

LABORATOIRE



INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

ON THE SEQUENTIAL CONSTRUCTION OF OPTIMUM BOUNDED DESIGNS

Luc Pronzato

Projet TOpModel

Rapport de recherche
ISRN I3S/RR-2004-20-FR

Août 2004

On the sequential construction of optimum bounded designs

Luc Pronzato

*Laboratoire I3S, CNRS-UNSA, Bât. Euclide, Les Algorithmes,
2000 route des Lucioles, BP 121, 06903 Sophia Antipolis cedex, France*

Abstract

We consider a parameter estimation problem with independent observations, where the experimental conditions X_k form a sequence of independent random variables, distributed with a probability measure μ , and observed sequentially. The length of the sequence (X_k) is N but only $n < N$ observations can be made, with n a proportion of N ($n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$). For $\phi(\cdot)$ a regular optimality criterion, function of the Fisher information matrix, we show that a ϕ -optimum sequential strategy (by sequential we mean that the decision to observe Y_k or not is taken immediately after X_k is known) can be obtained from the construction of a ϕ -optimum constrained design measure $\xi_\alpha^* \leq \mu/\alpha$, a problem for which we derive an equivalence theorem of the minimax form. We show that a slight modification of this sequential procedure generates an empirical design measure ξ_k such that $\phi(\xi_k)$ tends to $\phi(\xi_\alpha^*)$, μ -almost surely, as k tends to infinity. It is therefore possible to sample asymptotically optimally (as if from ξ_α^*), without having to determine ξ_α^* beforehand, or even without knowing μ *a priori*. Some possible applications are indicated.

Key words: sequential design, sampling, constrained design measure, bounded design, equivalence theorem.

1 Introduction

We consider a parameter estimation problem, with θ the vector of parameters to be estimated and X the experimental variables. We assume that the observations are independent, so that the Fisher information matrix is the sum of rank-one matrices of the type $f(X)f^\top(X)$. For instance, this may correspond to estimation in a regression model, with independent observations

$$Y_k = \eta(\bar{\theta}, X_k) + \epsilon_k, \quad (1)$$

where the ϵ_k 's are independently identically distributed (i.i.d.) with $\mathbb{E}\{\epsilon_k\} = 0$ and $\bar{\theta} \in \Theta$ is the unknown true value of the model parameters to be estimated, with Θ an open subset of \mathbb{R}^d . In this case, we write $f(x) = \partial\eta(\theta, x)/\partial\theta_{\hat{\theta}^0}$, with $\hat{\theta}^0$ a given nominal value for θ ($\hat{\theta}^0$ need not be specified when $\eta(\theta, x)$ is linear in x). Another possible situation is when one observes independent binary responses $Y_k \in \{0, 1\}$, with $\text{Prob}(Y_k = 1|\theta, X_k) = \eta(\theta, X_k)$. In this case we write $f(x) = \partial\eta(\theta, x)/\partial\theta_{\hat{\theta}^0} \{\eta(\hat{\theta}^0, X_k)[1 - \eta(\hat{\theta}^0, X_k)]\}^{-1/2}$.

We shall assume that $f(\cdot)$ is continuous in x on \mathcal{X} . We consider design criteria $\Phi(\cdot)$ that are functions of the information matrix \mathbf{M} , with $\Phi[\mathbf{M}]$ to be maximized, and generalized designs ξ that are probability distributions on the set

Email address: pronzato@i3s.unice.fr (Luc Pronzato).

URL: <http://www.i3s.unice.fr/pronzato/> (Luc Pronzato).

\mathcal{X} . We denote Ξ the set of such designs,

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} f(x)f^\top(x)\xi(dx),$$

and $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$, $\xi \in \Xi$. We give in Appendix A a list of assumptions on Φ that will be used throughout the paper. We shall always assume that Φ is strictly concave (**H $_{\Phi}$ 1**), differentiable (**H $_{\Phi}$ 2**) and increasing (**H $_{\Phi}$ 3**). The assumptions are discussed in the same appendix.

We consider the situation where the experimental conditions $X_k \in \mathcal{X}$ form a sequence of i.i.d. variables, independent of (ϵ_k) , of length N . Only $n < N$ observations can be made. We focus on the sequential problem, where, as soon as a value X_k becomes available, we must decide whether to observe Y_k or not. Notice the difference with a standard experimental design problem where the X_k 's can be *chosen*: here we can only decide to accept or reject X_k . The paper is rather theoretically oriented, but many practical decision problems could be formulated in this way. For instance, for some experiments in nuclear physics events are selected according to the energy dissipated in a detector (C.E.R.N., 1994), but the selection could be based on the information content of the event, measured by its contribution to the information matrix; in phase-I clinical trials, the volunteers could be selected according to their size, weight, age, etc., all variables to be used to build a model for the tolerance dose. Other possible developments are given in Section 5. Notice that the case $\dim(\theta) = 1$ corresponds to a variant of the secretary problem, see (Pronzato, 2001a).

Let μ denote the probability measure of the X_k 's, with $(\mathcal{X}, \mu, \mathcal{F})$ a probability space over the σ -field \mathcal{F} of subsets of \mathcal{X} , $\int_{\mathcal{X}} \mu(dx) = 1$, and (\mathcal{F}_n) denote the family of σ -algebra generated by (X_k) , $0 \leq k \leq n$. In particular, we shall con-

sider, but not restrict our attention to, the case when μ is atomless (for any $\Delta\mathcal{X}$ exists $\Delta\mathcal{X}' \subset \Delta\mathcal{X}$ such that $\int_{\Delta\mathcal{X}'} \mu(dx) < \int_{\Delta\mathcal{X}} \mu(dx)$, with measures absolutely continuous w.r.t. the Lebesgue measure as a special case). We assume that μ is such that

$$\mathbf{M}(\mu) = \mathbf{E}\{f(X_1)f^\top(X_1)\} = \int_{\mathcal{X}} f(x)f^\top(x)\mu(dx)$$

exists, with $-\infty < \phi(\mu) < \infty$; a list of additional assumptions on μ is given in Appendix A.

Let (u_k) denote the decision sequence: $u_k = 1$ if we decide to observe Y_k , with experimental conditions X_k , and $u_k = 0$ otherwise, with, for any admissible policy,

$$u_j \in \mathcal{U}_j \subseteq \{0, 1\}, \quad j = 1, \dots, N, \quad \sum_{j=1}^N u_j = n. \quad (2)$$

(Note that X_k is known when u_k is chosen.) The associated information matrix $\mathbf{M}_{N,n}$ is given by $\mathbf{M}_{N,n} = \sum_{k=1}^N u_k f(X_k)f^\top(X_k)$. The sequential problem for N finite corresponds to

$$\text{maximise } \mathbf{E}\{\Phi(\mathbf{M}_{N,n}/n)\} \quad (3)$$

with respect to (u_j) satisfying (2), the expectation $\mathbf{E}\{\cdot\}$ being with respect to the product measure $\mu^{\otimes N}$ of X_1, \dots, X_N (we shall see that the concavity and increasing properties of Φ and $\phi(\mu) < \infty$ imply that $\mathbf{E}\{\Phi(\mathbf{M}_{N,n}/n)\} < \infty$ for any $N, n, 0 < n \leq N$, and any \mathcal{F}_j -measurable sequence (u_j)).

In the multidimensional case $d > 1$, in general, the optimal solution cannot be obtained in closed form. Open-loop feedback-optimal control is used in (Pronzato, 1999) and a heuristic one-step ahead decision rule in (Pronzato, 2001b), without any result on the asymptotic performance of these strategies.

The asymptotics considered in this paper will only concern the case $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$. The fact that n tends to infinity at the same speed as N means that we shall obtain asymptotic performances that are achieved μ -almost surely (a.s.), contrary to (Pronzato, 2001a) which concerns expected performances. The construction of a ϕ -optimum constrained design measure $\xi_\alpha^* \leq \mu/\alpha$ is considered in Section 2, where we derive an equivalence theorem of the minimax form for ϕ -optimality. In Section 2.2 we show that this construction can be used to obtain an asymptotically optimal sequential strategy. Conversely, we show in Section 3 that a slight modification of this sequential strategy yields a procedure that generates a design measure ξ_k such that $\phi(\xi_k) \rightarrow \phi(\xi_\alpha^*)$, μ -a.s., without requiring the prior knowledge of the measure μ . Illustrative examples are given in Section 4. Section 5 gives some concluding remarks and suggests some extensions.

2 Constrained design measures and asymptotically optimum sequential strategies

Our candidate strategy will be obtained through the construction of a constrained ϕ -optimum design measure (or optimum bounded design, or submeasure). We first give some results about constrained measures.

2.1 Optimum constrained design measures

Let $\mathcal{D}(\mu, \alpha)$ denote the set of admissible measures satisfying

$$\xi(dx) \leq \mu(dx)/\alpha, \quad (5)$$

with $\int_{\mathcal{X}} \xi(dx) = 1$, and ξ_α^* denote a ϕ -optimum constrained design measure: ξ_α^* maximises $\phi(\xi)$, $\xi \in \mathcal{D}(\mu, \alpha)$, with $\mathbf{M}(\xi) = \int_{\mathcal{X}} f(x)f^\top(x)\xi(dx)$. When $\Phi(\cdot)$ is strictly concave ($\mathbf{H}_\Phi \mathbf{1}$), $\mathbf{M}(\xi_\alpha^*)$ is unique although ξ_α^* is not necessarily. We shall denote

$$\phi_\alpha^* = \phi(\xi_\alpha^*) = \Phi[\mathbf{M}(\xi_\alpha^*)].$$

($\mathbf{H}_\mu \mathbf{1}$) implies $\phi_\alpha^* \geq \phi(\mu) > -\infty$ and ($\mathbf{H}_\Phi \mathbf{3}$) implies $\phi_\alpha^* \leq \phi(\mu/\alpha) < \infty$.

The main result (Wynn, 1982; Sahm and Schwabe, 2001), presented in the following theorem, states that \mathcal{X} can be partitioned into three subsets $\mathcal{X}_{1,\alpha}^*$, $\mathcal{X}_{2,\alpha}^*$ and $\mathcal{X}_{3,\alpha}^* = \mathcal{X} \setminus (\mathcal{X}_{1,\alpha}^* \cup \mathcal{X}_{2,\alpha}^*)$, with $\xi_\alpha^* = 0$ on $\mathcal{X}_{1,\alpha}^*$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{X}_{2,\alpha}^*$ and the directional derivative $F(\xi_\alpha^*, x)$ (see Appendix A) constant on $\mathcal{X}_{3,\alpha}^*$.

Theorem 1 *The following statements are equivalent:*

- (i) ξ_α^* is a ϕ -optimum constrained design measure;
- (ii) there exists a number c such that $F(\xi_\alpha^*, x) \geq c$ for ξ_α^* -almost all x and $F(\xi_\alpha^*, x) \leq c$ for $(\mu - \alpha\xi_\alpha^*)$ -almost all x ;
- (iii) there exist two subsets $\mathcal{X}_{1,\alpha}^*$ and $\mathcal{X}_{2,\alpha}^*$ of \mathcal{X} such that

- $\xi_\alpha^* = 0$ on $\mathcal{X}_{1,\alpha}^*$ and $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{X}_{2,\alpha}^*$,
- $\inf_{x \in \mathcal{X}_{2,\alpha}^*} F(\xi_\alpha^*, x) \geq c \geq \sup_{x \in \mathcal{X}_{1,\alpha}^*} F(\xi_\alpha^*, x)$,
- $F(\xi_\alpha^*, x) = c$ on $\mathcal{X}_{3,\alpha}^* = \mathcal{X} \setminus (\mathcal{X}_{1,\alpha}^* \cup \mathcal{X}_{2,\alpha}^*)$.

When μ is atomless, $\mu(\mathcal{X}_{3,\alpha}^*) = 0$ and ξ_α^* belongs to the following subclass of $\mathcal{D}(\mu, \alpha)$:

$$\mathcal{D}^*(\mu, \alpha) = \{\xi \in \mathcal{D}(\mu, \alpha) / \exists \mathcal{A} \in \mathcal{F}, \xi(\mathcal{A}) = \mu(\mathcal{A})/\alpha, \xi(\mathcal{X} \setminus \mathcal{A}) = 0\},$$

see (Wynn, 1982; Fedorov, 1989; Fedorov and Hackl, 1997). The condition (ii) of Theorem 1 is then formulated as $F_\Phi(\xi_\alpha^*, x)$ separating the two sets \mathcal{X}_α^* and $\mathcal{X} \setminus \mathcal{X}_\alpha^*$, with

$$\mathcal{X}_\alpha^* = \text{supp } \xi_\alpha^* = \{x \in \mathcal{X} / \xi_\alpha^*(x) > 0\}.$$

Also, we have in that case $\int_{\mathcal{X}_\alpha^*} F_\Phi(\xi_\alpha^*, x) \mu(dx) = \int_{\mathcal{X}} F_\Phi(\xi_\alpha^*, x) \xi_\alpha^*(dx) = 0$, see Fedorov (1989); Fedorov and Hackl (1997). Iterative algorithms of the exchange type for the construction of an optimal constrained measure ξ_α^* are presented in the same references.

As can be seen from Theorem 1, the situation is more complicated when μ has discrete components. For a given ξ , consider the random variable $F_\Phi(\xi, X_1)$ and let $\mathbb{F}_\xi(\cdot)$ denote the corresponding distribution function,

$$\mathbb{F}_\xi(s) = \mu\{x / F_\Phi(\xi, x) \leq s\}. \quad (6)$$

The presence of atoms raises difficulty when there exists no s such that $\mathbb{F}_{\xi_\alpha^*}(s) = 1 - \alpha$. Then, defining $c_\alpha(\xi)$ as

$$c_\alpha(\xi) = \min\{s / \mathbb{F}_\xi(s) \geq 1 - \alpha\} \quad (7)$$

and

$$\begin{aligned} \mathcal{X}_{1,\alpha}(\xi) &= \{x / F_\Phi(\xi, x) < c_\alpha(\xi)\}, \\ \mathcal{X}_{2,\alpha}(\xi) &= \{x / F_\Phi(\xi, x) > c_\alpha(\xi)\}, \\ \mathcal{X}_{3,\alpha}(\xi) &= \{x / F_\Phi(\xi, x) = c_\alpha(\xi)\}, \end{aligned} \quad (8)$$

we get $\mathcal{X}_{j,\alpha}^* = \mathcal{X}_{j,\alpha}(\xi_\alpha^*)$, $j = 1, 2, 3$, and $c_\alpha(\xi_\alpha^*)$ is the constant c of Theorem 1.

Consider the following transformation, which plays a key role in the definition of ϕ -optimum constrained design measures and for constructing asymptotically optimum sequential strategies, as we shall see in Sections 2.2 and 3.

$$T_{\Phi,\alpha} : \xi \in \Xi \rightarrow T_{\Phi,\alpha}(\xi) \in \mathcal{D}(\mu, \alpha), \quad T_{\Phi,\alpha}(\xi) = \begin{cases} \mu/\alpha \text{ on } \mathcal{X}_{2,\alpha}(\xi), \\ \frac{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi)]}{\mu[\mathcal{X}_{3,\alpha}(\xi)]} \mu/\alpha \text{ on } \mathcal{X}_{3,\alpha}(\xi), \\ 0 \text{ on } \mathcal{X}_{1,\alpha}(\xi). \end{cases}$$

Next theorem adds equivalent statements to those in Theorem 1, closer to the standard formulation of the Kiefer-Wolfowitz (1960) Equivalence Theorem.

Theorem 2 *The following statements are equivalent:*

- (i) ξ_α^* is a ϕ -optimum constrained design measure;
- (ii) $F_\Phi[\xi_\alpha^*; T_{\Phi,\alpha}(\xi_\alpha^*)] = 0$;
- (iii) ξ_α^* minimises $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)]$, $\xi \in \mathcal{D}(\mu, \alpha)$;
- (iv) ξ_α^* minimises $\max_{\nu \in \mathcal{D}(\mu, \alpha)} F_\Phi(\xi; \nu)$, $\xi \in \mathcal{D}(\mu, \alpha)$.

Proof. From Theorem 1, (i) gives $T_{\Phi,\alpha}(\xi_\alpha^*) = \xi_\alpha^*$ which implies (ii). Take any $\xi \in \Xi$ and denote ξ_α^* an optimum constrained design measure. From the definition of $T_{\Phi,\alpha}$,

$$F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq F_\Phi(\xi; \nu) \text{ for any } \nu \in \mathcal{D}(\mu, \alpha) \quad (9)$$

and thus, $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq F_\Phi(\xi; \xi_\alpha^*)$. Concavity of Φ then gives

$$\forall \xi \in \Xi, \quad F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq \phi_\alpha^* - \phi(\xi). \quad (10)$$

Since $\phi_\alpha^* \geq \phi(\xi)$, $\xi \in \mathcal{D}(\mu, \alpha)$, (ii) implies (i). Equivalence between (i) and (iii) is obvious from (10); (9) implies the equivalence between (iii) and (iv). \square

Remark 3 *Theorem 2 suggests algorithms for constructing optimum constrained design measures, either (i) by successive applications of the transformation $T_{\Phi, \alpha}$, that is, using a recurrence of the form $\xi^{k+1} = T_{\Phi, \alpha}(\xi^k)$, or (ii) by repeating steps of suitable length γ_k in the direction of $T_{\Phi, \alpha}(\xi)$, that is, $\xi^{k+1} = (1 - \gamma_k)\xi^k + \gamma_k T_{\Phi, \alpha}(\xi^k)$. While (i) may produce non-converging oscillations, (ii) corresponds to a more classical steepest ascent algorithm, in the same line as the procedure suggested by Fedorov (1989). (Note, however, that the true steepest ascent direction is obtained by using the directional derivative G_Φ instead of F_Φ , see Molchanov and Zuyev (2001).) Since the main objective of the paper is the construction of a procedure that samples asymptotically from ξ_α^* without having to construct ξ_α^* , we do not develop this algorithmic construction of ξ_α^* , which may, however, deserve further study.*

The optimum criterion value ϕ_α^* , considered as a function of α , satisfies a Lipschitz condition, as shown below. This property will be used in Section 2.2 and 3 to obtain asymptotically optimum strategies.

Theorem 4 *For any α, β in $(0, 1)$, the associated ϕ -optimum constrained design measures ξ_α^* and ξ_β^* satisfy*

$$|\phi_\alpha^* - \phi_\beta^*| \leq |\alpha - \beta| \max \left[\frac{|c_\alpha(\xi_\alpha^*)|}{\beta}, \frac{|c_\beta(\xi_\beta^*)|}{\alpha} \right],$$

with $c_\alpha(\xi)$ defined by (7).

Proof. Assume that $\beta \leq \alpha$, which gives $\phi_\beta^* \geq \phi_\alpha^*$.

Suppose first that $c_\alpha(\xi_\alpha^*) = c_\beta(\xi_\alpha^*)$. This implies $\mathcal{X}_{j,\alpha}(\xi_\alpha^*) = \mathcal{X}_{j,\beta}(\xi_\alpha^*)$ for $j = 2, 3$, and

$$\begin{aligned} \mathbf{M}(\xi_\alpha^*) &= \frac{\beta}{\alpha} \mathbf{M}[T_{\Phi,\beta}(\xi_\alpha^*)] + \left\{ \frac{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi_\alpha^*)]}{\alpha \mu[\mathcal{X}_{3,\alpha}(\xi_\alpha^*)]} - \frac{\beta - \mu[\mathcal{X}_{2,\beta}(\xi_\alpha^*)]}{\alpha \mu[\mathcal{X}_{3,\beta}(\xi_\alpha^*)]} \right\} \\ &\quad \times \int_{\mathcal{X}_{3,\alpha}(\xi_\alpha^*)} f(x) f^\top(x) \mu(dx) \\ &= \frac{\beta}{\alpha} \mathbf{M}[T_{\Phi,\beta}(\xi_\alpha^*)] + \frac{\alpha - \beta}{\alpha \mu[\mathcal{X}_{3,\alpha}(\xi_\alpha^*)]} \int_{\mathcal{X}_{3,\alpha}(\xi_\alpha^*)} f(x) f^\top(x) \mu(dx). \end{aligned} \quad (11)$$

Suppose now that $c_\alpha(\xi_\alpha^*) < c_\beta(\xi_\alpha^*)$ and define $\mathcal{V}_{\alpha,\beta} = \{x / c_\alpha(\xi_\alpha^*) < F_\Phi(\xi_\alpha^*, x) < c_\beta(\xi_\alpha^*)\}$. We obtain

$$\begin{aligned} \mathbf{M}(\xi_\alpha^*) &= \frac{\beta}{\alpha} \mathbf{M}[T_{\Phi,\beta}(\xi_\alpha^*)] + \frac{1}{\alpha} \left\{ 1 - \frac{\beta - \mu[\mathcal{X}_{2,\beta}(\xi_\alpha^*)]}{\mu[\mathcal{X}_{3,\beta}(\xi_\alpha^*)]} \right\} \int_{\mathcal{X}_{3,\beta}(\xi_\alpha^*)} f(x) f^\top(x) \mu(dx) \\ &\quad + \frac{1}{\alpha} \int_{\mathcal{V}_{\alpha,\beta}} f(x) f^\top(x) \mu(dx) \\ &\quad + \frac{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi_\alpha^*)]}{\alpha \mu[\mathcal{X}_{3,\alpha}(\xi_\alpha^*)]} \int_{\mathcal{X}_{3,\alpha}(\xi_\alpha^*)} f(x) f^\top(x) \mu(dx). \end{aligned} \quad (12)$$

$F_\Phi(\xi_\alpha^*, x) \geq c_\alpha(\xi_\alpha^*)$ gives $G_\Phi(\xi_\alpha^*, x) \geq G_\Phi(\xi_\alpha^*; \xi_\alpha^*) + c_\alpha(\xi_\alpha^*)$, so that, in both cases, (11) and (12), we get

$$G_\Phi(\xi_\alpha^*; \xi_\alpha^*) \geq \frac{\beta}{\alpha} G_\Phi[\xi_\alpha^*; T_{\Phi,\beta}(\xi_\alpha^*)] + \frac{\alpha - \beta}{\alpha} [G_\Phi(\xi_\alpha^*; \xi_\alpha^*) + c_\alpha(\xi_\alpha^*)],$$

that is

$$F_\Phi[\xi_\alpha^*; T_{\Phi,\beta}(\xi_\alpha^*)] \leq -\frac{\alpha - \beta}{\beta} c_\alpha(\xi_\alpha^*).$$

Concavity of $\Phi(\cdot)$ and (9) imply

$$\phi_\beta^* \leq \phi_\alpha^* + F_\Phi(\xi_\alpha^*; \xi_\beta^*) \leq \phi_\alpha^* + F_\Phi[\xi_\alpha^*; T_{\Phi,\beta}(\xi_\alpha^*)].$$

Therefore,

$$0 \leq \phi_\beta^* - \phi_\alpha^* \leq -\frac{\alpha - \beta}{\beta} c_\alpha(\xi_\alpha^*).$$

The case $\beta > \alpha$ can be treated similarly. \square

Remark 5 *Developments similar to those used in the proof of Theorem 4 give for $0 < \beta < \alpha < 1$*

$$-\frac{c_\beta(\xi_\beta^*)}{\alpha} \leq \frac{\phi_\beta^* - \phi_\alpha^*}{\alpha - \beta} \leq -\frac{c_\alpha(\xi_\alpha^*)}{\beta}$$

so that ϕ_α^ is differentiable at α when $c_\alpha(\xi_\alpha^*)$ is continuous at the same point (see Section 4 for an example).*

2.2 An asymptotically optimum sequential strategy

For any strategy $\mathcal{S}_{N,n}$ used to solve (3), with $0 < n \leq N$, we denote $\Psi(\mathcal{S}_{N,n}) = \Phi(\mathbf{M}_{N,n}/n)$. We shall follow the same line as in (Albright and Derman, 1972), which concerns the case $d = 1$, and use as a benchmark the infeasible, but better-than-optimal, *non sequential* strategy $\mathcal{S}_{N,n}^*$, obtained by selecting the n design points X_{k_1}, \dots, X_{k_n} that maximise $\Phi(\mathbf{M}_{N,n}/n)$ after the N points X_1, \dots, X_N have been observed. $\mathcal{S}_{N,n}^*$ is thus a ϕ -optimum design algorithm that generates an exact n -point ϕ -optimum design in the finite design space $\{X_1, \dots, X_N\}$. Obviously, for any N, n and any strategy $\mathcal{S}_{N,n}$, sequential or not,

$$\Psi(\mathcal{S}_{N,n}) \leq \Psi(\mathcal{S}_{N,n}^*). \quad (13)$$

Also, $\mathcal{S}_{N,n}^*$ satisfies

$$\forall \alpha \in (0, 1), \limsup_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*) \leq \phi(\xi_\alpha^*), \quad \mu\text{-a.s.} \quad (14)$$

with $\xi_\alpha^* \leq \mu/\alpha$ an optimum constrained design measure. Consider now the sequential strategy $\mathcal{S}_{N,n}^\epsilon$ defined by

$$\mathcal{S}_{N,n}^\epsilon : \left\{ \begin{array}{l} \text{if } a_k = n, \text{ reject } X_k; \\ \\ \text{if } N - k + 1 \leq n - a_k, \text{ accept } X_k; \\ \\ \text{otherwise, } \left\{ \begin{array}{l} \text{accept } X_k \text{ if } X_k \in \mathcal{X}_{2,\alpha+\epsilon}^*; \\ \\ \text{accept } X_k \text{ with probability} \\ \\ P_{3,\alpha+\epsilon}^* = (\alpha + \epsilon - \mu_{2,\alpha+\epsilon}^*)/\mu_{3,\alpha+\epsilon}^* \text{ if } X_k \in \mathcal{X}_{3,\alpha+\epsilon}^*; \\ \\ \text{reject } X_k \text{ if } X_k \in \mathcal{X}_{1,\alpha+\epsilon}^*; \end{array} \right. \end{array} \right. \quad (15)$$

where $0 \leq \epsilon < 1 - \alpha$, a_k is given by (4), $\mathcal{X}_{j,\alpha+\epsilon}^* = \mathcal{X}_{j,\alpha+\epsilon}(\xi_{\alpha+\epsilon}^*)$ and $\mu_{j,\alpha+\epsilon}^* = \mu(\mathcal{X}_{\alpha+\epsilon}^*)$, $j = 1, 2, 3$, see (8) and Theorem 1. In fact, we shall consider the following, simpler, strategy, that may accept less than n points:

$$\bar{\mathcal{S}}_{N,n}^\epsilon : \left\{ \begin{array}{l} \text{if } a_k = n, \text{ reject } X_k; \\ \\ \text{otherwise, } \left\{ \begin{array}{l} \text{accept } X_k \text{ if } X_k \in \mathcal{X}_{2,\alpha+\epsilon}^*; \\ \\ \text{accept } X_k \text{ with probability} \\ \\ P_{3,\alpha+\epsilon}^* = (\alpha + \epsilon - \mu_{2,\alpha+\epsilon}^*)/\mu_{3,\alpha+\epsilon}^* \text{ if } X_k \in \mathcal{X}_{3,\alpha+\epsilon}^*; \\ \\ \text{reject } X_k \text{ if } X_k \in \mathcal{X}_{1,\alpha+\epsilon}^*. \end{array} \right. \end{array} \right.$$

Since there may be less than n terms in the information matrix obtained with $\bar{\mathcal{S}}_{N,n}^\epsilon$ and $\Phi(\cdot)$ is increasing, we have

$$\Psi(\bar{\mathcal{S}}_{N,n}^\epsilon) \leq \Psi(\mathcal{S}_{N,n}^\epsilon).$$

Denote by $N_{\alpha+\epsilon}$ the number of points that would have been accepted if ignoring the constraint $a_k \leq n$ (notice that these points are sampled from $T_{\Phi, \alpha+\epsilon}(\xi_{\alpha+\epsilon}^*) = \xi_{\alpha+\epsilon}^*$). It satisfies $N_{\alpha+\epsilon}/N \rightarrow \alpha + \epsilon$, μ -a.s. as N tends to infinity. Take $n = \lfloor \alpha N \rfloor$ and let N tend to infinity. The probability that $N_{\alpha+\epsilon} \leq n$ infinitely often (i.o.) is zero. Therefore, asymptotically, $\bar{\mathcal{S}}_{N,n}^\epsilon$ stops after $n = \lfloor \alpha N \rfloor$ samples from $\xi_{\alpha+\epsilon}^*$ have been collected, $n < N_{\alpha+\epsilon}$, and

$$\Psi(\bar{\mathcal{S}}_{N, \lfloor \alpha N \rfloor}^\epsilon) \rightarrow \phi_{\alpha+\epsilon}^*, \mu\text{-a.s.}$$

which gives

$$\liminf_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^\epsilon) \geq \phi_{\alpha+\epsilon}^* \mu\text{-a.s.}$$

We may then let ϵ tend to zero and use the continuity of ϕ_α^* with respect to α , see Theorem 4, to obtain the following property directly from (13) and (14).

Theorem 6 *The strategy $\mathcal{S}_{N,n}^0$ defined by (15) is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $N \rightarrow \infty$, $0 < \alpha < 1$:*

$$\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^0) = \lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*) = \phi(\xi_\alpha^*), \mu\text{-a.s.}$$

where $\xi_\alpha^* \leq \mu/\alpha$ is a ϕ -optimum constrained measure and $\mathcal{S}_{N,n}^*$ is the optimum non sequential strategy that selects n points among N after these N points have been observed.

The strategy $\mathcal{S}_{N,n}^0$ is asymptotically optimum for the problem (3) when $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$ and $N \rightarrow \infty$. However, it requires the construction of the sets $\mathcal{X}_{j\alpha}^*$, $j = 1, 2, 3$, and thus of the ϕ -optimum constrained design measure $\xi_\alpha^* \leq \mu/\alpha$. We shall see in Section 3 that it is possible to avoid this construction while preserving asymptotic optimality.

2.3 A concentration inequality for $\Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*)$

Theorem 6 does not say anything about the finite sample behavior of $\Psi(\mathcal{S}_{N,n}^*)$. The following result on $\mathbf{E}\{\Psi(\mathcal{S}_{N,n}^*)\}$ holds under very general conditions. The proof is given in Appendix B.

Lemma 7 *For a concave criterion $\Phi(\cdot)$, the non sequential strategy $\mathcal{S}_{N,n}^*$ (that is, a ϕ -optimum algorithm for an exact design with n point in $\{X_1, \dots, X_N\}$) satisfies*

$$\forall(n, N), 0 < n \leq N, \mathbf{E}\{\Psi(\mathcal{S}_{N,n}^*)\} \leq \phi(\xi_{n/N}^*), \quad (16)$$

with $\xi_{n/N}^*$ a ϕ -optimum constrained design measure in $\mathcal{D}(\mu, n/N)$.

H $_{\mu}$ 1 and **H $_{\Phi}$ 3** imply $\phi(\xi) < \infty$ for any design measure $\xi \leq \mu/\alpha$, α given in $(0, 1)$, so that Lemma 7 implies that for any (n, N) , $0 < n \leq N$, and any strategy $\mathcal{S}_{N,n}$:

$$\mathbf{E}\{\Phi[\mathbf{M}_{N,n}/n]\} = \mathbf{E}\{\Psi(\mathcal{S}_{N,n})\} \leq \mathbf{E}\{\Psi(\mathcal{S}_{N,n}^*)\} \leq \phi(\xi_{n/N}^*) < \infty. \quad (17)$$

It may be noticed that the upper bound $\phi(\xi_{n/N}^*)$ for the expected performance $\mathbf{E}\{\Psi(\mathcal{S}_{N,n})\}$ is not necessarily achievable when n is finite, $N \rightarrow \infty$ and the strategy $\mathcal{S}_{N,n}$ is sequential, even in the case $d = \dim(\theta) = 1$, see (Pronzato, 2001a).

Under more restrictive assumptions on μ and $\Phi(\cdot)$, a concentration inequality can be obtained for $\Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*)$, to be used in Section 3. The proof is given in Appendix B.

Theorem 8 *Let (α_k) be a sequence in $(0, 1)$, with $\lim_{k \rightarrow \infty} \alpha_k = \alpha \in (0, 1)$.*

Then, under $\mathbf{H}_\Phi\mathbf{1}$ - $\mathbf{H}_\Phi\mathbf{5}$ and $\mathbf{H}_\mu\mathbf{1}$ - $\mathbf{H}_\mu\mathbf{3}$,

$$\text{Prob} \left\{ \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*) \geq \mathbb{E}[\Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*)] + \frac{1}{\lfloor \alpha N \rfloor^\beta} \text{ i.o.} \right\} = 0 \quad (18)$$

for any β such that $0 < \beta < 1/2 - 2/\gamma$, with γ the constant in $\mathbf{H}_\mu\mathbf{2}$.

3 Sampling asymptotically from a constrained measure

Consider the following modification of the strategy $\mathcal{S}_{N,n}^0$ defined by (15): at step k , we simply substitute the sets $\mathcal{X}_{j,\alpha}(\xi_{k-1})$ for $\mathcal{X}_{j,\alpha}^*$, $j = 1, 2, 3$, with $\mathcal{X}_{j,\alpha}(\xi)$ defined by (8) and ξ_{k-1} the empirical measure defined by the a_k design points already selected. We thus define:

$$\mathcal{S}_\alpha(\mu) : \left\{ \begin{array}{l} \text{accept } X_k \text{ if } X_k \in \mathcal{X}_{2,\alpha}(\xi_{k-1}); \\ \text{accept } X_k \text{ with probability} \\ P_k(\alpha) = \{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi_{k-1})]\} / \mu[\mathcal{X}_{3,\alpha}(\xi_{k-1})] \text{ if } X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1}); \\ \text{reject } X_k \text{ if } X_k \in \mathcal{X}_{1,\alpha}(\xi_{k-1}). \end{array} \right. \quad (19)$$

(In practise, the first samples X_k are always accepted until $\mathbf{M}(\xi_{k-1})$ becomes nonsingular, but this initialization has no effect on the asymptotic behavior of $\mathcal{S}_\alpha(\mu)$).

Obviously, the sequence (ϕ_k) , with $\phi_k = \phi(\xi_k)$, is not monotonically increasing, since (i) the step-length $1/(1+a_k)$ when X_k is accepted and ξ_{k-1} updated is predetermined and (ii) X_k is random. While (i) is standard in the construction of optimum designs, (ii) is less common and forms a specific feature of

the context considered here. In order to eliminate the unboundedness case encountered in the dichotomous theorem of Wu and Wynn (1978), which is usually the main issue raised by (i), we introduce the assumption $\mathbf{H}_\mu\mathbf{4}$ on μ , see Appendix A.

Theorem 9 *Under $\mathbf{H}_\Phi\mathbf{1}$ - $\mathbf{H}_\Phi\mathbf{6}$ and $\mathbf{H}_\mu\mathbf{1}$ - $\mathbf{H}_\mu\mathbf{4}$ the empirical measure ξ_k defined by the points accepted by the strategy $\mathcal{S}_\alpha(\mu)$ satisfies*

$$\lim_{k \rightarrow \infty} \phi_k = \phi(\xi_\alpha^*), \mu\text{-a.s.}, \quad (20)$$

with $\xi_\alpha^* \leq \mu/\alpha$ a ϕ -optimum constrained design measure.

The proof is given in Appendix B.

Remark 10 $\mathcal{S}_\alpha(\mu)$ takes a simpler form when μ is atomless: we accept X_k when $\mathbb{F}_{\xi_{k-1}}(X_k) > 1 - \alpha$ and reject X_k otherwise, with $\mathbb{F}_\xi(s)$ given by (6).

The behavior is asymptotically the same if we modify $\mathcal{S}_\alpha(\mu)$ as follows: any $X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1})$ is accepted if $a_k/k < \alpha$ and is rejected otherwise.

Coming back to the original problem, with n points only to be accepted among N , we can adapt α and use, at step k , the strategy $\mathcal{S}_{\alpha_k}(\mu)$, with

$$\alpha_k = \frac{n - a_k}{N - k + 1}, \quad (21)$$

and a_k defined by (4). It coincides with the one-step-ahead rule suggested in (Pronzato, 2001b): when $\alpha_k = 0$, $\mathcal{X}_{2,\alpha_k}(\xi_{k-1})$ is empty, $\mu[\mathcal{X}_{2,\alpha_k}(\xi_{k-1})] = 0$ and X_k is always rejected; when $\alpha_k \geq 1$, $\mathcal{X}_{2,\alpha_k}(\xi_{k-1}) = \mathcal{X}$ and X_k is always accepted. From Theorems 6 and 9, although it is myopic this rule is asymptotically optimal when $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$.

When μ is unknown, we can use $\mathcal{S}_\alpha(\hat{\mu}_k)$, or $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$, at step k , with $\hat{\mu}_k$ the empirical version of μ (or a kernel estimate, or a parametric representation $\mu_{\hat{\beta}_k}$, with $\hat{\beta}_k$ estimated from X_1, \dots, X_k). The estimation of μ does not depend on the strategy that is used (which corresponds to a *separation property* in control theory), and Theorem 9 still holds: we can thus asymptotically sample from ξ_α^* , without constructing ξ_α^* beforehand and even without knowing μ in advance. Illustrative examples are presented in the next section.

Consider finally a nonlinear situation where the information matrix \mathbf{M} depends on the model parameters θ . When θ can be estimated on line, it is natural to replace the nominal value $\hat{\theta}^0$ by some estimate, $\hat{\theta}^k$ at step k . When using least squares estimation in a nonlinear regression problem, consistency and asymptotic normality of $\hat{\theta}^k$ will hold under $\mathbf{H}_\mu 4$ (and additional conditions on higher order derivatives of $\eta(\theta, x)$ with respect to θ and their tail cross product, see Jennrich (1969)), and the sampling strategy $\mathcal{S}_\alpha(\mu)$ will be such that ϕ_k converges μ -a.s. to $\phi[\xi_\alpha^*(\bar{\theta})]$, with $\xi_\alpha^*(\bar{\theta}) \leq \mu/\alpha$ an optimum constrained design measure for the *true value* $\bar{\theta}$ of the model parameters in (1). This is illustrated by Example 13 below.

4 Examples

We take $\Phi(\cdot) = \log \det(\cdot)$ in all the examples below.

Example 11 *We consider the quadratic regression model $\eta(\theta, X) = \theta_0 + \theta_1 X + \theta_2 X^2$. Let μ_n correspond to the normal distribution $\mathcal{N}(0, 1)$ and μ_d be the discrete measure supported at the points $\{-1, -1/2, 0, 1/2, 1\}$ with respective weights $(1/8, 1/4, 1/4, 1/4, 1/8)$. We take $\mu = 0.5\mu_n + 0.5\mu_d$ and $\alpha = 0.1$.*

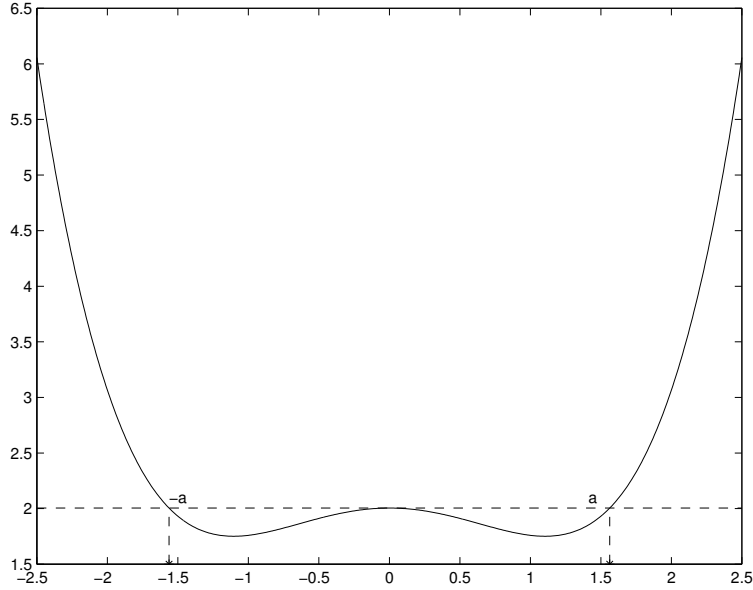


Fig. 1. Sensitivity function $d(\xi_\alpha^*, x)$ for the optimal constrained measure ξ_α^* in Example 11.

Easy calculations show that the ϕ -optimum constrained measure ξ_α^* is equal to μ/α on $\mathcal{X}_\alpha^* = (-\infty, -a] \cup [a, \infty)$, with $a \simeq 1.5625$, and puts the rest of its weight, approximately 0.4091, at zero. Figure 1 presents a plot of the sensitivity function $d(\xi_\alpha^*, x) = F_\Phi(\xi_\alpha^*, x) + 3 = f^\top(x)\mathbf{M}^{-1}(\xi_\alpha^*)f(x)$ and illustrates the optimality of ξ_α^* . Figure 2, left, gives an histogram of the first 1,000 samples accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the X_k 's. The right part of the figure presents ϕ_k as a function of k , with the optimal value $\phi_{0.1}^*$ indicated by the dashed line. (Note the fast convergence of ϕ_k , the large value of N being required only to have enough points for the histogram plot.)

Example 12 We consider again the case of a measure μ having both discrete and continuous components. The response of the regression model is $\eta(\theta, X) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, with $\theta = (\theta_0, \theta_1, \theta_2)$ and X being two-dimensional, $X = (x_1, x_2)$. The continuous component μ_n corresponds to the normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, with \mathbf{I}_2 the 2-dimensional identity matrix, the discrete component μ_d

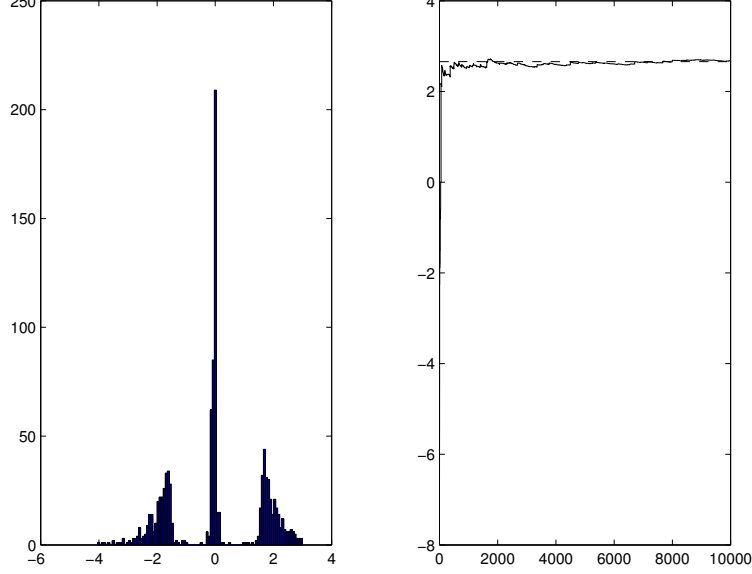


Fig. 2. Left: histogram of the 1,000 first samples X_k accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$ in Example 11. Right: ϕ_k as a function of k ; the value of $\phi_{0.1}^*$ corresponds to the dashed line.

puts weight $1/4$ at each one of the points $(\pm 1, \pm 1)$. Define $\mathcal{B}(a) = \{\mathbf{x} / \|\mathbf{x}\| \geq a\}$. When $\mu = (1/2)(\mu_n + \mu_d)$ the results are as follows.

For $0 < \alpha \leq 1/(2e)$, with $e = \exp(1)$, ξ_α^* coincides with $\mu_n/(2\alpha)$ on $\mathcal{B}(a_\alpha)$ with $a_\alpha = \sqrt{2 \log[1/(2\alpha)]} \geq \sqrt{2}$; $\phi_\alpha^* = 2 \log(m_\alpha)$ with $m_\alpha = [1/(2\alpha)](1 + a_\alpha^2/2) \exp(-a_\alpha^2/2)$.

For $1/(2e) < \alpha \leq 1/(2e) + 1/2$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{B}(\sqrt{2})$ and $\xi_\alpha^* = 2[\alpha - 1/(2e)]\mu/\alpha$ on the four points $(\pm 1, \pm 1)$; $\phi_\alpha^* = 2 \log[1 + 1/(2e\alpha)]$.

For $1/(2e) + 1/2 < \alpha \leq 1$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{B}(b_\alpha)$ with $b_\alpha = \sqrt{2 \log[1/(2\alpha - 1)]} < \sqrt{2}$; $\phi_\alpha^* = 2 \log(m_\alpha)$ with $m_\alpha = [1/(2\alpha)](1 + b_\alpha^2/2) \exp(-b_\alpha^2/2) + 1/(2\alpha)$.

Figure 3 presents ϕ_α^* as a function of α (although it does not appear clearly from the figure, ϕ_α^* tends to infinity when α tends to zero). Note that $c_\alpha(\xi_\alpha)$ is a continuous function of α , so that ϕ_α^* is differentiable with respect to α . (It is not always so. For instance, when μ is the mixture of the two discrete distributions

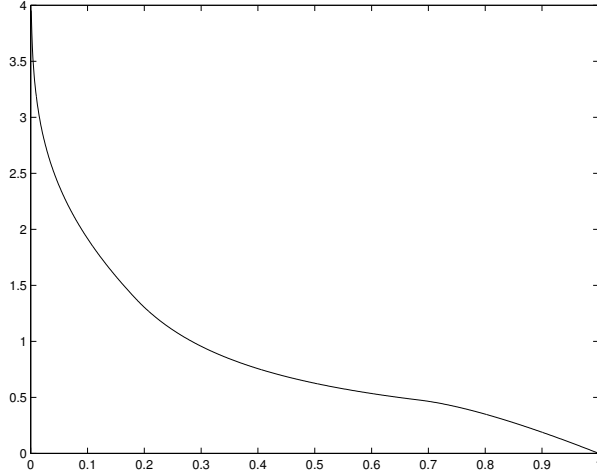


Fig. 3. ϕ_α^* as a function of α in Example 12.

μ_d , as above, and μ'_d that puts weights $1/4$ at each of the points $(\pm 2, \pm 2)$, with $\mu = (1/2)(\mu_d + \mu'_d)$, c_α^* is not continuous, and ϕ_α^* not differentiable, at $\alpha = 1/2$).

We take $N = 10,000$, $n = 2,000$ and use $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$, see (19,21), with $\hat{\mu}_k$ the empirical measure of the X_k 's. Figure 4 (top) presents an histogram of the 2,000 points accepted by $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$, Figure 4 (bottom) gives ϕ_k as a function of k , with the optimal value $\phi_{0.2}^* \simeq 1.3043$ indicated by the dashed line. (Again, notice the fast convergence of ϕ_k , the large value of N being required only to have enough points for the histogram plot.)

Example 13 We consider now the nonlinear model used in (Box and Lucas, 1959), where

$$\eta(\theta, x) = \frac{\theta_1}{\theta_1 - \theta_2} [\exp(-\theta_2 x) - \exp(-\theta_1 x)].$$

We estimate $\theta = (\theta_1, \theta_2)$ by LS, and use at step k the information matrix corresponding to the current estimated value $\hat{\theta}^k$. The numerical values used to generate the observations y_k correspond to $\bar{\theta} = (0.7, 0.2)$ and the standard deviation of the measurement errors is 0.2. (With this value for $\bar{\theta}$, the

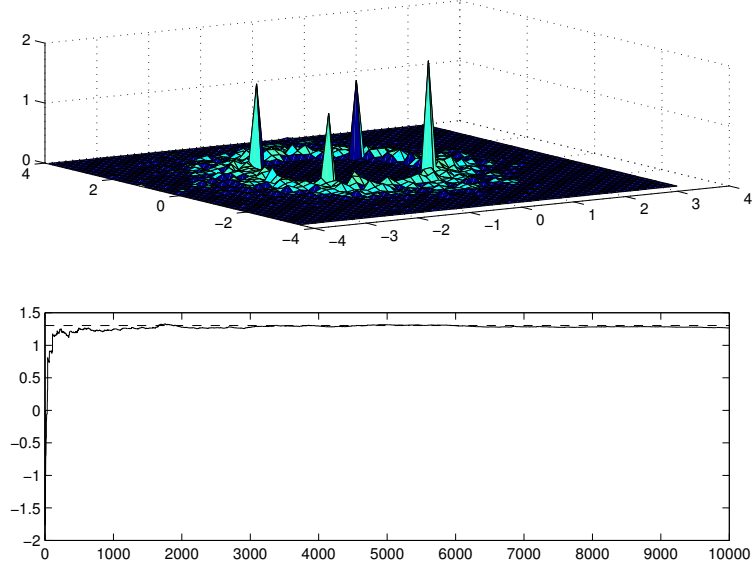


Fig. 4. Top: histogram of the 2,000 points accepted by $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$ in a sequence of 10,000 points; bottom: ϕ_k as a function of k , the value of $\phi_{0.2}^*$ corresponds to the dashed line (Example 12).

D-optimal experiment corresponds to performing the same number of observations at $x \simeq 1.23$ and 6.86 .)

We assume that the experimental variables X_k have a lognormal distribution: $\log(X_k)$ is normally distributed $\mathcal{N}(1, 0.25)$. Figure 5 gives an histogram of the first 2,500 points accepted by $\mathcal{S}_{\alpha}(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the X_k 's, when $\alpha = 0.5$. Easy calculation shows that $\phi_{\alpha}^ = 2\mu$ on $[0, a] \cup [b, \infty)$, with $a \simeq 1.996$ and $b \simeq 3.922$, and $\phi_{\alpha}^* \simeq -2.143$.*

5 Concluding remarks and further developments

Possible extensions of these results, that will be the subject of future work, include the following situations.

First, there are cases where the design variables are not directly observed: an

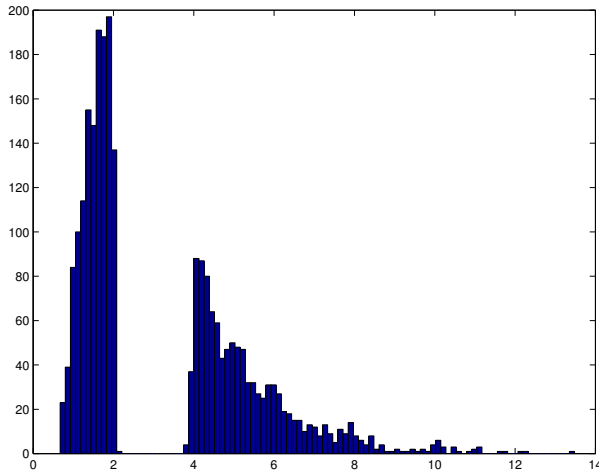


Fig. 5. Histogram of the 2,500 first samples X_k accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$ in Example 13. example is when one observes covariates Z_k , with (\mathcal{Z}_n) the family of σ -algebra generated by (Z_k) , $0 \leq k \leq n$, and the conditional probability measure $\mu(\cdot|Z_k)$ for the experimental conditions X_k is known for any k . A sequential selection strategy for this problem might reveal useful in phase-I clinical trials, where covariates Z_k such as the size, weight and age of volunteers could be used for their selection, in order to build a model of the tolerance dose as a function of (unobserved) pharmacokinetic/pharmacodynamic variables X_k .

Second, applications to parameter estimation in dynamical systems, nonlinear in particular, call for an extension to correlated design variables X_k . A simple example is when $X_k = (U_k, U_{k-1}, \dots, U_{k-m})$, with (U_i) a random input sequence for the system. Note, however, that when the model contains an autoregressive part, that is, when $X_k = (U_k, U_{k-1}, \dots, U_{k-m}, Y_{k-1}, \dots, Y_{k-l})$, the decision not to observe Y_k implies that l future experimental conditions are unknown, which makes the problem much different from the one we considered here and will require specific developments.

In the one dimensional case $d = \dim(\theta) = 1$, asymptotic optimality of a strategy similar to $\mathcal{S}_{\alpha_k}(\mu)$, see (19,21), is proved in (Pronzato, 2001a) for $N \rightarrow \infty$

with n fixed, provided the tail of μ is thin enough (von Mises distribution). Extending Theorem 6 to the situation where n is fixed but $d > 1$ remains an open issue. Note in particular that in this case, although (17) gives an upper bound on the expected performance, the optimal performance achievable by a sequential strategy is unknown (not to speak about the optimal strategy itself). Also, all intermediate situations, between n fixed and $n = \lfloor \alpha N \rfloor$, such as $n = \lfloor \log N \rfloor$, or $\lfloor N^\beta \rfloor$ with $\beta < 1$, etc., are of interest. A possible application concerns the construction of optimum design algorithms. Indeed, classical algorithms for the determination of ξ^* (unconstrained) that maximizes $\phi(\cdot)$ rely on the determination at iteration k of a design point X_k that maximizes $F_\Phi(\xi_{k-1}, x)$ with respect to $x \in \mathcal{X}$, with ξ_{k-1} the current design measure. This (global) maximisation problem may prove cumbersome, especially if \mathcal{X} is high dimensional, so that it is sometimes recommended to accept any X_k such that $F_\Phi(\xi_{k-1}, X_k) > \delta$, with δ some small positive number, see Fedorov and Hackl (1997), p. 49. A consequence of the results presented above is that generating candidates X_k randomly with some suitably chosen measure μ , with an acceptance rule such as (19), will ensure the convergence of ξ_k to ξ^* , provided α tends to zero at a proper speed. Whether or not efficient algorithms can be obtained in this way remains an open issue.

A Appendix (Assumptions and notations)

H_Φ1: Φ is strictly concave and $\Phi(\mathbf{M}) > -\infty$ for non singular \mathbf{M} .

H_Φ2: Φ is linearly differentiable; that is, the directional derivative

$$\mathcal{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\epsilon \rightarrow 0^+} \{\Phi[(1 - \epsilon)\mathbf{M}_1 + \epsilon\mathbf{M}_2] - \Phi(\mathbf{M}_1)\} / \epsilon$$

satisfies $F_\Phi(\xi_1; \xi_2) = \mathcal{F}_\Phi[\mathbf{M}(\xi_1), \mathbf{M}(\xi_2)] = \int_{\mathcal{X}} F_\Phi(\xi_1, x) \xi_2(dx)$ for any ξ_1, ξ_2 with $\phi(\xi_1) > -\infty$, where $F_\Phi(\xi, x) = F_\Phi(\xi; \delta_x)$ and δ_x is the Dirac measure supported at x . Similarly, we denote $\mathcal{G}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\epsilon \rightarrow 0^+} \{\Phi[\mathbf{M}_1 + \epsilon\mathbf{M}_2] - \Phi(\mathbf{M}_1)\} / \epsilon$, $G_\Phi(\xi_1; \xi_2) = \mathcal{G}_\Phi[\mathbf{M}(\xi_1), \mathbf{M}(\xi_2)]$ and $G_\Phi(\xi, x) = G_\Phi(\xi; \delta_x)$.

H_Φ3: Φ is increasing: $\mathbf{M}_2 - \mathbf{M}_1$ non negative definite implies $\Phi(\mathbf{M}_2) \geq \Phi(\mathbf{M}_1)$.

H_Φ4: $\|f\| \leq R$ and $\lambda_{\min}(\mathbf{M}) \geq l > 0$ imply $\mathcal{G}_\Phi(\mathbf{M}, ff^\top) \leq \delta_1(l)R^2$ for some $\delta_1(l) < \infty$.

H_Φ5: One of the following properties is satisfied:

- (i) there exists a function $g_1(\cdot)$, from \mathbb{R}^+ to \mathbb{R} , such that $\lim_{a \rightarrow 1^+} g_1(a) = 0$, and $\Phi(a\mathbf{M}) \leq g_1(a) + \Phi(\mathbf{M})$ for any non-negative definite \mathbf{M} and any $a \geq 1$;
- (ii) there exists a function $g_2(\cdot)$, from \mathbb{R}^+ to \mathbb{R}^+ , such that $\lim_{a \rightarrow 1^+} g_2(a) = 1$, and $\Phi(a\mathbf{M}) \leq g_2(a) \Phi(\mathbf{M})$ for any non-negative definite \mathbf{M} and any $a \geq 1$.

H_Φ6: Φ is two times differentiable. We denote by $\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2)$ the second order directional derivative

$$\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2) = \frac{\partial^2 \Phi[(1 - \gamma)\mathbf{M}_1 + \gamma\mathbf{M}_2]}{\partial \gamma^2} \Big|_{\gamma=0^+}.$$

It satisfies, for any $\|f\| \leq R$ and \mathbf{M} such that $\lambda_{\min}(\mathbf{M}) \geq l > 0$: $\nabla_\Phi^2(\mathbf{M}, ff^\top) > -\delta_0(l) - \delta_2(l)R^4$, for some $\delta_0(l) < \infty$ and $\delta_2(l) < \infty$.

H_μ1: $-\infty < \phi(\mu) < \infty$.

H_μ2: $\text{Prob}\{\|f(X_1)\| > R\} = \mathcal{O}(1/R^\gamma)$ for some $\gamma > 4$. Note that this implies $\mu_4 = \int_{\mathcal{X}} \|f(x)\|^4 \mu(dx) < \infty$, since $\mu_4 = 4 \int_0^\infty R^3 \text{Prob}\{\|f(X_1)\| > R\} dR$, see,

e.g., Shorack (2000, p. 117).

H_μ3: For any given $\alpha \in (0, 1)$, $\lambda_{\min}[\mathbf{M}(\xi_\alpha^*)] = l_\alpha^* > 0$.

H_μ4: For any given $\alpha \in (0, 1)$, there exists $\underline{l} > 0$ and $\epsilon \in (0, \alpha)$ such that for any design measure $\xi \leq \mu/(\alpha - \epsilon)$, $\lambda_{\min}[\mathbf{M}(\xi)] > \underline{l}$.

Discussion of the assumptions.

In the case of D -optimality, where $\Phi(\mathbf{M}) = \log \det \mathbf{M}$, we have $F_\Phi(\xi, x) = f^\top(x)\mathbf{M}^{-1}(\xi)f(x) - d$, with $d = \dim(\theta)$, and $G_\Phi(\xi, x) = f^\top(x)\mathbf{M}^{-1}(\xi)f(x)$.

H_Φ1-H_Φ4 are satisfied, with $\delta_1(l) = 1/l$ in **H_Φ4**. **H_Φ5(i)** is satisfied with $g_1(a) = d \log a$. (More generally, note that concavity implies $\Phi(a\mathbf{M}) = \Phi[\mathbf{M} + (a - 1)\mathbf{M}] \leq \Phi(\mathbf{M}) + (a - 1)\mathcal{G}_\Phi(\mathbf{M}, \mathbf{M})$ so that **H_Φ5(i)** is satisfied when $\mathcal{G}_\Phi(\mathbf{M}, \mathbf{M})$ is bounded for any non-negative definite \mathbf{M} .) Concerning **H_Φ6**, we have $\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2) = -\text{trace}\{[\mathbf{M}_1^{-1}(\mathbf{M}_2 - \mathbf{M}_1)]^2\}$ so that $\nabla_\Phi^2(\mathbf{M}, ff^\top) = -(f^\top \mathbf{M}^{-1} f)^2 + 2(f^\top \mathbf{M}^{-1} f) - d$ and **H_Φ6** is satisfied with $\delta_0 = d$, $\delta_2(l) = 1/l^2$.

Take now $\Phi(\mathbf{M}) = -\text{trace}(\mathbf{A}\mathbf{M}^{-p})$, with p a positive integer and \mathbf{A} a positive definite matrix (Kiefer's Φ_p -class of optimality criteria). We can assume that $\mathbf{A} = \mathbf{I}_d$, the d -dimensional identity matrix, by a linear transformation on \mathcal{X} . We have $\mathcal{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = p \text{trace}[\mathbf{M}_1^{-(p+1)}(\mathbf{M}_2 - \mathbf{M}_1)]$ and $\mathcal{G}_\Phi(\mathbf{M}, ff^\top) = pf^\top \mathbf{M}^{-(p+1)} f$. **H_Φ1-H_Φ4** are satisfied, with $\delta_1(l) = p/l^{p+1}$ in **H_Φ4**. **H_Φ5(ii)** is satisfied with $g_2(a) = a^{-d}$. We also have $\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2) = -p \sum_{a+b=p+2, a, b \geq 1} \text{trace}[\mathbf{M}_1^{-a}(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{M}_1^{-b}(\mathbf{M}_2 - \mathbf{M}_1)]$, see Wu and Wynn (1978), and easy calculation gives $\nabla_\Phi^2(\mathbf{M}, ff^\top) = -p \sum_{a+b=p+2, a, b \geq 1} [(f^\top \mathbf{M}^{-b} f)(f^\top \mathbf{M}^{-a} f)] + 2p(p+1)f^\top \mathbf{M}^{-(p+1)} f - p(p+1) \text{trace} \mathbf{M}^{-p}$. Therefore, **H_Φ6** is satisfied with $\delta_0(l) = p(p+1)/l^p$, $\delta_2(l) = p/l^{p+2}$.

H_μ1-H_μ4 are satisfied for instance when: (i) the functions $f(x)$ are linearly

independent on any open subset of \mathcal{X} with finite Lebesgue measure, (ii) μ has a component absolutely continuous with respect to the Lebesgue measure, with density φ , (iii) the mass of this component of μ is larger than $1 - \alpha + \epsilon$ (that is, the mass of the discrete components is less than $\alpha - \epsilon$), and (iv) \mathcal{X} is compact or $\varphi(x)$ is exponentially decreasing when $\|x\| \rightarrow \infty$.

B Appendix (Proofs)

Proof of Lemma 7. The strategy $\mathcal{S}_{N,n}^*$ satisfies $\Psi(\mathcal{S}_{N,n}^*) = \Phi(\mathbf{M}_{N,n}^*)$ with

$$\mathbf{M}_{N,n}^* = \frac{1}{n} \sum_{i=1}^N f(X_i) f^\top(X_i) J_n[X_j(j=1, \dots, N, j \neq i), X_i],$$

where $J_n \in \{0, 1\}$ defines the decision function (the decision to accept X_i depends on the rest of the sequence). Therefore,

$$\mathbf{E}\{\mathbf{M}_{N,n}^*\} = \frac{1}{n} \int_{\mathcal{X}^N} \left\{ \sum_{i=1}^N f(x_i) f^\top(x_i) J_n[x_j(j=1, \dots, N, j \neq i), x_i] \right\} \prod_{j=1}^N \mu(dx_j).$$

Using the fact that decisions are invariant by any permutation of the sequence, we get

$$\mathbf{E}\{\mathbf{M}_{N,n}^*\} = \int_{\mathcal{X}} f(x) f^\top(x) \xi_{N,n}^*(dx) = \mathbf{M}(\xi_{N,n}^*)$$

where

$$\xi_{N,n}^*(dx) = \frac{N}{n} \left(\int_{\mathcal{X}^{N-1}} J_n[x_j(j=1, \dots, N-1), x] \prod_{j=1}^{N-1} \mu(dx_j) \right) \mu(dx).$$

Since $J_n \in \{0, 1\}$, $\xi_{N,n}^*(dx)$ satisfies $\xi_{N,n}^*(dx) \leq (N/n)\mu(dx)$. This implies $\Phi[\mathbf{M}(\xi_{N,n}^*)] \leq \Phi(\xi_{n/N}^*)$. Concavity of $\Phi(\cdot)$ finally gives

$$\mathbf{E}\{\Phi(\mathbf{M}_{N,n}^*)\} \leq \Phi[\mathbf{E}\{\mathbf{M}_{N,n}^*\}] \leq \Phi(\xi_{n/N}^*)$$

which concludes the proof. \square

Proof of Theorem 8.

We shall need the following lemma.

Lemma 14 *Let (α_k) be a sequence in $(0, 1)$, with $\lim_{k \rightarrow \infty} \alpha_k = \alpha \in (0, 1)$.*

*Then, under **H_Φ1-H_Φ5**,*

$$\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, [\alpha_N N]}^*) = \phi_\alpha^*, \quad \mu\text{-a.s.}$$

Proof of Lemma 14. From **H_Φ3**, for $n_1 \leq n_2 \leq N$, $\Psi(\mathcal{S}_{N, n_1}^*) = \Phi(\mathbf{M}_{N, n_1}^*) \leq \Phi[(n_2/n_1)\mathbf{M}_{N, n_2}^*]$, with $\mathbf{M}_{N, n}^*$ defined as in the proof of Lemma 7. Since $\alpha_N \rightarrow \alpha$, for any ϵ such that $0 < \epsilon < \min(\alpha, 1 - \alpha)$ and for N large enough, $\alpha - \epsilon < \alpha_N < \alpha + \epsilon$. This implies that there exists N_0 such that, for any $N > N_0$,

$$\Phi(\mathbf{M}_{N, [\alpha_N N]}^*) \leq \Phi\left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \mathbf{M}_{N, [(\alpha + \epsilon)N]}^*\right)$$

and

$$\Phi(\mathbf{M}_{N, [(\alpha - \epsilon)N]}^*) \leq \Phi\left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \mathbf{M}_{N, [\alpha_N N]}^*\right).$$

H_Φ5(i) then gives

$$\begin{aligned} \Phi(\mathbf{M}_{N, [(\alpha - \epsilon)N]}^*) - g_1\left(\frac{\alpha + \epsilon}{\alpha - \epsilon}\right) &\leq \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) \\ &\leq \Phi(\mathbf{M}_{N, [(\alpha + \epsilon)N]}^*) + g_1\left(\frac{\alpha + \epsilon}{\alpha - \epsilon}\right) \end{aligned}$$

and thus, from Theorem 6,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) &\leq \phi_{\alpha + \epsilon}^* + g_1\left(\frac{\alpha + \epsilon}{\alpha - \epsilon}\right) \\ \phi_{\alpha - \epsilon}^* - g_1\left(\frac{\alpha + \epsilon}{\alpha - \epsilon}\right) &\leq \liminf_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*), \end{aligned}$$

μ -a.s. The continuity of ϕ_α^* with respect to α , see Theorem 4, and $\lim_{a \rightarrow 1^+} g_1(a) = 0$ give $\lim_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) = \phi_\alpha^*$, μ -a.s. When **H $_\Phi$ 5(ii)** is satisfied, we obtain

$$\begin{aligned} \limsup_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) &\leq g_2 \left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \right) \phi_{\alpha + \epsilon}^* \\ \left[g_2 \left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \right) \right]^{-1} \phi_{\alpha - \epsilon}^* &\leq \liminf_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*), \end{aligned}$$

μ -a.s., and the continuity of ϕ_α^* with respect to α together with $\lim_{a \rightarrow 1^+} g_2(a) = 1$ give $\lim_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) = \phi_\alpha^*$, μ -a.s. \square

We return to the proof of Theorem 8.

Consider the matrix $\mathbf{M}_{N, [\alpha_N N]}^*$ associated with the strategy $\mathcal{S}_{N, [\alpha_N N]}^*$. Strict concavity of $\Phi(\cdot)$ (**H $_\Phi$ 1**) and $\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, [\alpha_N N]}^*) = \phi_\alpha^*$ μ -a.s., see Lemma 14, imply $\mathbf{M}_{N, [\alpha_N N]}^* \rightarrow \mathbf{M}(\xi_\alpha^*)$ μ -a.s. **H $_\mu$ 3** then gives $\lambda_{\min}(\mathbf{M}_{N, [\alpha_N N]}^*) \rightarrow \lambda_{\min}[\mathbf{M}(\xi_\alpha^*)] = l_\alpha^*$, μ -a.s. Therefore, there exists N_0 (μ -a.s.), such that for any $N > N_0$, $\lambda_{\min}(\mathbf{M}_{N, [\alpha_N N]}^*) > l_\alpha^*/2$.

Define $\mathbf{M}'_{N, [\alpha_N N]}(i, X)$ as the matrix obtained from $\mathbf{M}_{N, [\alpha_N N]}^*$ by substituting X for the point X_i of the sequence. **H $_\Phi$ 1-H $_\Phi$ 3** imply

$$\Phi[\mathbf{M}'_{N, [\alpha_N N]}(i, X)] \leq \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) + \frac{1}{[\alpha_N N]} \mathcal{G}_\Phi[\mathbf{M}_{N, [\alpha_N N]}^*, f(X)f^\top(X)].$$

H $_\Phi$ 4 gives for $N > N_0$

$$\Phi[\mathbf{M}'_{N, [\alpha_N N]}(i, X)] \leq \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) + \frac{1}{[\alpha_N N]} \delta_1(l_\alpha^*/2) \|f(X)\|^2. \quad (\text{B.1})$$

Consider the event $\mathcal{E}(A, N) = (\max_{i=1, \dots, N} \|f(X_i)\| < AN^{\gamma'})$, with $\gamma' \in (1/\gamma, 1/4)$ where γ is the constant in **H $_\mu$ 2** ($\gamma > 4$).

Conditionally on $\mathcal{E}(A, N)$,

$$\begin{aligned}\lambda_{\min}[\mathbf{M}_{N, [\alpha_N N]}^*(i, X)] &\geq \lambda_{\min}(\mathbf{M}_{N, [\alpha_N N]}^*) - \frac{1}{[\alpha_N N]} \max_{i=1, \dots, N} \max_{\|u\|=1} [u^\top f(X_i)]^2 \\ &\geq \lambda_{\min}(\mathbf{M}_{N, [\alpha_N N]}^*) - \frac{1}{[\alpha_N N]} A^2 N^{2\gamma'}\end{aligned}$$

Since $\gamma' < 1/2$, there exists N_1 (μ -a.s.) such that

$$\forall N > N_1, \lambda_{\min}[\mathbf{M}_{N, [\alpha_N N]}^*(i, X)] \geq l_\alpha^*/4.$$

Using **H_Φ1-H_Φ4** again, we get

$$\Phi(\mathbf{M}_{N, [\alpha_N N]}^*) \leq \Phi[\mathbf{M}_{N, [\alpha_N N]}^*(i, X)] + \frac{1}{[\alpha_N N]} \delta_1(l_\alpha^*/4) A^2 N^{2\gamma'}. \quad (\text{B.2})$$

Combining (B.1) and (B.2), we get on $\mathcal{E}(A, N)$ for $N > N_1$ and $\|f(X)\| < AN^{\gamma'}$:

$$|\Phi[\mathbf{M}_{N, [\alpha_N N]}^*(i, X)] - \Phi(\mathbf{M}_{N, [\alpha_N N]}^*)| < \frac{1}{[\alpha_N N]} \delta_1(l_\alpha^*/4) A^2 N^{2\gamma'}.$$

The bounded difference inequality, see (McDiarmid, 1989) and Theorem 2.2 of Devroye and Lugosi (2001), then gives, on $\mathcal{E}(A, N)$:

$$\text{Prob} \left[\Phi(\mathbf{M}_{N, [\alpha_N N]}^*) \geq \mathbf{E}\{\Phi(\mathbf{M}_{N, [\alpha_N N]}^*)\} + t \right] \leq \exp[-h(N, t)],$$

with

$$h(N, t) = \frac{2t^2 \lfloor \alpha_N N \rfloor}{A^4 N^{4\gamma'} \delta_1^2(l_\alpha^*/4)}.$$

Assume that there exists A such that $\mathcal{E}(A, N)$ is true for any N . Then, taking $t = t_N = \lfloor \alpha_N N \rfloor^{-\beta}$ with $0 < \beta < (1 - 4\gamma')/2$, we obtain

$$\text{Prob} \left\{ \Phi(\mathbf{M}_{N, [\alpha_N N]}^*) \geq \mathbf{E}\{\Phi(\mathbf{M}_{N, [\alpha_N N]}^*)\} + \frac{1}{[\alpha_N N]^\beta} \text{ i.o.} \right\} = 0.$$

We finally show that such an A exists μ -a.s. Assumption **H_μ2** implies

$$\text{Prob}\{\|f(X_1)\| > N^{\gamma'}\} \leq \frac{T}{N^{\gamma\gamma'}}$$

for some $T > 0$. Since $\gamma\gamma' > 1$ [$\gamma' \in (1/\gamma, 1/4)$], Theorem 3.5.1, p. 169 of Embrechts et al. (1997) gives

$$\text{Prob}\left\{\max_{i=1,\dots,N} \|f(X_i)\| > N^{\gamma'} \text{ i.o.}\right\} = 0,$$

which completes the proof. \square

Proof of Theorem 9.

From **H $_{\Phi}$ 2**, **H $_{\Phi}$ 6**, the empirical design measure ξ_k generated by $\mathcal{S}_\alpha(\mu)$ satisfies the recurrence

$$\begin{aligned} \phi_k = \phi(\xi_k) = \phi_{k-1} + & \left[\frac{1}{1+a_k} F_\Phi(\xi_{k-1}, X_k) + \frac{1}{2(1+a_k)^2} H_\Phi(\xi_{k-1}, X_k, \gamma) \right] \times \\ & \left[\mathcal{I}_{\mathcal{X}_{2,k}}(X_k) + \mathcal{I}_{[0, P_k(\alpha)]}(Z) \mathcal{I}_{\mathcal{X}_{3,k}}(X_k) \right], \end{aligned}$$

where a_k is defined by (4), Z is a random variable uniformly distributed in $[0, 1]$, $\mathcal{I}_{\mathcal{A}}(\cdot)$ denotes the indicator function of the set \mathcal{A} , and

$$H_\Phi(\xi_{k-1}, X_k, \gamma) = \nabla_\Phi^2[(1-\gamma)\mathbf{M}(\xi_{k-1}) + \gamma f(X_k) f^\top(X_k), f(X_k) f^\top(X_k)]$$

for some $\gamma \in [0, 1/(1+a_k)]$.

Since X_k is accepted with probability α , $a_k/k \rightarrow \alpha$ μ -a.s. as $k \rightarrow \infty$.

From **H $_{\mu}$ 4**, there exists K_0 (μ -a.s.) such that for any $k > K_0$, $\lambda_{\min}[\mathbf{M}(\xi_{k-1})] > \underline{l}/2$. Since $\lambda_{\min}[(1-\gamma)\mathbf{M}(\xi_{k-1}) + \gamma f(X_k) f^\top(X_k)] \geq [1-1/(1+a_k)]\lambda_{\min}[\mathbf{M}(\xi_{k-1})]$, there exists K_1 (μ -a.s.) such that for $k > K_1$, $\lambda_{\min}[(1-\gamma)\mathbf{M}(\xi_{k-1}) + \gamma f(X_k) f^\top(X_k)] > \underline{l}/4$. From **H $_{\Phi}$ 6**, this implies $H_\Phi(\xi_{k-1}, X_k, \gamma) > -\delta_0(\underline{l}/4) - \delta_2(\underline{l}/4) \|f(X_k)\|^4$. We can now compute a lower bound on $\mathbf{E}\{\phi_k | \mathcal{F}_{k-1}\}$. Notice that

$$\begin{aligned} \mathbf{E}\{F_\Phi(\xi_{k-1}, X_k) \left[\mathcal{I}_{\mathcal{X}_{2,k}}(X_k) + \mathcal{I}_{[0, P_k(\alpha)]}(Z) \mathcal{I}_{\mathcal{X}_{3,k}}(X_k) \right] | \mathcal{F}_{k-1}\} = \\ \alpha F_\Phi[\xi_{k-1}; T_{\Phi, \alpha}(\xi_{k-1})] \end{aligned}$$

which gives for $k > K_1$,

$$\mathbb{E}\{\phi_k | \mathcal{F}_{k-1}\} > \phi_{k-1} + \frac{\alpha}{1+a_k} F_{\Phi}[\xi_{k-1}; T_{\Phi, \alpha}(\xi_{k-1})] - \alpha \frac{\delta_0(L/4) + \delta_2(L/4)\mu_4}{2(1+a_k)^2}$$

with $\mu_4 = \int_{\mathcal{X}} \|f(x)\|^4 \mu(dx)$, and $\mu_4 < \infty$ from **H _{μ 2}**. Using (10), we obtain

$$\mathbb{E}\{\phi_k | \mathcal{F}_{k-1}\} > \phi_{k-1} + \frac{\alpha}{1+a_k} (\phi_{\alpha}^* - \phi_{k-1}) - \alpha \frac{\delta_0(L/4) + \delta_2(L/4)\mu_4}{2(1+a_k)^2}.$$

Assume that $\limsup \phi_k < \phi_{\alpha}^* - \delta$ for some $\delta > 0$. This gives

$$\mathbb{E}\{\phi_k | \mathcal{F}_{k-1}\} > \phi_{k-1} + \frac{\alpha\delta}{2(1+a_k)} \quad (\text{B.3})$$

for k large enough. $\mathbb{E}\{\phi_k\}$ is bounded [from (17), **H _{Φ 1}** and **H _{μ 4}**]. The martingale convergence theorem then says that ϕ_k converges μ -a.s. to a finite limit, which contradicts (B.3). Therefore, $\limsup_{k \rightarrow \infty} \phi_k \geq \phi_{\alpha}^*$. At the same time, $\limsup_{k \rightarrow \infty} \phi_k \leq \lim_{k \rightarrow \infty} \Psi(\mathcal{S}_{k, a_{k+1}}^*) = \phi_{\alpha}^*$ μ -a.s. from Lemma 14, so that

$$\limsup_{k \rightarrow \infty} \phi_k = \phi_{\alpha}^*, \quad \mu\text{-a.s.} \quad (\text{B.4})$$

Define $T(B) = \inf\{k / \phi_k > \phi_{\alpha}^* + B/k^{\beta'}\}$, for some $\beta' > 0$. On $(T(B) = \infty)$ and for $k > K_1$, we have

$$\mathbb{E}\{\phi_{\alpha}^* - \phi_k | \mathcal{F}_{k-1}\} < \phi_{\alpha}^* - \phi_{k-1} + \frac{\alpha}{1+a_k} \frac{B}{k^{\beta'}} + \alpha \frac{\delta_0(L/4) + \delta_2(L/4)\mu_4}{2(1+a_k)^2}.$$

Introduce a new variable ω_k defined by

$$\omega_k = \phi_{\alpha}^* - \phi_k - \alpha \sum_{i=1}^k \left[\frac{B}{i^{\beta'}(1+a_i)} + \frac{\delta_0(L/4) + \delta_2(L/4)\mu_4}{2(1+a_i)^2} \right];$$

it satisfies $\sup_k \mathbb{E}\{|\omega_k|\} < \infty$ and $\mathbb{E}\{\omega_k | \mathcal{F}_{k-1}\} < \omega_{k-1}$, therefore, from the martingale convergence theorem, there exists $\omega_{\infty} < \infty$ such that $\omega_k \rightarrow \omega_{\infty}$ (μ -a.s.) as $k \rightarrow \infty$. Since $\sum_{i=1}^k 1/[i^{\beta'}(1+a_i)]$ and $\sum_{i=1}^k 1/(1+a_i)^2$ converge

μ -a.s., on $(T(B) = \infty)$ ϕ_k tends to a finite limit ϕ_∞ , μ -a.s., and $\phi_\infty = \phi_\alpha^*$ from (B.4).

Finally, we only need to show that $\text{Prob}\{\cup_{B=1}^\infty (T(B) = \infty)\} = 1$ to complete the proof. Define $\alpha_k = a_{k+1}/k$. Using Theorem 8 and (16), we get

$$\text{Prob}\left\{\phi_k \geq \phi_{\alpha_k}^* + \frac{1}{[\alpha k]^\beta} \text{ i.o.}\right\} = 0. \quad (\text{B.5})$$

Kolmogorov's Theorem, see, e.g., Shiryaev (1996, p. 389), gives

$$(a_{k+1} - k\alpha)/(\sqrt{k} \log k) \rightarrow 0 \text{ } \mu\text{-a.s.}$$

Therefore, there exists K_2 (μ -a.s.) such that, for any $k > K_2$, $|\alpha_k - \alpha| \leq \log k/\sqrt{k}$. Theorem 4 then gives

$$\phi_{\alpha_k}^* \leq \phi_\alpha^* + \frac{1}{k^{1/4}}$$

for k larger than some K_3 . Combining this with (B.5) we get

$$\text{Prob}\left\{\phi_k \geq \phi_\alpha^* + \frac{1}{[\alpha k]^\beta} + \frac{1}{k^{1/4}} \text{ i.o.}\right\} = 0,$$

so that, $\text{Prob}\{\cup_{B=1}^\infty (T(B) = \infty)\} = 1$ when $\beta' = \min(1/4, \beta)$, which completes the proof. \square

Acknowledgements

This work originated from a comment by V.V. Fedorov (GlaxoSmithKline, USA) at a meeting on experimental design held in Cardiff in April 2000, see Pronzato (2001b); he is gratefully acknowledged for pointing out the connection between the results presented in this meeting and optimum design with constraints. Also, stimulating discussion with H.P. Wynn (Warwick University

and London School of Economics, UK) and E. Thierry (Laboratoire I3S) on different occasions proved extremely useful.

References

- Albright, S., Derman, C., 1972. Asymptotically optimal policies for the stochastic sequential assignment problem. *Management Science* 19 (1), 46–51.
- Box, G., Lucas, H., 1959. Design of experiments in nonlinear situations. *Biometrika* 46, 77–90.
- C.E.R.N., 1994. The compact muon solenoid, technical proposal, trigger and data acquisition. Tech. Rep. CERN/LHCC 94-38, European Laboratory for Particle Physics.
- Devroye, L., Lugosi, G., 2001. *Combinatorial Methods in Density Estimation*. Springer, Heidelberg.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer, Berlin.
- Fedorov, V., 1989. Optimal design with bounded density: optimization algorithms of the exchange type. *J. Statist. Planning and Inference* 22, 1–13.
- Fedorov, V., Hackl, P., 1997. *Model-Oriented Design of Experiments*. Springer, Berlin.
- Jennrich, R., 1969. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.* 40, 633–643.
- Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. *Canadian Journal of Mathematics* 12, 363–366.
- McDiarmid, C., 1989. On the method of bounded differences. *Survey in Combinatorics*, Cambridge University Press, 148–188.

- Molchanov, I., Zuyev, S., 2001. Variational calculus in the space of measures and optimal design. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), *Optimum Design 2000*. Kluwer, Dordrecht, Ch. 8, pp. 79–90.
- Pronzato, L., 1999. Sequential selection of observations in randomly generated experiments. In: Proc. ProbaStat'98, Smolenice (Slovaquie), Feb. 98, *Tatra Mountains Mathematical Publications*. Vol. 17. pp. 167–175.
- Pronzato, L., 2001a. Optimal and asymptotically optimal decision rules for sequential screening and resource allocation. *IEEE Transactions on Automatic Control* 46 (5), 687–697.
- Pronzato, L., 2001b. Sequential construction of an experiment design from and i.i.d. sequence of experiments without replacement. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), *Optimum Design 2000*. Kluwer, Dordrecht, Ch. 11, pp. 113–122.
- Sahm, M., Schwabe, R., 2001. A note on optimal bounded designs. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), *Optimum Design 2000*. Kluwer, Dordrecht, Ch. 13, pp. 131–140.
- Shiryaev, A., 1996. *Probability*. Springer, Berlin.
- Shorack, G., 2000. *Probability for Statisticians*. Springer, Berlin.
- Wu, C., Wynn, H., 1978. The convergence of general step-length algorithms for regular optimum design criteria. *Annals of Statistics* 6 (6), 1273–1285.
- Wynn, H., 1982. Optimum submeasures with applications to finite population sampling. In: Gupta, S., Berger, J. (Eds.), *Statistical Decision Theory and Related Topics III*. Proc. 3rd Purdue Symp., vol. 2. Academic Press, New York, pp. 485–495.