On the sequential construction of optimum bounded designs
Luc Pronzato

To cite this version:
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 one samples from a finite population of independent and identically distributed experimental conditions $X$. The size of the population is $N$ but only $n$ samples, a proportion $\alpha$ of $N$, can be used. The quality of a sample is measured by a regular optimality criterion $\phi(\cdot)$ based on the information matrix, such as the $D$-criterion. The construction of an optimal approximate design bounded by $\mu/\alpha$, with $\mu$ the probability measure of $X$, can be used to construct a sampling strategy which is asymptotically optimum (when the size $N$ of the population tends to infinity). We show that a sequential strategy which does not require any information on $\mu$ is also asymptotically optimum. Some possible applications are indicated.

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

1991 MSC: 62K05, 62L05

________________________

Email address: pronzato@i3s.unice.fr (Luc Pronzato).
URL: http://www.i3s.unice.fr/ pronzato/ (Luc Pronzato).

Preprint submitted to Elsevier Science 20 August 2004
1 Introduction

We consider a parameter estimation problem, with $\theta$ the vector of parameters to be estimated and $X$ the experimental variables, $X \in \mathcal{X} \subset \mathbb{R}^q$. We assume that the observations are independent, so that the Fisher information matrix is the sum of rank-one matrices of the type $f(X)f^T(X)$. For instance, this may correspond to estimation in a linear regression model, with independent observations $Y_k = f^T(X_k)\bar{\theta} + \varepsilon_k$, where the $\varepsilon_k$'s are independently identically distributed (i.i.d.) with $\mathbb{E}\{\varepsilon_k\} = 0$ and $\bar{\theta} \in \mathbb{R}^d$ is the unknown true value of the model parameters to be estimated.

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

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

and $\phi(\xi) = \Phi[M(\xi)]$, $\xi \in \Xi$. We give in Appendix A a list of assumptions on $\Phi$ that will be used throughout the paper. We shall always assume that $\Phi$ is strictly concave ($H_\Phi 1$), differentiable ($H_\Phi 2$) and increasing ($H_\Phi 3$). The assumptions are discussed in the same appendix; they are satisfied in particular when $\Phi(M) = \log \det M$ or $\Phi(M) = -\text{trace}(AM^{-p})$, with $p$ a positive integer and $A$ a positive definite matrix.

We consider the situation where the experimental conditions $X_k \in \mathcal{X}$ form a sequence of i.i.d. variables sampled in a population of size $N$, and only $n < N$ samples can be used. We focus on the sequential problem, where, as soon as a new sample $X_k$ becomes available, we must decide whether to use it or not,
that is, 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 6. Notice that the case $\text{dim}(\theta) = 1$ corresponds to a variant of the
secretary problem, see (Pronzato, 2001a).

Let $\mu$ denote the probability measure of $X_1$, with $\int_\mathcal{X} \mu(dx) = 1$. We assume
that $\mu$ is such that

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

exists, with $-\infty < \phi(\mu) < \infty$; a list of additional assumptions on $\mu$ is given
in Appendix A. The sequence of decisions will be denoted $(u_k)$: $u_k = 1$ if we
decide to observe $Y_k$, with experimental conditions $X_k$, and $u_k = 0$ otherwise,
with, for any admissible policy,

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

(Nota que $X_k$ is known when $u_k$ is chosen.) The associated information matrix
$M_{N,n}$ (normalized, per sample collected) is given by

$$
M_{N,n} = \frac{1}{n} \sum_{k=1}^N u_k f(X_k)f^\top(X_k). \ (2)
$$
(It may be singular for small \(n\), but \(\Phi(M_{N,n}) > -\infty\) for \(n\) large enough if the \(X_k\)’s are selected in a suitable population, see \(H_1, H_1\).)

For \(N\) finite, we shall consider the following sequential problem

\[
\text{maximise } E\{\Phi(M_{N,n})\}
\]

with respect to \((u_j)\) satisfying (1), the expectation \(E\{\cdot\}\) being with respect to the product measure \(\mu^\otimes N\) of \(X_1, \ldots, X_N\) (we shall see, c.f. (12), that the concavity and increasing properties of \(\Phi\) and \(\phi(\mu) < \infty\) imply that \(E\{\Phi(M_{N,n})\} < \infty\) for any \(N, n, 0 < n \leq N\), and any sequence \((u_j)\)).

For any sequence \((u_j)\) and any step \(k, 1 \leq k \leq N\), \(a_k\) will denote the number of observations already made; that is,

\[
a_k = \sum_{j=1}^{k-1} u_j,\]

with \(a_1 = 0\). For each \(k \in \{1, \ldots, N\}\), the optimal decision at step \(k\) is obtained by solving:

\[
\max_{u_k \in \mathcal{U}_k} \mathbb{E}_{X_{k+1}} \left\{ \max_{u_{k+1} \in \mathcal{U}_{k+1}} \mathbb{E}_{X_{k+2}} \left\{ \max_{u_{k+2} \in \mathcal{U}_{k+2}} \left[ \ldots \right] \right\} \right\} \ldots \\
\mathbb{E}_{X_{N-1}} \left\{ \max_{u_{N-1} \in \mathcal{U}_{N-1}} \mathbb{E}_{X_N} \left\{ \max_{u_N \in \mathcal{U}_N} \left[ \Phi((1/n) \sum_{i=1}^N u_i f(X_i)f^\top(X_i))] \right] \ldots \right\} \right\},
\]

where \(\mathbb{E}_{X_j}\{\cdot\}\) denotes the expectation with respect to \(X_j\), distributed with the measure \(\mu\), and the sets \(\mathcal{U}_j\) satisfy

\[
\mathcal{U}_j = \mathcal{U}_j(a_j) = \begin{cases} 
0 & \text{if } a_j = n, \\
1 & \text{if } a_j + N - j + 1 = n, \\
\{0, 1\} & \text{otherwise.}
\end{cases}
\]
In the first case \((a_j = n)\) the maximum number of samples allowed has already been collected, the second one \((a_j + N - j + 1 = n)\) means that all remaining samples of the population must be accepted in order to reach a total of \(n\) samples selected. It is only in the last case that some freedom exists for decisions.

The case \(d = \text{dim}(\theta) = 1\) is considered in (Pronzato, 2001a). The optimal (closed-loop) solution is given by a backward recurrence equation. Using results on extreme value distributions, a simple open-loop solution is proved to be asymptotically optimal for \(N \to \infty\) with \(n\) fixed (for measures \(\mu\) absolutely continuous with respect to the Lebesgue measure and such that the associated distribution function is a von Mises function). This extends the results of Albright and Derman (1972) which concern the case \(n = \lfloor \alpha N \rfloor, \alpha \in (0, 1)\).

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 \(n = \lfloor \alpha N \rfloor, \alpha \in (0, 1), N \to \infty\). The fact that \(n\) tends to infinity at the same speed as \(N\) means that we shall obtain asymptotic performances that are achieved \(\mu\)-almost surely (a.s.), contrary to (Pronzato, 2001a) which concerns expected performances.

The different types of strategies to be considered (with finite or infinite horizon, stationary, non-adaptive, randomized) are defined in Section 2. The value \(\phi(\xi^*_\alpha)\), with \(\xi^*_\alpha \leq \mu/\alpha\) a \(\phi\)-optimum bounded design measure, is shown to form an upper bound on the asymptotic performance of any strategy. A trivial
asymptotically optimal strategy $\mathcal{S}_{N,n}(\alpha)$ is proposed in Section 3: $\mathcal{S}_{N,n}(\alpha)$ is the truncated version (finite horizon) of a stationary strategy $\mathcal{S}(\xi^*_\alpha)$ which samples from $\xi^*_\alpha$. We show in Section 4 that a slight modification of $\mathcal{S}(\xi^*_\alpha)$ leads to an asymptotically optimal strategy that does not require the prior knowledge of the measure $\mu$. Illustrative examples are presented in Section 5. Section 6 gives some concluding remarks and suggests some extensions.

2 Strategies, performances and asymptotic optimality

2.1 Strategies

We shall consider sequential strategies for which at step $k$ only $X_1, \ldots, X_k$ have been observed. When $\mu$ is known, $\Omega_k = (k, a_k, M_{k-1,a_k}, X_k)$ then summarizes all the information necessary to make a decision about the acceptance of $X_k$, and the problem (3) corresponds to a discrete-time stochastic control problem, where $k$ represents time, $(k, a_k, M_{k-1,a_k}, X_k)$ and $u_k \in \mathcal{U}_k \subseteq \{0, 1\}$ respectively represent the state\(^1\) and control at time $k$. A sequential strategy $\mathcal{S}_{N,n}$ is then defined by a sequence of mappings $(k, a, M, X) \mapsto u \in \{0, 1\}$, $k = 1, \ldots, N$.

For instance, the optimal strategy defined by (5) is sequential.

Strategies are non-sequential when the selection of the $n$ $X_k$’s is made after all the population $X_1, \ldots, X_N$ has been observed. A non-sequential strategy will be used to obtain an upper bound on the performance of sequential strategies.

\(^1\) When $\mu$ is unknown, a case considered in Section 4, at step $k$ an estimate $\hat{\mu}_k$ of $\mu$ (empirical version, parametric representation, etc.) based on $X_1, \ldots, X_k$ must also be used to decide about $X_k$; the “state” $\Omega_k$ should then be extended to include this $\hat{\mu}_k$. \[6\]
A sequential strategy $\mathcal{S}$ will be called *stationary* when the mapping is from $(M, X)$ to $u \in \{0, 1\}$. Stationary strategies correspond to the case of infinite horizon ($N = \infty$) where we only require that $a_k/k \to \alpha$ when $k \to \infty$. We shall denote $\xi_{k-1}$ the empirical measure of the $a_k$ samples already accepted at step $k$, with $M(\xi_{k-1})$ the current value of the (normalized) information matrix,

$$M(\xi_{k-1}) = \frac{1}{a_k} \sum_{j=1}^{k-1} u_j f(X_j) f^\top(X_j),$$

and write

$$\phi_k = \phi(\xi_k) = \Phi[M(\xi_k)].$$

*Non-adaptive* strategies are special cases of stationary strategies for which the decision at step $k$ only depends on the value of $X_k$, and $\mathcal{S}$ samples from some measure $\xi \leq \mu/\alpha$; we shall then write $\mathcal{S} = \mathcal{S}(\xi)$. We consider non-adaptive strategies as sequential, although the information carried by previous samples is not used.

The non-adaptive strategies $\mathcal{S}(\xi)$ that we shall consider have the property that $\xi = 0$ on $\mathcal{X}_1$, $\xi = \mu/\alpha$ on $\mathcal{X}_2$ and $\xi = \beta \mu/\alpha$, $0 < \beta < 1$, on $\mathcal{X} \setminus (\mathcal{X}_1 \cup \mathcal{X}_2)$ for some subsets $\mathcal{X}_1, \mathcal{X}_2$ of $\mathcal{X}$. This means that $\mathcal{S}(\xi)$ should reject every $X_k$ falling in $\mathcal{X}_1$, accept all the samples in $\mathcal{X}_2$ but only a fraction $\beta$ of those in $\mathcal{X}_3$. Therefore, the strategy is completely defined when $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$ but, when $\mathcal{X}_3 \neq \emptyset$, we must specify how the fraction $\beta$ of the samples in $\mathcal{X}_3$ is selected.

We shall then consider *randomized* strategies, which lead to a simple analysis: a sample in $\mathcal{X}_3$ will be accepted with probability $\beta$, by tossing a biased coin. The mapping $X \mapsto u \in \{0, 1\}$ is then a random function when $X \in \mathcal{X}_3$.

A stationary strategy, defined for infinite horizon, will be implemented when the horizon is finite through a truncation procedure, which will be shown to
preserve asymptotic optimality (when \(n = \lfloor \alpha N \rfloor, \alpha \in (0, 1), N \to \infty\)).

### 2.2 Performances

For any strategy \(S_{N,n}\) used to solve (3), with \(0 < n \leq N\), we shall denote \(\Psi(S_{N,n}) = \Phi(M_{N,n})\), which measures the performance of \(S_{N,n}\) (for a particular realization of the sequence \(X_1, \ldots, X_N\)).

Following the same line as in (Albright and Derman, 1972), which concerns the case \(d = 1\), we can use as a benchmark the infeasible, but better-than-optimal, non-sequential strategy \(S_{N,n}^*\), obtained by selecting the \(n\) design points \(X_{k_1}, \ldots, X_{k_n}\) that maximise \(\Phi(M_{N,n})\) after the \(N\) points \(X_1, \ldots, X_N\) have been observed. \(S_{N,n}^*\) is thus a \(\phi\)-optimum design algorithm that generates an exact \(n\)-point \(\phi\)-optimum design in the finite design space \(\{X_1, \ldots, X_N\}\). Obviously, for any \(N\), \(n\) and any strategy \(S_{N,n}\), sequential or not,

\[
\Psi(S_{N,n}) \leq \Psi(S_{N,n}^*) .
\]  

(6)

Also, \(S_{N,n}^*\) satisfies

\[
\forall \alpha \in (0, 1), \limsup_{N \to \infty} \Psi(S_{N,\lfloor \alpha N \rfloor}^*) \leq \phi_{\alpha}^*, \ \text{\(\mu\)-a.s.}
\]

(7)

with \(\phi_{\alpha}^* = \phi(\xi_{\alpha}^*)\) and \(\xi_{\alpha}^* \leq \mu/\alpha\) a \(\phi\)-optimum constrained design measure: \(\xi_{\alpha}^*\) maximises \(\phi(\xi), \xi \in \mathcal{D}(\mu, \alpha)\), with \(\mathcal{D}(\mu, \alpha) \subset \Xi\) the set of admissible measures satisfying

\[
\xi(dx) \leq \mu(dx)/\alpha .
\]

When \(\Phi(\cdot)\) is strictly concave (H\(_\Phi\)1), \(\mathbf{M}(\xi_{\alpha}^*)\) is unique (although \(\xi_{\alpha}^*\) is not necessarily unique). (H\(_\Phi\)1) implies \(\phi_{\alpha}^* \geq \phi(\mu) > -\infty\) and (H\(_\Phi\)3) implies
\( \phi_\alpha^* \leq \phi(\mu/\alpha) < \infty. \) Other properties of \( \xi_\alpha^* \) will be presented in Section 3.1, see also Wynn (1982); Fedorov (1989); Fedorov and Hackl (1997); Sahm and Schwabe (2001); Pronzato (2004).

Following (6, 7), a strategy \( S_{N,n} \) will be called \textit{asymptotically optimal} (for \( n = [\alpha N], \alpha \in (0, 1), N \to \infty \)) when \( \lim_{N \to \infty} \Psi(S_{N,[\alpha N]}) = \phi^*_\alpha, \mu\text{-a.s.} \).

Although sampling from a \( \phi \)-optimum constrained measure \( \xi_\alpha^* \) yields \( \phi_k \to \phi^*_\alpha, \mu\text{-a.s.}, k \to \infty \) (we shall see in Section 3.2 how this can be implemented), it corresponds to a non-adaptive strategy that does not satisfy constraint (1). We show below how enforcing (1) can preserve asymptotic optimality.

For any \( \alpha \in (0, 1) \) let \( S_\alpha \) be a stationary strategy such that \( a_k/k \to \alpha \) and \( \phi_k \to \phi^*_\alpha, \mu\text{-a.s.}, k \to \infty. \) In particular, it may correspond to the non-adaptive strategy \( S(\xi_\alpha^*) \) which samples from a \( \phi \)-optimum constrained measure \( \xi_\alpha^* \leq \mu/\alpha. \) To any such \( S_\alpha \) we associate the following truncation, which defines a sequential strategy \( S_{N,n}(S_\alpha) \) satisfying the constraint \( a_{N+1} = n \) for any finite \( N \geq n: \)

\[
S_{N,n}(S_\alpha) : \begin{cases} 
\text{if } a_k = n, \text{ reject } X_k; \\
\text{if } N - k + 1 = n - a_k, \text{ accept } X_k; \\
\text{otherwise, apply } S_\alpha.
\end{cases}
\]

(8)

We show that \( S_{N,n}(S_\alpha) \) is asymptotically optimal for \( n = [\alpha N], \alpha \in (0, 1), N \to \infty. \) Replace \( \alpha \) by \( \alpha + \epsilon, \) with \( \epsilon > 0, \) and consider the following strategy,
simpler than \( S_{N,n}(S_{a+\epsilon}) \):

\[
\tilde{S}_{N,n}(S_{a+\epsilon}) : \begin{cases} 
  \text{if } a_k = n, \text{ reject } X_k; \\
  \text{otherwise, apply } S_{a+\epsilon}.
\end{cases}
\]

Let \( \bar{n} \) denote the number of samples accepted by \( \tilde{S}_{N,n}(S_{a+\epsilon}) \). We have \( \bar{n} \leq n \), and, if \( M_{N,n}(S_{a+\epsilon}) \) and \( \tilde{M}_{N,n}(S_{a+\epsilon}) \) denote the information matrices associated with both strategies, \( M_{N,n}(S_{a+\epsilon}) = \tilde{M}_{N,n}(S_{a+\epsilon}) \) when \( \bar{n} = n \) and \( M_{N,n}(S_{a+\epsilon}) = \tilde{M}_{N,n}(S_{a+\epsilon}) + (1/n) \sum_{j=1}^{n-\bar{n}} f(X'_j)f'(X'_j) \) when \( \bar{n} < n \), where the \( X'_j \)'s denote the points selected by \( S_{N,n}(S_{a+\epsilon}) \) and not \( \tilde{S}_{N,n}(S_{a+\epsilon}) \). Since \( \Phi(\cdot) \) is increasing, we have

\[
\Psi[\tilde{S}_{N,n}(S_{a+\epsilon})] \leq \Psi[S_{N,n}(S_{a+\epsilon})]. \tag{9}
\]

Let \( n_{a+\epsilon} \) denote the number of points accepted by \( S_{a+\epsilon} \) among the \( N \) samples of the sequence (these are points that would have been accepted by \( \tilde{S}_{N,n}(S_{a+\epsilon}) \) if ignoring the constraint \( a_k \leq n \)). It satisfies \( n_{a+\epsilon}/N \to \alpha + \epsilon, \mu\text{-a.s. as } N \) tends to infinity. Take \( n = \lfloor \alpha N \rfloor \) and let \( N \) tend to infinity. The probability that \( n_{a+\epsilon} \leq n \) infinitely often is zero when \( \epsilon > 0 \). Therefore, asymptotically, \( \tilde{S}_{N,\lfloor \alpha N \rfloor}(S_{a+\epsilon}) \) stops after \( n = \lfloor \alpha N \rfloor \) samples from \( S_{a+\epsilon} \) have been collected and

\[
\Psi[\tilde{S}_{N,\lfloor \alpha N \rfloor}(S_{a+\epsilon})] \to \phi^*_a, \mu\text{-a.s.}, N \to \infty.
\]

Together with (9), it implies

\[
\liminf_{N \to \infty} \Psi[S_{N,\lfloor \alpha N \rfloor}(S_{a+\epsilon})] \geq \phi^*_a, \mu\text{-a.s.}, N \to \infty.
\]

On the other hand, from (6, 7), \( \limsup_{N \to \infty} \Psi[S_{N,\lfloor \alpha N \rfloor}(S_{a+\epsilon})] \leq \phi^*_a, \mu\text{-a.s.} \)

We may then let \( \epsilon \) tend to zero and use the continuity of \( \phi^*_a \) with respect to
$\alpha$, see Theorem 6, to obtain the following property.

**Theorem 1** For $\alpha \in (0, 1)$ let $S_\alpha$ be a stationary strategy such that $a_k/k \to \alpha$ and $\phi_k \to \phi_\alpha^*$, $\mu$-a.s., $k \to \infty$. The strategy $S_{N,n}(S_\alpha)$ defined by (8) is asymptotically optimal for $n = [\alpha N]$, $N \to \infty$.

Section 3.1 will present some properties of $\phi$-optimum constrained design measures, to be used in Section 3.2 to implement the stationary strategy $S(\xi_\alpha^*)$. When $S(\xi_\alpha^*)$ is plugged in (8), the resulting strategy $S_{N,n}[S(\xi_\alpha^*)]$ is asymptotically optimal from the previous theorem. Notice that from (6, 7) the existence of such an asymptotically optimal strategy implies

$$\lim_{N \to \infty} \Psi(S_{N,\lfloor \alpha N \rfloor}^*) = \phi_\alpha^*, \mu\text{-a.s.} \quad (10)$$

This property can be extended into the following (see Appendix B), which will be used in the proof of Theorem 9.

**Lemma 2** Let $(\alpha_k)$ be a sequence in $(0, 1)$, with $\lim_{k \to \infty} \alpha_k = \alpha \in (0, 1)$. Then, under $H_\mu, 1$, $H_\phi 1-H_\delta 3$ and $H_\delta 5$,

$$\lim_{N \to \infty} \Psi(S_{N,\lfloor \alpha N \rfloor}^*) = \phi_\alpha^*, \mu\text{-a.s.}$$

Whereas (10) and Lemma 2 only concern the limiting behavior of $\Psi(S_{N,n}^*)$, the following result on $E\{\Psi(S_{N,n}^*)\}$ holds for any $N$ under very general conditions. The proof is given in Appendix B.

**Lemma 3** For a concave criterion $\Phi(\cdot)$, the non sequential strategy $S_{N,n}^*$ (that is, a $\phi$-optimum algorithm for an exact design with $n$ point in $\{X_1, \ldots, X_N\}$)
satisfies
\begin{equation}
\forall (n, N), \ 0 < n \leq N, \ E\{\Psi(S_{N,n}^*)\} \leq \phi(\xi_{n/N}^*),
\end{equation}

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

$\textbf{H}_\phi 1$ and $\textbf{H}_\phi 3$ imply $\phi(\xi) < \infty$ for any design measure $\xi \leq \mu/\alpha$, $\alpha$ given in $(0, 1)$, so that Lemma 3 implies that for any $(n, N)$, $0 < n \leq N$, and any strategy $S_{N,n}$
\begin{equation}
E\{\Psi(S_{N,n})\} \leq E\{\Psi(S_{N,n}^*)\} \leq \phi(\xi_{n/N}^*) < \infty.
\end{equation}

(When $S_{N,n}$ involves randomized decisions, see e.g. Section 3.2, the first expectation in (12) is also with respect to them.)

It may be noticed that the upper bound $\phi(\xi_{n/N}^*)$ for the expected performance $E\{\Psi(S_{N,n})\}$ is not necessarily achievable when $n$ is finite, $N \rightarrow \infty$ and the strategy $S_{N,n}$ is sequential, even in the case $d = \text{dim}(\theta) = 1$, see Example 3 of (Pronzato, 2001a).

3 Constrained design measures and asymptotically optimum sequential strategies

3.1 Optimum constrained design measures

The main result (Wynn, 1982; Sahm and Schwabe, 2001), presented in the following theorem, states that when $\phi_{\alpha}^*$ is a $\phi$-optimum constrained measure, $\mathcal{X}$ 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 \( \xi_a = 0 \) on \( \mathcal{X}_{1,a}^* \), \( \xi_a = \mu/\alpha \) on \( \mathcal{X}_{2,a}^* \) and the directional derivative \( F(\xi_a, x) \) (see Appendix A) constant on \( \mathcal{X}_{3,a}^* \).

**Theorem 4** The following statements are equivalent:

(i) \( \xi_a^* \) is a \( \phi \)-optimum constrained design measure;

(ii) there exists a number \( c \) such that \( F(\xi_a^*, x) \geq c \) for \( \xi_a^* \)-almost all \( x \) and \( F(\xi_a^*, x) \leq c \) for \( (\mu - \alpha \xi_a^*) \)-almost all \( x \);

(iii) there exist two subsets \( \mathcal{X}_{1,a}^* \) and \( \mathcal{X}_{2,a}^* \) of \( \mathcal{X} \) such that

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

The construction of the sets \( \mathcal{X}_{1,a}^* \) and \( \mathcal{X}_{2,a}^* \) is important in order to be able to sample from \( \xi_a^* \); also, we must precise the value of \( \xi_a^* \) on \( \mathcal{X}_{3,a}^* = \mathcal{X}\backslash(\mathcal{X}_{1,a}^* \cup \mathcal{X}_{2,a}^*) \).

This is considered in the rest of this section.

For a given \( \xi \), consider the random variable \( F_\phi(\xi, X_1) \) and let \( F_\xi(\cdot) \) denote the corresponding distribution function,

\[
F_\xi(s) = \mu \{ x / F_\phi(\xi, x) \leq s \}.
\]  

(13)

Define \( c_\alpha(\xi) \) as

\[
c_\alpha(\xi) = \min \{ s / \; F_\xi(s) \geq 1 - \alpha \} \tag{14}
\]

and

\[
\mathcal{X}_{1,a}(\xi) = \{ x / F_\phi(\xi, x) < c_\alpha(\xi) \},
\]

\[
\mathcal{X}_{2,a}(\xi) = \{ x / F_\phi(\xi, x) > c_\alpha(\xi) \},
\]

\[
\mathcal{X}_{3,a}(\xi) = \{ x / F_\phi(\xi, x) = c_\alpha(\xi) \}.
\]  

(15)
We then obtain $X_{j,\alpha} = X_{j,\alpha}(\xi^*_\alpha)$, $j = 1, 2, 3$, and $c_{\alpha}(\xi^*_\alpha)$ is the constant $c$ of Theorem 4.

Consider now the following transformation

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

Notice that $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] = \max_{\nu \in D(\mu, \alpha)} F_\Phi(\xi; \nu)$, with $D(\mu, \alpha)$ the set of admissible measures on $\mathcal{X}$ satisfying $\nu \leq \mu/\alpha$ (indeed, $T_{\Phi,\alpha}(\xi)$ distributes its mass on $\mathcal{X}$ where $F_\Phi(\xi; x)$ takes its highest values).

The next theorem complements Theorem 4 by a minimax formulation similar to the Kiefer-Wolfowitz (1960) Equivalence Theorem.

**Theorem 5** The following statements are equivalent:

(i) $\xi^*_\alpha$ is a $\phi$-optimum constrained design measure;

(ii) $F_\Phi[\xi^*_\alpha; T_{\Phi,\alpha}(\xi^*_\alpha)] = 0$;

(iii) $\xi^*_\alpha$ minimises $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)]$, $\xi \in D(\mu, \alpha)$;

(iv) $\xi^*_\alpha$ minimises $\max_{\nu \in D(\mu, \alpha)} F_\Phi(\xi; \nu)$, $\xi \in D(\mu, \alpha)$.

**Proof.** From Theorem 4, (i) gives $\phi[T_{\Phi,\alpha}(\xi^*_\alpha)] = \phi(\xi^*_\alpha)$ which implies (ii). Take any $\xi \in \Xi$ and denote $\xi^*_\alpha$ an optimum constrained design measure. From the definition of $T_{\Phi,\alpha}$,

$$\forall \nu \in D(\mu, \alpha), \quad F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq F_\Phi(\xi; \nu) \quad (16)$$
and thus $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq F_\Phi(\xi; \xi^*_\alpha)$. Concavity of $\Phi$ then gives

$$\forall \xi \in \Xi, \quad F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq \phi^*_\alpha - \phi(\xi). \quad (17)$$

Since $\phi^*_\alpha \geq \phi(\xi), \xi \in D(\mu, \alpha)$, (ii) implies (i). Equivalence between (i) and (iii) is obvious from (17); (16) implies the equivalence between (iii) and (iv). \hfill \Box

We conclude this section by mentioning the following Lipschitz property, see (Pronzato, 2004), which shows that $\phi^*_\alpha$, is continuous in $\alpha$.

**Theorem 6** For any $\alpha, \beta$ in $(0, 1)$, the associated $\phi$-optimum constrained design measures $\xi^*_\alpha$ and $\xi^*_\beta$ 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 (14).

### 3.2 An asymptotically optimum sequential strategy

The implementation of a non-adaptive strategy that samples from $\xi^*_\alpha$ is straightforward when $\mu$ has no atoms, that is, when for any $\Delta \mathcal{X}$ with $\mu(\Delta \mathcal{X}) > 0$ exists $\Delta \mathcal{X}' \subset \Delta \mathcal{X}$ such that $0 < \mu(\Delta \mathcal{X}') < \mu(\Delta \mathcal{X})$, with measures absolutely continuous w.r.t. the Lebesgue measure as a special case. Indeed, in that case there exists a $\xi^*_\alpha$ such that the set $\mathcal{X}^*_3,\alpha$ is empty, see (Wynn, 1982; Fedorov, 1989; Fedorov and Hackl, 1997), and a trivial implementation of $S(\xi^*_\alpha)$ then consists in accepting the samples that fall into $\mathcal{X}^*_2,\alpha$ and rejecting the others.

We shall use the transformation $T_{\Phi,\alpha}$ of previous section to deal with the general situation, and sample from $T_{\Phi,\alpha}(\xi^*_\alpha)$ (which is equivalent to sampling
from $\xi^*_\alpha$ since $\phi[T_{n,\alpha}(\xi^*_\alpha)] = \phi(\xi^*_\alpha)$, see the proof of Theorem 5). Since only the fraction $[\alpha - \mu(\mathcal{X}^*_3)]/\mu(\mathcal{X}^*_3)$ of the samples in $\mathcal{X}^*_3$ must be accepted, we randomize the selection procedure and define

$$
\mathcal{S}(\xi^*_\alpha) : \begin{cases} 
\text{if } X_k \in \mathcal{X}^*_2 \text{ accept } X_k; \\
\text{if } X_k \in \mathcal{X}^*_3 \text{ accept } X_k \text{ with probability } P^*_3 = [\alpha - \mu(\mathcal{X}^*_2)]/\mu(\mathcal{X}^*_3); \\
\text{otherwise reject } X_k.
\end{cases}
$$

(18)

From Theorem 1, when the non-adaptive strategy $\mathcal{S}(\xi^*_\alpha)$ is plugged in (8) the resulting $\mathcal{S}_{N,n}[\mathcal{S}(\xi^*_\alpha)]$ is asymptotically optimal for $n = [\alpha N]$, $N \to \infty$. However, it requires the construction of the sets $\mathcal{X}^*_j$, $j = 1, 2, 3$, and thus of a $\phi$-optimum constrained design measure $\xi^*_\alpha \leq \mu/\alpha$. We show in the next section that it is possible to avoid this construction while preserving asymptotic optimality.

**Remark 7** A deterministic version of (18) consists in accepting a $X_k$ that falls into $\mathcal{X}^*_3$ only when $a_k/k < \alpha$. One can then easily show that $a_k/k \to \alpha$ and $\phi_k \to \phi^*_\alpha$ as $k \to \infty$, $\mu$-a.s. (however, the strategy is then non stationary, see Section 2.1).

4 Sampling asymptotically from a constrained measure

Consider the following modification of the strategy $\mathcal{S}(\xi^*_\alpha)$ defined by (18): at step $k$, we simply substitute the sets $\mathcal{X}_{j,\alpha}(\xi_{k-1})$ for $\mathcal{X}^*_j$, $j = 1, 2, 3$, with $\mathcal{X}_{j,\alpha}(\xi)$ defined by (15) and $\xi_{k-1}$ the empirical measure defined by the $a_k$ design points
already selected. We thus define the following adaptive strategy:

\[
S_\alpha(\mu) : \begin{cases} 
\text{if } X_k \in \mathcal{X}_{2,\alpha}(\xi_{k-1}) \text{ accept } X_k; \\
\text{if } X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1}) \text{ accept } X_k \text{ with probability } P_k(\alpha) = \frac{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi_{k-1})]}{\mu[\mathcal{X}_{3,\alpha}(\xi_{k-1})]}; \\
\text{otherwise reject } X_k.
\end{cases}
\] (19)

**Remark 8**

1. Notice that the sets \( \mathcal{X}_{j,\alpha}(\xi_{k-1}) \) are obtained from the function \( F_\phi(\xi_{k-1}, x) \), that is, from \( M(\xi_{k-1}) \). Hence, the (randomized) decision at step \( k \) only depends on \( X_k \) and \( M(\xi_{k-1}) \) and \( S_\alpha(\mu) \) is stationary.

2. In practice, the first samples \( X_k \) are always accepted until \( M(\xi_{k-1}) \) becomes nonsingular, but this initialization has no effect on the asymptotic behavior of \( S_\alpha(\mu) \). Also, for technical reasons, we can assume that \( X_k \) is always accepted when \( a_k/k < \alpha/C \), with \( C \) an arbitrarily large constant. This has no practical importance for the asymptotic behavior of the strategy (since \( a_k/k \to \alpha \) \( \mu \)-a.s., see Theorem 9 below) and for that reason is not mentioned in the definition of \( S_\alpha(\mu) \). On the other hand, it implies \( a_{k+1}/k > \alpha/C \) for any \( k \) which permits to guarantee that \( \mathbb{E}\{\phi_k\} < \infty \), a property used in the proof of asymptotic optimality, see Theorem 9.

3. \( S_\alpha(\mu) \) takes a simpler form when \( F_\xi(s) \) given by (13) is continuous in \( s \) for any \( \xi \) (so that \( \mathcal{X}_{3,\alpha}(\xi_{k-1}) \) is always empty): we accept \( X_k \) when \( F_{\xi_{k-1}}(X_k) > 1 - \alpha \) and reject \( X_k \) otherwise, and the strategy is fully deterministic.

4. We observe the same behavior when \( S_\alpha(\mu) \) is modified as follows: any \( X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1}) \) is accepted if \( a_k/k < \alpha \) and is rejected otherwise; see Remark 7.
However, the strategy is then non stationary and its asymptotic behavior is more difficult to analyse.

5. Obviously, the sequence \((\phi_k)\) generated by (19) is not monotonically increasing, since (i) the step-length \(1/(1 + a_k)\) when \(X_k\) is accepted and \(\xi_{k-1}\) 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 \(H_{\mu,3}\) on \(\mu\), see Appendix A.

The asymptotic behavior of (19) satisfies the following.

**Theorem 9** Under \(H_{\Phi1-H_{\Phi6}}\) and \(H_{\mu1-H_{\mu3}}\) the empirical measure \(\xi_k\) defined by the points accepted by the stationary strategy \(S_\alpha(\mu)\) satisfies \(a_k/k \to \alpha\), \(\mu\)-a.s., and

\[
\lim_{k \to \infty} \phi(\xi_k) = \phi(\xi_\alpha^*), \mu\text{-a.s.,}
\]

as \(k \to \infty\), with \(\xi_\alpha^* \leq \mu/\alpha\) a \(\phi\)-optimum constrained design measure.

The proof is given in Appendix B. When the strategy is truncated as indicated in (8), the resulting \(S_{N,n}[S_\alpha(\mu)]\) is asymptotically optimal for \(n = \lfloor \alpha N \rfloor\), \(N \to \infty\), from Theorem 1. Another possible truncation consists in adapting \(\alpha\) and using, at step \(k\), the strategy \(S_{\alpha_k}(\mu)\), with

\[
\alpha_k = \frac{n - a_k}{N - k + 1},
\]

18
and \( a_k \) defined by (4). \( S_{\alpha_k}(\mu) \) coincides with the one-step-ahead rule suggested in (Pronzato, 2001b): when \( \alpha_k = 0 \), \( X_{2,\alpha_k}(\xi_k-1) \) is empty, \( \mu[X_{2,\alpha_k}(\xi_k-1)] = 0 \) and \( X_k \) is always rejected; when \( \alpha_k \geq 1 \), \( X_{2,\alpha_k}(\xi_k-1) = \mathcal{X} \) and \( X_k \) is always accepted.

When \( \mu \) is unknown, we can use \( S_{\alpha}(\hat{\mu}_k) \), or \( S_{\alpha_k}(\hat{\mu}_k) \), at step \( k \), with \( \hat{\mu}_k \) the empirical version of \( \mu \) (or a kernel estimate, or a parametric representation \( \mu_{\hat{\beta}_k} \), with \( \hat{\beta}_k \) estimated from \( X_1, \ldots, X_k \)). The estimation of \( \mu \) 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 \( \xi^*_\alpha \), without constructing \( \xi^*_\alpha \) beforehand and even without knowing \( \mu \) 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 \( \theta \), so that local optimum design is based on \( \mathbf{M}(\hat{\theta}^0) \) with \( \hat{\theta}^0 \) a nominal value for \( \theta \). When \( \theta \) can be estimated on line, it is natural to replace \( \hat{\theta}^0 \) by an estimate, \( \hat{\theta}^k \) at step \( k \). When using least squares estimation in a nonlinear regression problem with independent observations \( Y_k = \eta(X_k, \bar{\theta}) + \varepsilon_k \), where \( (\varepsilon_k) \) is a sequence of i.i.d. errors with zero mean and finite variance, consistency and asymptotic normality of \( \hat{\theta}^k \) will hold under \( \mathbf{H}_\mu 3 \) (and additional conditions on higher order derivatives of \( \eta(\theta, x) \) with respect to \( \theta \) and their tail cross product, see Jennrich (1969)). The sampling strategy \( S_{\alpha}(\mu) \) will then ensure \( \phi_k \rightarrow \phi[\xi^*_\alpha(\bar{\theta})], \mu\text{-a.s., } k \rightarrow \infty, \) with \( \xi^*_\alpha(\bar{\theta}) \leq \mu/\alpha \) a \( \phi \)-optimum constrained design measure for the true value \( \bar{\theta} \) of the model parameters. This is illustrated by Example 12 below.
5 Examples

We take $\Phi(\cdot) = \log \det(\cdot)$ in all the examples below. Note that all the conditions $H_\Phi$ are then satisfied, see Appendix A.

**Example 10** We consider the quadratic regression model $\eta(\theta, X) = \theta_0 + \theta_1 X + \theta_2 X^2$. Let $\mu_c$ correspond to the normal distribution $N(0, 1)$ and $\mu_d$ 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 $\mu = 0.5\mu_c + 0.5\mu_d$ and $\alpha = 0.1$.

Easy calculations show that the $\phi$-optimum constrained measure $\xi^*_\alpha$ is equal to $\mu/\alpha$ on $X^*_\alpha = (-\infty, -a] \cup [a, \infty)$, with $a \approx 1.5625$, and puts the rest of its weight, approximately 0.4091, at zero. Figure 1 presents a plot of the sensitivity function $d(\xi^*_\alpha, x) = F_\Phi(\xi^*_\alpha, x) + 3 = f^\top(x)M^{-1}(\xi^*_\alpha)f(x)$ and illustrates the optimality of $\xi^*_\alpha$. Figure 2, left, gives a histogram of the first 1000 samples accepted by $S_{\alpha}(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the $X_k$’s. The right part of the figure presents $\phi_k$ as a function of $k$ (log scale), with the optimal value $\phi^*_{0.1}$ indicated by the dashed line. (Note the fast convergence of $\phi_k$, the large value of $N$ being required only to have enough points for the histogram plot.)

**Example 11** We consider again the case of a measure $\mu$ 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 $X$ is two-dimensional, $X = (x_1, x_2)$. The continuous component $\mu_c$ corresponds to the normal distribution $N(0, I_2)$, with $I_2$ the 2-dimensional identity matrix, the discrete component $\mu_d$ puts weight $1/4$ at each one of the points $(\pm 1, \pm 1)$. Define $B(a) = \{x \mid \|x\| > a\}$. When $\mu = (1/2)(\mu_c + \mu_d)$ the results are as follows.
Fig. 1. Sensitivity function $d(\xi^*_a, x)$ for the optimal constrained measure $\xi^*_a$ in Example 10.

Fig. 2. Left: histogram of the 1000 first samples $X_k$ accepted by $S_\alpha(\hat{\mu}_k)$ in Example 10. Right: $\phi_k$ as a function of $k$ (log scale); the value of $\phi^*_0.1$ corresponds to the dashed line.
Fig. 3. $\phi^*_\alpha$ as a function of $\alpha$ in Example 11.

For $0 < \alpha \leq 1/(2e)$, with $e = \exp(1)$, $\xi^*_\alpha = \text{coincides with } \mu_c/(2\alpha)$ on $B(a_\alpha)$ with $a_\alpha = \sqrt{2 \log[1/(2\alpha)]}$ \geq \sqrt{2}; $\phi^*_\alpha = 2 \log[1 - \log(2\alpha)]$.

For $1/(2e) < \alpha \leq 1/(2e) + 1/2$, $\xi^*_\alpha = \mu/\alpha$ on $B(\sqrt{2})$ and $\xi^*_\alpha = 2[\alpha - 1/(2e)]\mu/\alpha$ on the four points $(\pm1, \pm1)$; $\phi^*_\alpha = 2 \log[1 + 1/(2e\alpha)]$.

For $1/(2e) + 1/2 < \alpha \leq 1$, $\xi^*_\alpha = \mu/\alpha$ on $B(b_\alpha)$ with $b_\alpha = \sqrt{2 \log[1/(2\alpha - 1)])} < \sqrt{2}; \phi^*_\alpha = 2 \log\{1 - [(2\alpha - 1) \log(2\alpha - 1)]/(2\alpha\})

Figure 3 presents $\phi^*_\alpha$ as a function of $\alpha$ (although it does not appear clearly from the figure, $\phi^*_\alpha$ tends to infinity when $\alpha$ tends to zero). Here $c_\alpha(\xi_\alpha)$ is a continuous function of $\alpha$, so that $\phi^*_\alpha$ is differentiable with respect to $\alpha$, see Pronzato (2004).

We take $N = 10000$, $n = 2000$ and use $S_{\alpha_k}(\hat{\mu}_k)$, see (19,21), with $\hat{\mu}_k$ the empirical measure of the $X_k$'s. (Note that $\alpha = 0.2$ is related to the intermediate case $1/(2e) < \alpha \leq 1/(2e) + 1/2$.) In the simulation presented $M(\xi_k)$ has full rank after the first three samples $X_1, X_2, X_3$ have been accepted, and Figure 4 (top) presents a histogram of the remaining 1997 points accepted by $S_{\alpha_k}(\hat{\mu}_k)$. 
Fig. 4. Top: histogram of the 2000 points accepted by $S_{\alpha k}(\hat{\mu}_k)$ in a sequence of 10,000 points; bottom: $\phi_k$ as a function of $k$ (log scale), the value of $\phi_{0.2}^*$ corresponds to the dashed line (Example 11).

Figure 4 (bottom) gives $\phi_k$ as a function of $k = 4, \ldots, 2000$ (log scale), with the optimal value $\phi_{0.2}^* \simeq 1.3043$ indicated by the dashed line. (Again, notice the fast convergence of $\phi_k$, the large value of $N$ being required only to have enough points for the histogram plot.)

**Example 12** We consider now the nonlinear regression model used in (Box and Lucas, 1959), where

$$ \mathbb{E}\{Y_k | X_k = x; \theta\} = \eta(\theta, x) = \frac{\theta_1}{\theta_1 - \theta_2} \left[ \exp(-\theta_2 x) - \exp(-\theta_1 x) \right]. $$

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 $D$-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 $N(1, 0.25)$. Figure 5 gives a histogram of the first 2,500 points accepted by $S_\alpha(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the $X_k$’s, when $\alpha = 0.5$. Easy calculations show that $\xi_\alpha^*(\hat{\theta}) = 2\mu$ on $[0, a] \cup [b, \infty)$, with $a \simeq 1.996$ and $b \simeq 3.922$, and $\phi[\xi_\alpha^*(\hat{\theta})] \simeq -2.143$.

6 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 example is when one observes covariates $Z_k$ 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.
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 $X_k = (U_k, U_{k-1}, \ldots, U_{k-m})$, with $(U_i)$ a random input sequence for the system. Note, however, that when the model contains an autoregressive part, that is, when $X_k = (U_k, U_{k-1}, \ldots, U_{k-m}, Y_{k-1}, \ldots, Y_{k-l})$, 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 $S_{\alpha_k}(\mu)$, see (19, 21), is proved in (Pronzato, 2001a) for $N \to \infty$ with $n$ fixed, provided the distribution function of $X$ is a von Mises function (see, e.g., Embrechts et al. 1997, p. 138), with a tail decreasing faster than any power law. Extending Theorem 1 to the situation where $n$ is fixed but $d > 1$ remains an open issue. Note in particular that in this case, although (12) 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 $n = [\alpha N]$, such as $n = [\log N]$, or $[N^\beta]$ with $\beta < 1$, etc., are of interest. A possible application concerns the construction of optimum design algorithms. Indeed, classical algorithms for the determination of $\xi^\ast$ (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 $\xi_{k-1}$ 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 $\delta$ 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 $\mu$, with an acceptance rule such as (19), will ensure the convergence of $\xi_k$ to $\xi^*$, provided $\alpha$ 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)

$H_\Phi 1$: $\Phi$ is strictly concave and $\Phi(M) > -\infty$ for non singular $M$.

$H_\Phi 2$: $\Phi$ is linearly differentiable; that is, the directional derivative

$$F_\Phi(M_1, M_2) = \lim_{\epsilon \to 0^+} \{\Phi[(1 - \epsilon)M_1 + \epsilon M_2] - \Phi(M_1)/\epsilon$$

satisfies $F_\Phi(\xi_1; \xi_2) = \mathcal{F}_\Phi[M(\xi_1), M(\xi_2)] = \int_X F_\Phi(\xi_1, x)\xi_2(dx)$ for any $\xi_1, \xi_2$ with $\phi(\xi_1) = \Phi[M(\xi_1)] > -\infty$, where $F_\Phi(\xi, x) = F_\Phi(\xi; \delta_x)$ and $\delta_x$ is the Dirac measure supported at $x$.

$H_\Phi 3$: $\Phi$ is increasing: $M_2 - M_1$ non negative definite implies $\Phi(M_2) \geq \Phi(M_1)$.

$H_\Phi 4$: $\lambda_{\min}(M) \geq l > 0$ implies $\mathcal{F}_\Phi(M, ff^T) \geq -\delta_l(l)$ for any $f$, for some $\delta_l(l) < \infty$.

$H_\Phi 5$: there exist a function $g_1(\cdot)$ from $\mathbb{R}^+$ to $\mathbb{R}$ and a function $g_2(\cdot)$ from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that $\lim_{a \to +1} g_1(a) = 0$, $\lim_{a \to +1} g_2(a) = 1$ and $\Phi(aM) \leq g_1(a) + g_2(a) \Phi(M)$ for any non-negative definite $M$ and any $a \geq 1$. Note that this assumption is satisfied for homogeneous criteria with $g_1 = 0$ and $g_2$ the identity. Also, if one wishes that maximizing $\Phi(M)$ be equivalent to maximiz-
ing $\Phi(aM)$ (which is often implicitly assumed when approximate designs are used), the assumption $\Phi(aM) = g_1(a) + g_2(a) \Phi(M)$ comes naturally.

**H₆:** $\Phi$ is two times differentiable. We denote by $\nabla^2_{\Phi}(M_1, M_2)$ the second order directional derivative

$$\nabla^2_{\Phi}(M_1, M_2) = \frac{\partial^2 \Phi[(1 - \gamma)M_1 + \gamma M_2]}{\partial \gamma^2} \bigg|_{\gamma=0^+}.$$ 

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

**H₁:** $-\infty < \phi(\mu) < \infty$.

**H₂:** $\mu_4 = \int x^4 \mu(dx) < \infty$. Note that Chebyshev’s inequality implies $\text{Prob}\{\|f(X_1)\| \geq R\} \leq \mu_4/R^4$.

**H₃:** For $\alpha \in (0, 1)$ the proportion of interest, there exists $l > 0$ and $\epsilon \in (0, \alpha)$ such that for any design measure $\xi \leq \mu/(\alpha - \epsilon)$, $\lambda_{\min}[M(\xi)] > l$.

**Discussion of the assumptions.**

In the case of $D$-optimality, where $\Phi(M) = \log \det M$, we have $F_\Phi(\xi, x) = f^T(x)M^{-1}(\xi)f(x) - d$ with $d = \dim(\theta)$. $\textbf{H}_Φ1$-$\textbf{H}_Φ4$ are satisfied, with $\delta_1(l) = d$ in $\textbf{H}_Φ4$. $\textbf{H}_Φ5$ is satisfied with $g_1(a) = d \log a$ and $g_2(a) = 1$. Finally, $\nabla^2_{\Phi}(M_1, M_2) = -\text{trace}\{[M_1^{-1}(M_2 - M_1)]^2\}$ so that

$$\nabla^2_{\Phi}(M, ff^T) = -(f^TM^{-1}f)^2 + 2(f^TM^{-1}f) - d$$

and $\textbf{H}_Φ6$ is satisfied with $\delta_0 = d$, $\delta_2(l) = 1/l^2$.

Take now $\Phi(M) = -\text{trace}(AM^{-p})$, with $p$ a positive integer and $A$ a positive definite matrix (Kiefer’s $\Phi_p$-class of optimality criteria). We can as-
sume that $A = I_d$, the $d$-dimensional identity matrix, by a linear transformation on the set $M(\Xi) = \{M(\xi), \xi \in \Xi\}$. We have $F(M_1, M_2) = p \text{trace}[M_1^{-(p+1)}(M_2 - M_1)]$. $H_5.1$-$H_5.4$ are satisfied, with $\delta_1(l) = p/l^p$ in $H_5.4$. $H_5.5$ is satisfied with $g_1(a) = 0$ and $g_2(a) = a^{-d}$. Finally, $\nabla^2_{\phi}(M_1, M_2) = -p \sum_{a+b=p+2, a,b \geq 1} \text{trace}[M_1^{-a}(M_2 - M_1)M_1^{-b}(M_2 - M_1)]$, see Wu and Wynn (1978), and easy calculations give $\nabla^2_{\phi}(M, f_{\phi}^T) = -p \sum_{a+b=p+2, a,b \geq 1} ((f^TM^{-b}f) (f^TM^{-a}f)] + 2p(p+1)f^TM^{-(p+1)}f - p(p+1) \text{trace}[M^{-p}]$. Therefore, $H_5.6$ is satisfied with $\delta_0(l) = p(p+1)/l^p$, $\delta_2(l) = p/l^{p+2}$.

$H_5.1$-$H_5.3$ are satisfied for instance when: (i) the functions $f(x)$ are linearly independent on any open subset of $X$ with finite Lebesgue measure, (ii) $\mu$ has a component $\mu_c$ absolutely continuous with respect to the Lebesgue measure, with density $\varphi$, and a finite number of discrete components, (iii) the mass of $\mu_c$ is larger than $1 - \alpha + \epsilon$ (that is, the mass of the discrete components is less than $\alpha - \epsilon$), and (iv) $X$ is compact or $\varphi(x)$ is exponentially decreasing when $\|x\| \to \infty$.

B Appendix (Proofs)

Proof of Lemma 2. From $H_3$, for $n_1 \leq n_2 \leq N$, $\Psi(S_{N,n_1}^c) = \Phi(M_{N,n_1}^c) \leq \Phi[(n_2/n_1)M_{N,n_2}^c]$, with $M_{N,n}^c$ the normalized information matrix associated with the $n$ points selected by $S_{N,n}^c$, see (2). Since $\alpha_N \to \alpha$, for any $\epsilon$ 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 $N > N_0$,

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

Denote \(a_\epsilon = (\alpha + \epsilon)/(\alpha - \epsilon)\), notice that \(a_\epsilon > 1\) for \(\alpha > 0\) and \(a_\epsilon \to 1\) when \(\epsilon \to 0\). H5 then gives
\[
\frac{\Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]}) - g_1(a_\epsilon)}{g_2(a_\epsilon)} \leq \Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]}) \leq g_2(a_\epsilon) \Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]}) + g_1(a_\epsilon)
\]
and thus, from (10),
\[
\lim_{N \to \infty} \Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]}) \leq g_2(a_\epsilon) \phi^*_\alpha + g_1(a_\epsilon)
\]
\[
\frac{\phi^*_\alpha - g_1(a_\epsilon)}{g_2(a_\epsilon)} \leq \liminf_{N \to \infty} \Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]})
\]
\(\mu\)-a.s. The continuity of \(\phi^*_\alpha\) with respect to \(\alpha\) (Theorem 6) and \(\lim_{n \to 1^+} g_1(a) = 0, \lim_{n \to 1^+} g_2(a) = 1\) give \(\lim_{N \to \infty} \Phi(M^*_{N,[(\alpha-\epsilon)N],(\alpha+\epsilon)N]}) = \phi^*_\alpha, \mu\)-a.s. \(\square\)

**Proof of Lemma 3.** The strategy \(S^*_{N,n}\) satisfies \(\Psi(S^*_{N,n}) = \Phi(M^*_{N,n})\) with
\[
M^*_{N,n} = \frac{1}{n} \sum_{i=1}^{N} f(X_i) f^\top(X_i) J_n[X_j(j = 1, \ldots, N, j \neq i), X_i],
\]
where \(J_n \in \{0, 1\}\) defines the decision function (the decision to accept \(X_i\) depends on the rest of the sequence). Therefore,
\[
E\{M^*_{N,n}\} = \frac{1}{n} \int_{\mathcal{X}^N} \left\{ \sum_{i=1}^{N} f(x_i) f^\top(x_i) J_n[x_j(j = 1, \ldots, N, j \neq i), x_i] \right\} \prod_{j=1}^{N} \mu(dx_j).
\]
Using the fact that decisions are invariant by any permutation of the sequence, we get
\[
E\{M^*_{N,n}\} = \int_{\chi} f(x) f^\top(x) \xi^*_{N,n}(dx) = M(\xi^*_{N,n})
\]
where
\[
\xi^*_{N,n}(dx) = n^{-1} \left( \int_{\mathcal{X}^{N-1}} J_n[x_j(j = 1, \ldots, N - 1), x] \prod_{j=1}^{N-1} \mu(dx_j) \right) \mu(dx).
\]
Since $J_n \in \{0, 1\}$, $\xi_{N,n}(dx)$ satisfies $\xi_{N,n}(dx) \leq (N/n) \mu(dx)$. This implies $\Phi[M(\xi_{N,n}^*)] \leq \phi(\xi_{n/N})$. Concavity of $\Phi(\cdot)$ finally gives

$$E[\Phi(M_{N,n}^*)] \leq \Phi[E(M_{N,n}^*)] \leq \phi(\xi_{n/N})$$

which concludes the proof.  

**Proof of Theorem 9.**

Since $X_k$ is accepted with probability $\alpha$, $a_k/k \to \alpha$ $\mu$-a.s. as $k \to \infty$.

The rest of the proof is decomposed in three steps. In (i), we construct a lower bound on $E(\phi_k|X_k^{k-1})$, where $X_k^{k-1}$ denotes the collection $X_1, \ldots, X_{k-1}$. In (ii) we show that $\limsup_{k \to \infty} \phi_k = \phi_0^*, \mu$-a.s. Finally, in (iii) we show that $\liminf_{k \to \infty} \phi_k = \phi_0^*, \mu$-a.s., by an approach similar to Doob’s upcrossing Lemma, see Williams (1991, p. 108).

(i) From $H_\Phi 2$, $H_\Phi 6$, criterion value $\phi_k$ associated with the empirical design measure $\xi_k$ generated by $S_\alpha(\mu)$ satisfies the recurrence

$$\phi_k = \phi_{k-1} + \frac{1}{1 + a_k} F\Phi(\xi_{k-1}, X_k) + \frac{1}{2(1 + a_k)^2} H\Phi(\xi_{k-1}, X_k, \gamma_k)$$

$$\times \left[ I_{X_{\gamma_k}}(X_k) + I_{[0, P_k(\alpha)]}(Z) I_{X_{\gamma_k}}(X_k) \right], $$

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

$$H\Phi(\xi_{k-1}, X_k, \gamma_k) = \nabla^2 \Phi[(1 - \gamma_k) \mathbf{M}(\xi_{k-1}) + \gamma_k f(X_k)f^T(X_k), f(X_k)f^T(X_k)]$$

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

Since $a_k/k \to \alpha$ $\mu$-a.s., from $H_\mu 3$ there exists $K_0$ ($\mu$-a.s.) such that for any $k > K_0$, $\lambda_{\min}[\mathbf{M}(\xi_{k-1})] > l/2$. Since $\lambda_{\min}[(1 - \gamma_k) \mathbf{M}(\xi_{k-1}) + \gamma_k f(X_k)f^T(X_k)] \geq$

30
\[ 1 - 1/(1 + a_k) \lambda_{\min}[M(\xi_{k-1})], \] there exists \( K_1 \) (\( \mu \)-a.s.) such that for \( k > K_1, \)
\[ \lambda_{\min}[1 - \gamma_k M(\xi_{k-1}) + \gamma_k f(X_k) f^T(X_k)] > 1/4. \] From \( \mathbf{H}_\Phi \), this implies \( H_\Phi(\xi_{k-1}, X_k, \gamma_k) > -\delta_0(1/4) - \delta_2(1/4) \| f(X_k) \|^4 \). We can now compute a lower bound on \( \mathbb{E}\{\phi_k | X_1^{k-1}\} \). Notice that

\[
\mathbb{E}\{F_\Phi(\xi_{k-1}, X_k) \left[ I_{X_{k+1}}(X_k) + I_{0,P_k(\alpha)}(Z)I_{X_{k+1}}(X_k) \right] | X_1^{k-1}\} = \alpha F_\Phi[\xi_{k-1}; T_{\Phi,\alpha}(\xi_{k-1})]
\]

which gives for \( k > K_1, \)

\[
\mathbb{E}\{\phi_k | X_1^{k-1}\} \geq \phi_{k-1} + \frac{\alpha}{1 + a_k} F_\Phi[\xi_{k-1}; T_{\Phi,\alpha}(\xi_{k-1})] - \frac{\delta_0(1/4) + \delta_2(1/4) \mu_4}{2(1 + a_k)^2}
\]

with \( \mu_4 = \int_X \| f(x) \|^4 \mu(dx) \), and \( \mu_4 < \infty \) from \( \mathbf{H}_\mu \). Using (17), we obtain

\[
\mathbb{E}\{\phi_k | X_1^{k-1}\} \geq \phi_{k-1} + \frac{\alpha}{1 + a_k} (\phi^*_\alpha - \phi_{k-1}) - \frac{A}{(1 + a_k)^2} \tag{B.2}
\]

for \( k > K_1 \), where \( A = \alpha [\delta_0(1/4) + \delta_2(1/4) \mu_4]/2 \).

(ii) Assume that \( \limsup_k \phi_k < \phi^*_\alpha - \delta \) for some \( \delta > 0 \), (B.2) gives

\[
\mathbb{E}\{\phi_k | X_1^{k-1}\} > \phi_{k-1} + \frac{\alpha \delta}{2(1 + a_k)} \tag{B.3}
\]

for \( k \) larger than some \( K_2 \). Since \( \phi_k < \Psi(S^*_{k,a_{k+1}}) \) and \( a_{k+1}/k > \alpha/C \) for any \( k \), see Remark 8.2, we have \( -\infty \leq \mathbb{E}\{\phi_k\} < \mathbb{E}\{\Psi(S^*_{k,a_{k+1}})\} < \phi(\xi_{a_{k+1}/k}) < \phi(\xi^*_{\alpha/C}) < \infty \) (see Lemma 3, \( \mathbf{H}_\Psi \), \( \mathbf{H}_\Phi \)) and \( \mathbb{E}\{\phi_k\} \) is well defined. Also, for \( k > K_2 \), (B.3) gives \( \mathbb{E}\{\phi_k\} > \mathbb{E}\{\phi_{K_2}\} > -\infty \) and thus \( \sup_k \mathbb{E}\{|\phi_k|\} < \infty \).

The martingale convergence theorem then says that \( \phi_k \) converges \( \mu \)-a.s. to a finite limit, which contradicts (B.3). Repeating the same arguments for a countable number of rational \( \delta \) gives \( \limsup_{k \to \infty} \phi_k \geq \phi^*_\alpha \). At the same time,
\[
\limsup_{k \to \infty} \phi_k \leq \lim_{k \to \infty} \Psi(S^*_{k,a_{k+1}}) = \phi^*_\alpha \text{ } \mu\text{-a.s. from Lemma 2, so that}
\]
\[
\limsup_{k \to \infty} \phi_k = \phi^*_\alpha, \text{ } \mu\text{-a.s.}
\]  \hspace{1cm} (B.4)

(iii) Assume that \(\liminf_{k \to \infty} \phi_k < \phi^*_\alpha - \epsilon\) for some \(\epsilon > 0\). We show that this event has probability zero.

First, we show that there exists \(K^* (\mu\text{-a.s.})\) such that for any \(k > K^*\),

\[
\phi_{k-1} < \phi^*_\alpha + \epsilon/6, \tag{B.5}
\]
\[
(\epsilon/6)[\alpha/(1 + a_k)] - A/(1 + a_k)^2 > 0, \tag{B.6}
\]
\[
\phi_k - \phi_{k-1} > -\epsilon/6. \tag{B.7}
\]

(B.5) follows from (B.4) and (B.6) from \(a_k/k \to \alpha \text{ } \mu\text{-a.s.} \). \(H_4\) implies that \(F_\Phi(\xi_{k-1}, X_k) \geq -\delta_1(l/2)\) for \(k > K_0\) (since \(\lambda_{\min}[M(\xi_{k-1})] > l/2\) for \(k > K_0\)) and (B.1) gives

\[
\phi_k - \phi_{k-1} \geq \frac{-\delta_1(l/2)}{1 + a_k} - \frac{\delta_0(l/4) + \delta_2(l/4)\|f(X_k)\|^4}{2(1 + a_k)^2}
\]

for \(k > K_1\). \(H_2\) implies \(\text{Prob}\{|\|f(X_k)\| > k^\gamma\} \leq \mu_4/k^{4\gamma}\) and therefore, from Borel-Cantelli, \(\text{Prob}\{|\|f(X_k)\| > k^\gamma \text{ infinitely often }\} = 0\) for any \(\gamma > 1/4\). It implies that exists \(K_3 (\mu\text{-a.s.})\) such that for any \(k > K_3\)

\[
\phi_k - \phi_{k-1} \geq \frac{-\delta_1(l/2)}{1 + a_k} - \frac{\delta_0(l/4) + \delta_2(l/4)k^{4\gamma}}{2(1 + a_k)^2}.
\]

Take \(\gamma < 1/2\), (B.7) then follows from \(a_k/k \to \alpha \text{ } \mu\text{-a.s.}\). We work conditionally on \(k > K^*\) in the rest of the proof.

Take \(\epsilon > 0\) arbitrarily small and define \(U^*_\xi\) as the number of down-crossings by \(\phi_k\) of the interval \((\phi^*_\alpha - \epsilon, \phi^*_\alpha - \epsilon/6)\) for \(k > K^*\): \(\liminf_{k \to \infty} < \phi^*_\alpha - \epsilon\) thus implies \(U^*_\infty = \infty\). Also define a previsible process \((C_k)\) as follows: set \(C_k\) to one when \(\phi_{k-1}\) gets above \(\phi^*_\alpha - \epsilon/6\) (strictly) and turn it to zero when \(\phi_{k-1}\)
gets below $\phi^*_\alpha - \epsilon$ (strictly). This defines a new process

$$V_k = \sum_{i=K^*}^{k} C_i(\phi_i - \phi_{i-1})$$

which satisfies

$$V_k < -(5\epsilon/6) U_k^\epsilon + \max\{0, \phi_k - (\phi^*_\alpha - \epsilon/6)\}.$$  \hspace{1cm} (B.8)

Define $k_j, \ldots, \bar{k}_j$ as the $j$th set of consecutive indices for which $C_i = 1$, that is, such that $C_{k_j-1} = 0, C_{k_j} = C_{k_j+1} = \ldots = C_{\bar{k}_j} = 1, C_{\bar{k}_j+1} = 0$. Within this set, we define $k^*_j$ as the largest index such that $\phi_{k^*_j-1} \leq \phi^*_\alpha - \epsilon/6$ for all $k^*_j \leq k \leq \bar{k}_j$. Note that necessarily $k^*_j > k_j$ (since $\phi_{k_j-1} > \phi^*_\alpha - \epsilon/6$) and $k^*_j < \bar{k}_j$ (since (B.7) is satisfied for $k > K^*$). We bound $\mathbb{E}\{V_{k_j} - V_{k_j}|X^{k_j-1}_1\}$ as follows.

We have $\phi_{k^*_j-2} > \phi^*_\alpha - \epsilon/6$ (from the definition of $k^*_j$) and $\phi_{k^*_j-1} - \phi_{k^*_j-2} > -\epsilon/6$ from (B.7). Also, (B.5) implies $\phi_{k_j-1} < \phi^*_\alpha + \epsilon/6$. Therefore, $\phi_{k^*_j-1} - \phi_{k_j-1} > -\epsilon/2$. Now, for all $k$ such that $k^*_j \leq k \leq \bar{k}_j$, $\phi_{k-1} \leq \phi^*_\alpha - \epsilon/6$ and thus $\mathbb{E}\{\phi_k - \phi_{k-1}|X^{k-1}_1\} > 0$ from (B.6) and (B.2). We obtain $\mathbb{E}\{V_{k_j} - V_{k_j}|X^{k_j-1}_1\} > -\epsilon/2$ and $\mathbb{E}\{V_k|X^{K^*-1}_1\} > -(\epsilon/2)\mathbb{E}\{U_k^\epsilon|X^{K^*-1}_1\}$, which, together with (B.8) gives

$$\mathbb{E}\{U_k^\epsilon|X^{K^*-1}_1\} < \frac{3}{\epsilon}\mathbb{E}\{\max\{0, \phi_k - (\phi^*_\alpha - \epsilon/6)\}|X^{K^*-1}_1\} < \infty$$

where the last inequality follows from $\mathbb{E}\{|\phi_k|\} < \infty$. It implies $\text{Prob}\{U_\infty = \infty\} = 0$, that is, $\lim\inf_{k \to \infty} \phi_k > \phi^*_\alpha - \epsilon$, $\mu$-a.s. Repeating the same arguments for a countable number of rational $\epsilon$ gives $\lim\inf_{k \to \infty} \phi_k = \phi^*_\alpha$, $\mu$-a.s., which completes the proof. $\Box$
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 very helpful, and the comments of two anonymous referees have been extremely useful to improve the paper.

References


