) = \log(\det G_\mu) - \log z = 0$ to those imposed above. If we use the fact that the mean is fixed to m, the gradient of this additional constraint is equal to

$$g(x, m, \mu) = x^\top G_\mu^{-1} x - x^\top G_\mu^{-1} m - m^\top G_\mu^{-1} x.$$

Then the first-order necessary condition becomes

$$\begin{cases} d_\Psi(x, \mu) = v{\cdot}x + u_0 + u_1 g(x, m, \mu) & \mu\text{-almost everywhere,} \\ d_\Psi(x, \mu) \geq v{\cdot}x + u_0 + u_1 g(x, m, \mu) & \text{for all } x \in X, \end{cases}$$

for some $u_0, u_1 \in \mathbb{R}$ and $v \in \mathbb{R}^d$. Note that $g(x, m, \mu) = x^2 - 2xm$ for $d = 1$.

Example 3.3. *(D-optimal designs with a limited total cost)* Let $h(x)$ determine the cost of taking an observation at point x. If the total cost is bounded by C, then the design measure μ should, in addition to $\mu(X) = 1$, satisfy the constraint $\int h(x)\mu(\mathrm{d}x) \leq C$. By Theorem 3.1, if μ provides a Ψ-optimal design under such constraints, then

$$\begin{cases} d_\Psi(x, \mu) = u + vh(x) & \mu\text{-almost everywhere,} \\ d_\Psi(x, \mu) \leq u + vh(x) & \text{for all } x \in X, \end{cases}$$

for some $u \in \mathbb{R}$ and $v \leq 0$ if $\int h(x)\mu(\mathrm{d}x) = C$ and $v = 0$ if $\int h(x)\mu(\mathrm{d}x) < C$.

Since the function Ψ in the D-optimality problem is a convex function of μ, the above necessary conditions become necessary and sufficient conditions, thus providing a full characterization of the optimal designs. It is fairly easy to replace D-optimality with some other goal functional by calculating the corresponding derivative and using the same approach to deal with the constraints.

Example 3.4. *(D-optimal designs with bounded densities)* As discussed in [6, 11], there are situations where the design measures need to have (possibly, bounded) densities with respect to a fixed reference measure λ. In our framework this corresponds to an optimisation problem in the space $\mathbb{M}_\lambda$ of measures absolutely continuous with respect λ or in the space

$$\mathbb{M}_\lambda^C = \left\{ \mu \in \mathbb{M}_\lambda : \phi_\mu(x) \leq C \right\}.$$

of measures whose Radon-Nikodym derivative $\phi_\mu(x) = \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x)$ with respect to λ exists and is bounded by C.

Recall the definitions of essential extrema:

$$\lambda\text{--ess-inf}_B f(x) = \sup_{E:\, \lambda(E)=0} \inf_{B \setminus E} f(x), \quad \lambda\text{--ess-sup}_B f(x) = \inf_{E:\, \lambda(E)=0} \sup_{B \setminus E} f(x).$$

The following result follows from [22] (Th. 5.1).

Theorem 3.3. *Let μ be a regular local minimum of $\Psi(\mu)$ over $\mu \in \mathbb{A}$ subject to (3.1). Assume that Ψ and H are continuous on $\mathbb{A}$ and Fréchet differentiable at μ and there exist measurable real-valued functions $d_\Psi(x, \mu)$ and $h_i(x, \mu)$, $i = 1, \dots, k$, such that (3.2) holds. Then there exists $u = (u_1, \dots, u_k) \in \mathbb{R}^k$, where $u_j \leq 0$ (resp. $u_j = 0$) for those $j \in \{m + 1, \dots, l\}$ satisfying $H_j(\mu) = 0$ (resp. $H_j(\mu) < 0$), such that*

(i) *in the case $\mathbb{A} = \mathbb{M}_\lambda$ one has*

$$\begin{cases} d_\Psi(x, \mu) - u{\cdot}h(x) = 0 & \mu\text{-almost everywhere,} \\ \lambda\text{--ess-inf}\left(d_\Psi(x, \mu) - u{\cdot}h(x)\right) \geq 0; \end{cases}$$

(ii) *and in the case* $\mathbb{A} = \mathbb{M}_\lambda^C$ *one has*

$$\begin{cases} d_\Psi(x,\mu) - u \cdot h(x) = 0 & \mu|_{\{\phi_\mu < C\}}\text{-}almost\ everywhere\,, \\ \lambda\text{--ess-sup}_{\{\phi_\mu = C\}}\left(d_\Psi(x,\mu) - u \cdot h(x)\right) \leq 0\,, \\ \lambda\text{--ess-inf}_{\{\phi_\mu < C\}}\left(d_\Psi(x,\mu) - u \cdot h(x)\right) \geq 0\,. \end{cases}$$

Another characterization theorem for optimal designs with bounded densities was obtained in [27] (Th. 11.8) in terms of the information matrix and under the only constraint on the total mass of μ, see also [10].

4. POISSON DESIGN

Every finite measure μ on X corresponds to a Poisson point process Π (with possibly multiple points) with μ being its intensity measure. We treat each realization of Π as a counting measure, so that $\Pi(B)$ equals the number of the points in a Borel set B. Note that $\Pi(B)$ follows a Poisson distribution with mean $\mu(B)$ and $\Pi(B_1), \ldots, \Pi(B_k)$ are mutually independent for disjoint $B_1, \ldots, B_k$. In what follows, $\mathbf{E}_\mu$ denotes the expectation with respect to the distribution of the Poisson process Π with intensity measure μ.

In this section we define a new class of design problems where the observation points are distributed according to a Poisson point process Π, and one can control them only through the process' intensity measure μ.

Such a situation is typical in bio-engineering experiments as described, for instance, in the paper [18]. Experiments are carried out studying the repopulation rate of mice bearing cancer cells as a function of a number of haematopoietic stem cells received through injection. The experimentator may vary the dilution of the liquid derived from stem cells' bearing material in a doze, but he cannot control the number of stem cells in it. As even a small amount of cells is capable of repopulating a mouse, the interest is to keep the concentration of cells so low that to have just a few cells per doze. Therefore the number of cells in a doze may be assumed to follow Poisson distribution with the parameter proportional to the inverse of the dilution rate. The P-optimal design is thus aims at identifying the set of dilutions in injected dozes and their corresponding frequencies in order to achieve the minimum of the goal function which along with statistical properties of the estimator should also include a very high cost of the specially elevated mouse should it dies as a result of the experiment.

In this case the fixed total number of points n is replaced by the constraint $\mu(X) = a$ on the fixed average total number of points. Since Poisson process arises as a limit in various point process convergence and thinning schemes, see [7] (Chap. 9), such a model naturally appears also when the locations of observations are subject to errors or measurements may fail at some originally suggested design points.

Consider a general continuous functional Ψ that depends on counting measures and assesses the quality of the standard least squares estimator $\widehat{\beta}$ in (1.1) obtained by choosing the points of Π as design points. Then it is natural to choose μ that maximizes $\mathbf{E}_\mu \Psi(\Pi)$ among all measures μ satisfying constraints (3.1). We will call such μ a *P-optimal* design.

It is possible to find the derivative of $\mathbf{E}_\mu \Psi(\Pi)$ with respect to μ by appealing to the following result proved in [23]. Assume that

$$|\Psi(\Pi)| \leq c^{\Pi(X)} \quad \text{for some } c > 0 \text{ and all realisations of } \Pi. \tag{4.1}$$

Then $f(\mu) = \mathbf{E}_\mu \Psi(\Pi)$ is Fréchet differentiable and

$$Df(\mu)[\eta] = \int d_\Psi(x,\mu)\eta(\mathrm{d}x)\,, \tag{4.2}$$

where the gradient is given by

$$d_\Psi(x,\mu) = \mathbf{E}_\mu[\Psi(\Pi + \delta_x) - \Psi(\Pi)]\,. \tag{4.3}$$

Note that $d_\Psi(x,\mu)$, called the *expected first difference*, is the expected increment of Ψ when a new point located at x (or unit mass measure δ_x concentrated at x) is added to the realization Π.

Using (4.2), we may now apply Theorem 3.1 to arrive at the following result:

Theorem 4.1. *Let the function $\Psi(\Pi)$ satisfy (4.1), and let μ provide a regular P-optimal design under constraints (3.1) satisfying (3.2). Then the expected first-order difference $d_\Psi(x, \mu)$ given by (4.3) satisfies (3.3). If Ψ is convex, then (3.3) is also sufficient for μ to be a P-optimal design.*

As an example, consider $\Psi(\Pi) = \det M(\Pi)$, where the random information matrix $M(\Pi)$ is given by

$$M(\Pi) = \int f(x)^\top f(x)\, \Pi(\mathrm{d}x) = \sum_{x_i \in \Pi} f^\top(x_i) f(x_i). \tag{4.4}$$

Functional Ψ is convex. Indeed, the same arguments as in the standard D-optimal case are applicable here, since the functional $(-\log \det M)$ is convex on the family of information matrices. The exponential boundedness condition (4.1) is satisfied if and only if all $f_j(\cdot)$ are bounded on X.

For this particular functional Ψ, it appears that the P-optimal design coincides with the D-optimal design. This is due to the fact that the Poisson process expectation and determinant of an information matrix commute:

$$\mathbf{E}_\mu \det M(\Pi) = \det \mathbf{E}_\mu M(\Pi) = \det M(\mu). \tag{4.5}$$

The last equality follows from the Campbell theorem, see, *e.g.* [7], p. 188):

$$\mathbf{E}_\mu M(\Pi) = \int f(x)^\top f(x)\, \mu(\mathrm{d}x) = M(\mu).$$

To justify the first equality, we write the definition of the determinant in terms of the sum over all permutations σ of k indexes

$$
\begin{aligned}
\det M(\Pi) &= \sum_\sigma (-1)^{|\sigma|} M_{1\sigma_1} \ldots M_{k\sigma_k} \\
&= \sum_\sigma (-1)^{|\sigma|} \left(\int f_1(x) f_{\sigma_1}(x) \Pi(\mathrm{d}x) \right) \times \ldots \times \left(\int f_k(x) f_{\sigma_k}(x) \Pi(\mathrm{d}x) \right) \\
&= \sum_\sigma (-1)^{|\sigma|} \int \cdots \int f_1(x_1) f_{\sigma_1}(x_1) \ldots f_k(x_k) f_{\sigma_k}(x_k)\, \Pi(dx_1) \ldots \Pi(\mathrm{d}x_k) \\
&= \sum_\sigma (-1)^{|\sigma|} \sum_{x_1, \ldots, x_k \in \Pi} f_1(x_1) f_{\sigma_1}(x_1) \ldots f_k(x) f_{\sigma_k}(x) \\
&= \sum_{x_1, \ldots, x_k \in \Pi} \det \Phi(x_1, \ldots, x_k), \tag{4.6}
\end{aligned}
$$

where $\Phi(x_1, \ldots, x_k)$ is a k by k matrix with elements $f_i(x_i) f_j(x_i)$ at the (i, j)th place and $|\sigma|$ is the rank of the permutation. Note that if $x_l = x_m$ for some $l \neq m$ then the lth and mth rows in the matrix are identical and thus its determinant vanishes. Therefore the last sum can only be taken over k-tuples of *distinct* points $x_1, \ldots, x_k \in \Pi$ thus giving

$$
\begin{aligned}
\mathbf{E} \det M(\Pi) &= \mathbf{E} \sum_{x_1, \ldots, x_k \in \Pi}^{\neq} \det \Phi(x_1, \ldots, x_k) \\
&= \int \cdots \int \det \Phi(x_1 \ldots \mathrm{d}x_k)\, \mu(\mathrm{d}x_1) \ldots \mu(\mathrm{d}x_k).
\end{aligned}
$$

The last equality reflects the total independence of the Poisson process points and can be found, *e.g.* in [30] (p. 44). As the last expression has the same form as the third line in (4.6) (with μ in place of Π) then it is simply $\det M(\mu)$.

Therefore the P-optimal design μ also maximizes $\det \mathbf{E}_\mu M(\Pi)$ and thus is the D-optimal design for $\det M(\mu)$. This fact also allows for another non-asymptotic interpretation of a D-design measure: it is the optimal intensity measure of the Poisson point process (with average total number of points equal to one) with the points of this process chosen as the design points.

5. Gradient method for search of optimal design

Various descent algorithms are widely studied in the optimal design literature as they allow one to find the optimal design numerically if the "exact" analytical solution is not feasible. The gradient descent method relies on knowledge of the derivative of the objective function. We first consider the general set-up, which will be specialized later for particular constraints and objective functionals. As before, it is assumed that $\Psi(\mu)$ is Fréchet differentiable and its derivative is representable in the integral form (3.2).

The most basic method of the gradient descent type used in the optimal design suggests moving from μ_n (the approximation on step n) to $\mu_{n+1} = (1 - \alpha_n)\mu_n + \alpha_n \zeta_n$, where $0 < \alpha_n < 1$ and ζ_n minimizes $\tilde{D}\Psi(\mu)[\zeta]$ over all probability measures ζ, see, e.g. [36]. It is easy to see that such ζ_n is concentrated at the points where the corresponding gradient function $d_\Psi(x, \mu)$ is minimised. Rearranging the terms, we obtain

$$\mu_{n+1} = \mu_n + \alpha_n(\zeta_n - \mu_n). \tag{5.1}$$

In this form the algorithm looks like a conventional descent algorithm that descends along the direction $\eta_n = \alpha_n(\zeta_n - \mu_n)$. In this particular case, the step size is $\|\eta_n\| \leq 2\alpha_n$, and

$$\eta_n^+ = \alpha_n \zeta_n, \quad \eta_n^- = \alpha_n \mu_n.$$

Such a choice of η_n ensures that $\mu_n + \eta_n$ remains a probability measure guaranteed by the fact that the negative part of η_n is proportional to μ_n with $\alpha_n < 1$.

However, if one does not restrict the choice for the descent direction to measures with the negative part proportional to μ_n, then it is possible to find a *steeper* descent direction than that given by (5.1). If the current value $\mu_n = \mu$, then the "steepness" of the descent direction η is characterized by the directional derivative

$$D\Psi(\mu)[\eta] = \int d_\Psi(x, \mu)\eta(\mathrm{d}x).$$

Then the true steepest descent direction must be chosen to minimise $D\Psi(\mu)[\eta]$ over all signed measures η with total variation $\|\eta\| \leq \alpha = 2\alpha_n$ such that $\mu + \alpha\eta$ is a non-negative measure satisfying all specified constraints. For example, if the only constraint is that the total mass $\mu(X) = a$, then η should deliver the zero total mass, whence $\eta(X) = 0$.

A straightforward generalisation of a single constraint on the total mass is the case of several linear constraints written in integral form:

$$H_i(\mu) = \int h_i(x)\mu(\mathrm{d}x) = a_i, \quad i = 0, 1, \ldots, k, \tag{5.2}$$

where $H_0(\mu) = \mu(X) = a = a_0$ is the fixed total mass condition and $a_1, \ldots, a_k$ are given real numbers. In vector form, $H(\mu) = \int h(x)\mathrm{d}\mu = A$, where $H = (H_0, \ldots, H_k)$, $h = (1, h_1, \ldots, h_k)$ and $A = (a, a_1, \ldots, a_k)$.

For a $\mu \in \mathbb{M}$ denote by Υ_μ the family of all signed measures $\eta \in \widetilde{\mathbb{M}}$ such that $\mu + \eta \in \mathbb{M}$ and $\mu + \eta$ satisfies the constraints (5.2). Further, ε denotes a positive constant that controls the size of the step. The following result provides a characterization of the steepest direction. Recall that vectors $w_0, w_1, \ldots, w_k$ are called *affinely independent* if $w_1 - w_0, \ldots, w_k - w_0$ are linearly independent.

Theorem 5.1. *The minimum of $D\Psi(\mu)[\eta]$ over all $\eta \in \Upsilon_\mu$ with $\|\eta\| \leq \varepsilon$ is achieved on a signed measure η such that η^+ has at most $k + 1$ atoms and η^- is representable as a convex combination $\sum_{i=0}^{k+1} t_i \mu|_{B_i}$ for*

some $0 \leq t_i \leq 1$, $\sum_{i=0}^{k+1} t_i = 1$ *and some measurable sets* B_i, *such that* $H(\mu|_{B_i})$, $i = 0, \ldots, k+1$ *are affinely independent vectors.*

Proof. Let η_0 minimise $D\Psi(\mu)[\eta]$ with respect to η and satisfy the required conditions. By [32] (Th. 2.1), there exists a measure η^+ with at most $k+1$ atoms such that

$$H(\eta^+) = H(\eta_0^+) \quad \text{and} \quad D\Psi(\mu)[\eta^+] = D\Psi(\mu)[\eta_0^+].$$

Next, the set $\mathbb{A}_\mu = \{\nu \in \mathbb{M} : \mu - \nu \in \mathbb{M}\}$ is a convex linearly compact subset of $\mathbb{M}$ and H is a linear map from $\mathbb{M}$ to $\mathbb{R}^{k+1}$. By [32] (Prop. 2.1) the extreme points of the set $\mathbb{H} = \mathbb{A}_\mu \cap H^{-1}(\eta_0^-)$ are contained in the convex combinations of the type $\sum_{i=0}^{k+1} t_i \, \mu|_{B_i}$ with affinely independent $H(\mu|_{B_i})$, $i = 0, \ldots, k+1$. Therefore, the maximum of a linear functional $D\Psi(\mu)[\nu]$ over all $\nu \in \mathbb{H}$ is attained at one of such measures which can then be taken to replace η_0^-. Therefore, $\eta^+ - \eta^-$ can be taken instead of η_0 without changing the value of the derivative $D\Psi(\mu)[\eta]$. □

Corollary 5.2. *If the only constraint is* $\mu(X) = a$, *then the minimum of* $D\Psi(\mu)[\eta]$ *over all* $\eta \in \Upsilon_\mu$ *with* $\|\eta\| \leq \varepsilon$ *is achieved on a signed measure* η *such that* η^+ *is the positive measure of total mass* $\varepsilon/2$ *concentrated on the points of the global minima of* $d_\Psi(x, \mu)$ *and* $\eta^- = \mu|_{M(t_\varepsilon)} + \delta\mu|_{M(s_\varepsilon) \setminus M(t_\varepsilon)}$, *where*

$$M(p) = \{x \in X : d_\Psi(x, \mu) \geq p\},$$

and

$$t_\varepsilon = \inf\{p : \mu(M(p)) < \varepsilon/2\},$$
$$s_\varepsilon = \sup\{p : \mu(M(p)) \geq \varepsilon/2\}.$$

The factor δ *is chosen in such a way that* $\mu(M(t_\varepsilon)) + \delta\mu(M(s_\varepsilon) \setminus M(t_\varepsilon)) = \varepsilon/2$.

Proof. By Theorem 5.1, we can search the optimal η among all signed measures with η^+ concentrated at a single point x_0 and η^- being the restriction of μ. This is equivalent to minimizing the integral

$$\frac{\varepsilon}{2} d_\Psi(x_0, \mu) - \int_B d_\Psi(x, \mu)\mu(\mathrm{d}x).$$

The first term is minimised if x_0 is a point of the set of global minima of $d_\Psi(x, \mu)$. To maximize the second term over $\eta = \mu|_B$, the set B should consist of the points where $d_\Psi(x, B)$ is as large as possible. Therefore, if the decreasing function $\mu(M(p))$ takes the value $\varepsilon/2$ for some $p = t_\varepsilon$, then $t_\varepsilon = s_\varepsilon$ and $\eta^- = \mu|_{M(t_\varepsilon)}$ provides the maximum. If $\mu(M(p))$ is discontinuous at the point t_ε then $\mu|_{M(t_\varepsilon)}$ has total mass smaller than $\varepsilon/2$ and the rest of the mass is provided by $\delta\mu|_{M(s_\varepsilon) \setminus M(t_\varepsilon)}$, leading to $\eta^- = \mu|_{M(t_\varepsilon)} + \delta\mu|_{M(s_\varepsilon) \setminus M(t_\varepsilon)}$. Both cases correspond to form of η^- given by Theorem 5.1 with $t_0 = 1$, $t_1 = 0$ in the first case and $(1-\delta)\mu|_{M(t_\varepsilon)} + \delta\mu|_{M(s_\varepsilon)}$ in the second. □

Remark 5.1. (*Step length and line search*)
1. As Corollary 5.2 determines the true steepest direction in the space of measures, it is possible to use a variety of well-established procedures from optimisation theory in order to find the optimal step sizes. For example, the Armijo method of choosing the step-length [26] may be used. All convergence results concerning gradient descent algorithms are then applied literally to finding optimal designs.

2. It is interesting to note that, without constraint $\|\eta\| \leq \varepsilon$ on the total variation norm of the increment measure, the steepest direction that preserves the total mass a is the measure $\eta = a\delta_{x_0} - \mu$, where x_0 is a global minimum point of $d_\Psi(x, \mu)$. The classical descent algorithm (5.1) based on the modified directional derivative (1.5) uses, in fact, the scaled variant of this direction.

3. On a discrete state space X, the descent along the steepest direction provided by Corollary 5.2 with a very small step size ε is equivalent to the descent by Vertex Exchange Method (VEM) suggested in [4] as η^- is concentrated on a single atom being the point of global maximum of $d_\Psi(x, \mu)$ for such ε. Thus VEM uses the scaled variant of the true steepest direction.

The necessary condition given in Theorem 3.1 can be used as a stopping rule for descent algorithms. For instance, if μ is optimal in the problem with a fixed mass then $d_\Psi(x, \mu)$ is constant on the support of μ. Therefore, the difference between the supremum of $d_\Psi(x, \mu)$ over the support of μ and the global infimum can be used to indicate how close is μ to the solution.

On discrete design spaces, which are the only ones used in numerical methods, minimisation of $D\Psi(\mu)[\eta]$ over all signed measures $\eta \in \Upsilon_\mu$ with $\|\eta\| = \varepsilon$ becomes a linear programming problem of dimension equal to the cardinality of X, see also [13]. Therefore, if the cardinality is large and/or in the presence of many constraints, it might be computationally more effective to use an approximation to the exact steepest direction described in [21]. An additional analysis is however necessary to ensure that the algorithm does converge to the minimal solution.

The steepest descent method described above has been programmed in the `SPLUS` and `R` languages. For more detailed discussion on the algorithms, numerical examples and codes, see [21].

References

[1] A.C. Atkinson and A.N. Donev, *Optimum Experimental Designs.* Clarendon Press, Oxford (1992).

[2] C.L. Atwood, Sequences converging to D-optimal designs of experiments. *Ann. Statist.* **1** (1973) 342-352.

[3] C.L. Atwood, Convergent design sequences, for sufficiently regular optimality criteria. *Ann. Statist.* **4** (1976) 1124-1138.

[4] D. Böhning, A vertex-exchange-method in D-optimal design theory. *Metrika* **33** (1986) 337-347.

[5] R. Cominetti, Metric regularity, tangent sets, and second-order optimality conditions. *Appl. Math. Optim.* **21** (1990) 265-287.

[6] D. Cook and V. Fedorov, Constrained optimization of experimental design. *Statistics* **26** (1995) 129-178.

[7] D.J. Daley and D. Vere–Jones, *An Introduction to the Theory of Point Processes.* Springer, New York (1988).

[8] N. Dunford and J.T. Schwartz, *Linear Operators. Part I: General Theory.* Wiley, New York (1988).

[9] V.V. Fedorov, *Theory of Optimal Experiments.* Academic Press, New York (1972).

[10] V.V. Fedorov, Optimal design with bounded density: Optimization algorithms of the exchange type. *J. Statist. Plan. Inf.* **22** (1989) 1-13.

[11] V.V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments.* Springer, New York, *Lecture Notes in Statist.* **125** (1997).

[12] I. Ford, *Optimal Static and Sequential Design: A Critical Review*, Ph.D. Thesis. Department of Statistics, University of Glasgow, Glasgow (1976).

[13] A. Gaivoronski, Linearization methods for optimization of functionals which depend on probability measures. *Math. Progr. Study* **28** (1986) 157-181.

[14] N. Gaffke and R. Mathar, On a Class of Algorithms from Experimental Design Theory. *Optimization* **24** (1992) 91-126.

[15] E. Hille and R.S. Phillips, *Functional Analysis and Semigroups.* American Mathematical Society, Providence, *AMS Colloquium Publications* **XXXI** (1957).

[16] J. Kiefer, General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** (1974) 849-879.

[17] J. Kiefer and J. Wolfowitz, The equivalence of two extremal problems. *Canad. J. Math.* **14** (1960) 363-366.

[18] P. Kumaravelu, L. Hook, A.M. Morrison, J. Ure, S. Zhao, S. Zuyev, J. Ansell and A. Medvinsky, Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): Role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development* **129** (2002) 4891-4899.

[19] E.P. Liski, N.K. Mandal, K.R. Shah and B.K. Singha, *Topics in Optimal Design.* Springer, New York, *Lect. Notes Statist.* **163** (2002).

[20] H. Maurer and J. Zowe, First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Programming* **16** (1979) 98-110.

[21] I. Molchanov and S. Zuyev, Steepest descent algorithms in space of measures. *Statist. and Comput.* **12** (2002) 115-123.

[22] I. Molchanov and S. Zuyev, Tangent sets in the space of measures: With applications to variational calculus. *J. Math. Anal. Appl.* **249** (2000) 539-552.

[23] I. Molchanov and S. Zuyev, Variational analysis of functionals of a Poisson process. *Math. Oper. Res.* **25** (2000) 485-508.

[24] C.H. Müller and A. Pázman, Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika* **48** (1998) 1-19.

[25] A. Pázman, Hilbert-space methods in experimantal design. *Kybernetika* **14** (1978) 73-84.

[26] E. Polak, *Optimization. Algorithms and Consistent Approximations.* Springer, New York (1997).

[27] F. Pukelsheim, *Optimal Design of Experiments.* Wiley, New York (1993).

[28] S.M. Robinson, First order conditions for general nonlinear optimization. *SIAM J. Appl. Math.* **30** (1976) 597-607.

[29] S.D. Silvey, *Optimum Design.* Chapman & Hall, London (1980).

[30] D. Stoyan, W.S. Kendall and J. Mecke, *Stochastic Geometry and its Applications*, Second Edition. Wiley, Chichester (1995).

[31] P. Whittle, Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** (1973) 123-130.

[32] G. Winkler, Extreme points of moment sets. *Math. Oper. Res.* **13** (1988) 581-587.

[33] C.-F. Wu, Some algorithmic aspects of the theory of optimal design. *Ann. Statist.* **6** (1978) 1286-1301.

[34] C.-F. Wu, Some iterative procedures for generating nonsingular optimal designs. *Comm. Statist. Theory Methods A* **7** (1978) 1399-1412.

[35] C.-F. Wu and H.P. Wynn, The convergence of general step-length algorithms for regular optimum design criteria. *Ann. Statist.* **6** (1978) 1273-1285.

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

[37] H.P. Wynn, Results in the theory and construction of *D*-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **34** (1972) 133-147.

[38] J. Zowe and S. Kurcyusz, Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.* **5** (1979) 49-62.