DESIGNING FROM A SEQUENCE OF I.I.D. EXPERIMENTS

LUC PRONZATO

ABSTRACT. We consider a regression problem, with observations $Y_k = \eta(\theta, X_k) + \epsilon_k$, where (ϵ_k) is a sequence of independent measurement errors and where the experimental conditions X_k form a sequence of independent random variables distributed with a probability measure μ, independent of (ϵ_k) and observed sequentially. The length of the sequence (X_k) is N but only $n < N$ observations can be made. As soon as a new experiment X_k is available, one must decide whether to observe Y_k or not, the objective being to estimate the parameters θ as precisely as possible. The optimal rule for the on-line selection of the X_k's can be constructed when θ is scalar, see [1], and suboptimal rules have been suggested for the case $p = \dim(\theta) > 1$ [2, 3]. We propose here a different solution, based on the construction of an optimal constrained design measure, and show that it is asymptotically optimal ($n = \lfloor \alpha N \rfloor$, $\alpha \in (0,1)$, $N \to \infty$). As a byproduct, we obtain a procedure that asymptotically samples from an optimal constrained measure $\xi_\alpha^* \leq \mu/\alpha$, without requiring neither the determination of ξ_α^* nor the knowledge of the measure μ.

1. INTRODUCTION

Consider a regression model, with observations

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

where the errors ϵ_k are independent with $\mathtt{E}\{\epsilon_k\} = 0$, $\mathtt{E}\{\epsilon_k^2\} = 1$, and $\bar{\theta} \in \Theta$ is the unknown true value of the model parameters to be estimated, with Θ an open subset of $I\!R^p$. The function $\eta(\theta, x)$ is assumed continuously differentiable in θ, uniformly in $x \in \mathcal{X} \subseteq I\!R^q$. We shall write $f(x) = \partial\eta(\theta, x)/\partial\theta_{\hat{\theta}^0}$, with $\hat{\theta}^0$ a given nominal value for θ, and assume that $f(\cdot)$ is continuous in x on $\mathcal{X}$ ($\hat{\theta}^0$ need not be specified when $\eta(\theta, x)$ is linear in x).

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. As soon as a value X_k becomes available, one must decide whether to observe Y_k or not, in order to estimate θ as precisely as possible.

Let μ denote the probability measure of X_1, 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), $0 \leq k \leq n$. We assume that μ is atomless, that is, 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 particular cases). The decision sequence will be denoted by (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,

$$u_j \in \mathcal{U}_j \subseteq \{0, 1\} \,, \; j = 1, \ldots, N \,, \; \sum_{j=1}^{N} u_j = n \,. \tag{2}$$

Note that X_k is known when u_k is chosen.

We consider design criteria $\Phi(\cdot)$ for the estimation of θ that are increasing functions of the information matrix, and we wish to maximise $\Phi[\mathbf{M}_N/n]$, with $\mathbf{M}_N = \sum_{k=1}^{N} u_k f(X_k) f^\top(X_k)$. Notice that $\mathbf{M}_N$ depends on $\hat{\theta}^0$ when $\eta(\theta, x)$ is nonlinear in θ (local design). We assume that μ is such that

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

Date: August 29, 2002
2000 Mathematics Subject Classification: Primary: 62K05; Secondary: 62L05.
Key words: sequential design, sampling, constrained design measure.

exists, with $-\infty < \Phi[\mathbf{M}(\mu)] < \infty$, and that $\mathtt{E}\{\Phi(\mathbf{M}_N/n)\}$ exists for any N and $n \geq p$ and any $\mathcal{F}_n$-measurable sequence (u_n), the expectation $\mathtt{E}\{\cdot\}$ being with respect to the product measure $\mu^{\otimes N}$ of $X_1, \ldots, X_N$. Note that this setup easily extends to the case $\mathtt{E}\{\epsilon_k^2\} = \sigma^2(X_k)$ with $\sigma^2(x)$ a known function such that $f(x)f^\top(x)/\sigma^2(x)$ is μ-integrable.

We consider generalized designs that are probability distributions on the set $\mathcal{X}$, and denote Ξ the set of such designs. The subset $\{\xi \;/\; \Phi[\mathbf{M}(\xi)] \geq A > -\infty\}$ will be denoted $\Xi(A)$. We assume that Φ is a concave linearly differentiable function on $\Xi(A)$ for any A, that is, the directional derivative $\mathcal{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\epsilon \to 0^+}\{\Phi[(1-\epsilon)\mathbf{M}_1 + \epsilon\mathbf{M}_2] - \Phi(\mathbf{M}_1)\}/\epsilon$ 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 in Ξ, with $F_\Phi(\xi, x) = F_\Phi(\xi; \delta_x)$ and δ_x the Dirac measure supported at x. For instance, in the case of D-optimality where $\Phi(\cdot) = \log\det(\cdot)$, one has $F_\Phi(\xi, x) = f^\top(x)\mathbf{M}^{-1}(\xi)f(x) - p$, with $p = \dim(\theta)$.

For N finite, the problem is

$$\text{maximise} \;\; \mathtt{E}\{\Phi(\mathbf{M}_N/n)\} \tag{3}$$

with respect to (u_j) satisfying (2). 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,

$$a_k = \sum_{j=1}^{k-1} u_j \,, \tag{4}$$

with $a_1 = 0$. This problem corresponds to a discrete-time stochastic control problem, where k represents time, $(a_k, \mathbf{M}_{k-1}, X_k)$ and $u_k \in \mathcal{U}_k \subseteq \{0,1\}$ respectively represent the state and control at time k. A strategy $S_{N,n}$ is defined by a mapping $(k, a, \mathbf{M}, X) \mapsto u \in \{0,1\}$. For each $k \in \{1, \ldots, N\}$, the optimal decision at step k is obtained by solving:

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

where $\mathtt{E}_{X_j}\{.\}$ denotes the expectation with respect to X_j, distributed with the measure μ, and,

$$\mathcal{U}_j = \mathcal{U}_j(a_j) = \begin{cases} \{0\} & \text{if } a_j = n \,, \\ \{1\} & \text{if } a_j + N - j + 1 \leq n \,, \\ \{0,1\} & \text{otherwise.} \end{cases}$$

The case $p = \dim(\theta) = 1$ is considered in [1]: the optimal (closed-loop) solution is given by a backward recurrence equation, and a simple open-loop solution is constructed, that is asymptotically optimal for $N \to \infty$ with n fixed (for measures μ absolutely continuous with respect to the Lebesgue measure and such that the associated distribution function is a von Mises function). The multidimensional case $p > 1$, for which the optimal solution cannot in general be obtained in close form, is considered in [2] and [3], where suboptimal solutions are proposed: open-loop feedback-optimal control is used in [2] and a heuristic one-step ahead decision rule in [3]. The strategies presented in this paper rely on the construction of an optimal constrained design measure $\xi_\alpha^* \leq \mu/\alpha$, a problem which is briefly described in Section 2. We show in Section 3 that from ξ_α^* one can easily define a strategy asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $\alpha \in (0,1)$, $N \to \infty$ (we use an approach similar to [4] which corresponds to the case $p = 1$). This in turn suggests a method for sampling asymptotically from an optimal constrained measure $\xi_\alpha^* \leq \mu/\alpha$, without having to determine ξ_α^*, or even without knowing μ in advance. This sampling strategy $S_\alpha(\mu)$ is presented in Section 4 where we prove that $\Phi[\mathbf{M}(\xi_k)] \to \Phi[\mathbf{M}(\xi_\alpha^*)]$, μ-almost surely, with ξ_k the empirical measure of sampled points. A simple adaptation of α as a function of k makes $S_\alpha(\mu)$ coincide with the heuristic rule in [3], which is thus shown to be asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $\alpha \in (0,1)$, $N \to \infty$. Illustrative examples are given in Section 5. Section 6 finally concludes and draws some perspectives.

2. Optimal constrained design measures

We forget for the moment the sequential character of the decisions, and assume that we simply have to select n design points X_i among N, distributed with the measure μ satisfying $\int_{\mathcal{X}} \mu(dx) = 1$. For $n = \lfloor \alpha N \rfloor$, $N \to \infty$, $0 < \alpha < 1$, this means that only a proportion α of the design points can be accepted, which puts a constraint of the form

$$\xi(dx) \leq \mu(dx)/\alpha \tag{5}$$

on the design measure ξ. We denote by $\mathcal{D}(\mu,\alpha)$ the set of admissible measures satisfying (5), with $\int_{\mathcal{X}} \xi(dx) = 1$. The issue is to determine a measure ξ^*_α in $\mathcal{D}(\mu,\alpha)$ that maximises $\Phi[\mathbf{M}(\xi)]$, where $\mathbf{M}(\xi) = \int_{\mathcal{X}} f(x)f^\top(x)\xi(dx)$.

Define the following subclass of $\mathcal{D}(\mu,\alpha)$:

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

The following theorem is proved in [5].

Theorem 1. *For any ξ in $\mathcal{D}(\mu,\alpha)$ there exist a ξ' in $\mathcal{D}^*(\mu,\alpha)$ such that $\mathbf{M}(\xi) = \mathbf{M}(\xi')$.*

Concavity of Φ and convexity of $\mathcal{D}(\mu,\alpha)$ imply that an optimal design measure ξ^*_α exists in $\mathcal{D}(\mu,\alpha)$. From Theorem 1, $\xi^*_\alpha \in \mathcal{D}^*(\mu,\alpha)$. Differentiability of Φ gives a characterization of this optimal measure in terms of a necessary and sufficient condition, see [5, 6, 7, 8].

Theorem 2. *A necessary and sufficient condition that ξ^*_α maximises $\Phi[\mathbf{M}(\xi)]$ over $\mathcal{D}(\mu,\alpha)$ is that exist a constant c such that $F_\Phi(\xi^*_\alpha, x) \geq c$ for ξ^*_α-almost all x, and $F_\Phi(\xi^*_\alpha, x) \leq c$ for $(\mu - \xi^*_\alpha)$-almost all x.*

In [5, 6], the condition is formulated as: $F_\Phi(\xi^*_\alpha, x)$ separates the two sets

$$\mathcal{X}^*_\alpha = \text{supp}\ \xi^*_\alpha = \{x \in \mathcal{X} \ / \ \xi^*_\alpha(x) > 0\} \tag{6}$$

and $\mathcal{X}\backslash\mathcal{X}^*_\alpha$. It is shown in [6, 7] that $\int_{\mathcal{X}^*_\alpha} F_\Phi(\xi^*_\alpha, x)\mu(dx) = \int_{\mathcal{X}} F_\Phi(\xi^*_\alpha, x)\xi^*_\alpha(dx) = 0$. An extension to the case where μ is not necessarily atomless is given in [8] (which also considers design measures bounded from below). Iterative algorithms of the exchange type for the construction of an optimal constrained measure ξ^*_α are presented in [6, 7]. See also [9] for other steepest ascent algorithms.

3. Asymptotically optimal decisions

For any strategy $S_{N,n}$ used for the maximisation of (3), we denote $\Psi(S_{N,n}) = \Phi(\mathbf{M}_N/n)$. We follow the same line as in [4], which concerns the case $p = 1$.

First, we compare the optimal strategy with an 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 $\Phi(\mathbf{M}_N/n)$ *after* the N points $X_1, \ldots, X_N$ have been observed. Obviously, for any N, n and any sequential strategy $S_{N,n}$

$$\Psi(S_{N,n}) \leq \Psi(S^*_{N,n})\,. \tag{7}$$

The strategy $S^*_{N,n}$ also satisfies the following.

Lemma 1. *For any $\alpha \in (0,1)$,*

$$\lim_{N\to\infty} \Psi(S^*_{N,\lfloor\alpha N\rfloor}) = \Phi[\mathbf{M}(\xi^*_\alpha)]\,,\ \mu\text{-a.s.} \tag{8}$$

and

$$\mathtt{E}\{\Psi(S^*_{N,n})\} \leq \Phi[\mathbf{M}(\xi^*_{n/N})]\,,\ \forall(N,n)\,, \tag{9}$$

with $\xi^*_\alpha \leq \mu/\alpha$ an optimal constrained measure.

Proof. Let $\hat{S}_{N,\lfloor\alpha N\rfloor}$ be the strategy defined as follows. Sample N times from μ, let N_α be the number of X_i's that fall in $\mathcal{X}^*_\alpha$. Accept $\min\{\lfloor\alpha N\rfloor, N_\alpha\}$ such points, and, if $\lfloor\alpha N\rfloor > N_\alpha$ complete this set by $\lfloor\alpha N\rfloor - N_\alpha$ other points from the sample. One has $\lim_{N\to\infty} \Psi(\hat{S}_{N,\lfloor\alpha N\rfloor}) = \Phi[\mathbf{M}(\xi^*_\alpha)]$, μ-a.s. At the same time, $\Psi(\hat{S}_{N,\lfloor\alpha N\rfloor}) \leq \Psi(S^*_{N,\lfloor\alpha N\rfloor})$ for any N, which gives

$$\liminf_{N\to\infty} \Psi(S^*_{N,\lfloor\alpha N\rfloor}) \geq \Phi[\mathbf{M}(\xi^*_\alpha)]\,,\ \mu\text{-a.s.}$$

On the other hand, for any N the strategy $S^*_{N,\lfloor \alpha N \rfloor}$ samples from a measure that belongs to $\mathcal{D}(\mu,\alpha)$, and thus

$$\limsup_{N\to\infty} \Psi(S^*_{N,\lfloor \alpha N \rfloor}) \leq \Phi[\mathbf{M}(\xi^*_\alpha)]\,,\ \mu\text{-a.s.}$$

which gives (8).

Let $\mathbf{M}^*_{N,n}$ denote the information matrix generated by $S^*_{N,n}$,

$$\mathbf{M}^*_{N,n} = \frac{1}{n}\sum_{i=1}^{N} f(X_i) f^\top(X_i) I_N(X_i)$$

where $I_N(X_i) = I(X_1,\ldots,X_N;X_i)$ equals 1 if X_i is accepted by $S^*_{N,n}$ and equals 0 otherwise. Repeat m times this strategy, with $\mathbf{M}^*_{N,n}(j)$ the matrix generated at the jth experiment, $j=1,\ldots,m$. One has for any N,n,m

$$\Phi\left[\frac{1}{m}\sum_{j=1}^{m}\mathbf{M}^*_{N,n}(j)\right] \leq \Phi(\mathbf{M}^*_{mN,mn})\,,$$

where $\mathbf{M}^*_{mN,mn}$ is obtained from the same non sequential strategy $S^*_{mN,mn}$ applied to the full sample of mN points. Now, let m tend to infinity: $\Phi(\mathbf{M}^*_{mN,mn})$ tends to $\Phi[\mathbf{M}(\xi^*_{n/N})]$ μ-a.s., see (8), and $(1/m)\sum_{j=1}^{m}\mathbf{M}^*_{N,n}(j)$ to $\mathtt{E}\{\mathbf{M}^*_{N,n}\}$, μ-a.s. Concavity of $\Phi(\cdot)$ gives (9). □

Consider now the strategy defined by

$$S^\epsilon_{N,n} : \begin{cases} \text{accept } X_k \text{ if } X_k \in \mathcal{X}^*_{\alpha+\epsilon}\,,\ \text{or } N-k+1 \leq n-a_k \\ \text{reject } X_k \text{ otherwise}\,, \end{cases} \tag{10}$$

with $0 \leq \epsilon < 1-\alpha$, a_k given by (4) and $\mathcal{X}^*_{\alpha+\epsilon}$ by (6). We simplify this strategy, and consider instead $\bar{S}^\epsilon_{N,n}$ defined by

$$\bar{S}^\epsilon_{N,n} : \begin{cases} \text{accept } X_k \text{ if } X_k \in \mathcal{X}^*_{\alpha+\epsilon}\,, \\ \text{reject } X_k \text{ otherwise.} \end{cases}$$

It is non admissible but satisfies $\Psi(S^\epsilon_{N,n}) \geq \Psi(\bar{S}^\epsilon_{N,n})$ since $\bar{S}^\epsilon_{N,n}$ may accept less than n design points. One has

$$\Psi(\bar{S}^\epsilon_{N,n}) = \Phi\left[\frac{1}{n}\sum_{i=1}^{M_n} f(X_i) f^\top(X_i) I_{\mathcal{X}^*_{\alpha+\epsilon}}(X_i)\right]$$

where $I_{\mathcal{A}}(\cdot)$ is the indicator function of the set $\mathcal{A}$ and $M_n = \min(U_n, N)$, with

$$U_n = \min\{K\ /\ \sum_{i=1}^{K} I_{\mathcal{X}^*_{\alpha+\epsilon}}(X_i) = n\}\,.$$

Now, $\text{Prob}(X_1 \in \mathcal{X}^*_{\alpha+\epsilon}) = \mu(\mathcal{X}^*_{\alpha+\epsilon}) = \alpha+\epsilon$, so that $(1/N)\sum_{i=1}^{N} I_{\mathcal{X}^*_{\alpha+\epsilon}}(X_i) \to \alpha+\epsilon$, μ-a.s. This implies for $\epsilon > 0$:

$$\forall \delta > 0\,,\ \exists N_0\ /\ \text{Prob}\left[\forall N > N_0\,,\ \sum_{i=1}^{N} I_{\mathcal{X}^*_{\alpha+\epsilon}}(X_i) > \lfloor \alpha N \rfloor\right] > 1-\delta\,,$$

and thus for $n = \lfloor \alpha N \rfloor$,

$$\text{Prob}[\forall N > N_0\,,\ M_n = U_n \leq N] > 1-\delta\,.$$

Therefore,

$$\text{Prob}\left\{\forall N > N_0\,,\ \Psi(\bar{S}^\epsilon_{N,n}) = \Phi\left[\frac{1}{n}\sum_{i=1}^{n} f(X_{k_i}) f^\top(X_{k_i})\right]\,,\ X_{k_i} \in \mathcal{X}^*_{\alpha+\epsilon}\,,\ i=1,\ldots,n\right\} > 1-\delta\,,$$

$$\Psi(\bar{S}^\epsilon_{N,\lfloor \alpha N \rfloor}) \to \Phi\left[\frac{1}{\alpha+\epsilon}\mathtt{E}\{f(X_1) f^\top(X_1) I_{\mathcal{X}^*_{\alpha+\epsilon}}(X_1)\}\right] = \Phi[\mathbf{M}(\xi^*_{\alpha+\epsilon})]\,,\ \mu\text{-a.s.}$$

and thus

$$\liminf_{N\to\infty} \Psi(S^\epsilon_{N,\lfloor \alpha N \rfloor}) \geq \Phi[\mathbf{M}(\xi^*_{\alpha+\epsilon})]\ \mu\text{-a.s.} \tag{11}$$

We may then let ϵ tend to zero and use the continuity of $\Phi[\mathbf{M}(\xi_\alpha^*)]$ with respect to α to obtain the following property directly from (7,8) and (11).

Theorem 3. *The strategy $S_{N,n}^0$ defined by (10) is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $N \to \infty$, $0 < \alpha < 1$:*

$$\lim_{N\to\infty} \Psi(S_{N,\lfloor \alpha N \rfloor}^0) = \lim_{N\to\infty} \Psi(S_{N,\lfloor \alpha N \rfloor}^*) = \Phi[\mathbf{M}(\xi_\alpha^*)]\,,\ \mu\text{-a.s.}$$

where $\xi_\alpha^* \leq \mu/\alpha$ is an optimal constrained measure and $S_{N,n}^*$ is an optimal non sequential strategy that selects n points among N after these N points have been observed.

4. Sampling asymptotically from a constrained measure

Theorem 2 and the fact that $\text{Prob}(X_1 \in \mathcal{X}_\alpha^*) = \mu(\mathcal{X}_\alpha^*) = \alpha$ imply that the asymptotically optimal strategy $S_{N,n}^0$ is equivalently defined by

$$S_{N,n}^0 : \begin{cases} \text{accept } X_k \text{ if } \mu\{x \ / \ F_\Phi(\xi_\alpha^*, x) > F_\Phi(\xi_\alpha^*, X_k)\} < \alpha\,, \text{ or } N-k+1 \leq n - a_k \\ \text{reject } X_k \text{ otherwise}\,. \end{cases}$$

This suggests the following procedure for sampling asymptotically from ξ_α^* without having to determine ξ_α^* beforehand: we simply substitute ξ_k for ξ_α^*, with ξ_k the empirical design measure at the current stage.

Strategy $S_\alpha(\mu)$**:**

- (0) Sample $X_1, X_2, \ldots, X_{n_0}$ from μ, with n_0 the first integer such that $\Phi[\sum_{i=1}^{n_0} f(X_i) f^\top(X_i)] > -\infty$. Set $k = 1$, $a_k = n_0$, $\xi_k = (1/n_0) \sum_{i=1}^{n_0} \delta_{X_i}$.
- (i) Sample X_{k1} from μ.
- (ii) Compute $P_k = P(X_k) = \mu\{x \ / \ F_\Phi(\xi_k, x) > F_\Phi(\xi_k, X_k)\}$,

$$\begin{cases} \text{if } P_k < \alpha\,,\ \text{accept } X_k\,,\ \text{set } a_{k+1} = a_k + 1\,,\ \xi_{k+1} = [1 - 1/(1+a_k)]\xi_k + 1/(1+a_k)\,\delta_{X_k}\,, \\ \text{otherwise, reject } X_k\,,\ \text{set } a_{k+1} = a_k\,,\ \xi_{k+1} = \xi_k\,. \end{cases}$$

$k \leftarrow k+1$, return to step (i).

Notice that this procedure does not guarantee that $a_k/k \leq \alpha$. However, this is easily obtained by adding the condition $a_k/k < \alpha$ at step (ii) for accepting X_k. Other initialisations for ξ, e.g. $\xi_1 = \mu$, could be used too.

Define the second order directional derivative

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

Besides the assumptions made in the introduction, we shall need the following.

H1: $\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2)$ is continuous in $\mathbf{M}_1$ and $\mathbf{M}_2$ for any finite $\mathbf{M}_1$ and

$$\int_{\mathcal{A}} \nabla_\Phi^2[\mathbf{M}_1, f(x) f^\top(x)] \mu(dx) > \Delta_\alpha(A) > -\infty$$

for any set $\mathcal{A}$ such that $\mu(\mathcal{A}) \geq \alpha$ and any $\mathbf{M}_1$ such that $\Phi(\mathbf{M}_1/2) > A$.

H2: $\exists A > -\infty$ and $\epsilon > 0$ such that $D^*(\mu, \alpha - \epsilon) \subset \Xi(A)$.

H1 holds for usual design criteria, for instance when $\mathcal{X}$ in bounded, or when $\eta(\theta, x)$ is polynomial in x and μ has a density $\phi(\cdot)$ with respect to the Lebesgue measure, with $\phi(x)$ exponentially decreasing when $x \to \infty$. H2 holds for instance if the functions $f(x)$ are independent on any subset $\mathcal{A}$ of $\mathcal{X}$ with $\mu(\mathcal{A}) > \alpha - \epsilon$ and μ has a density $\phi(\cdot)$ such that $\phi(x) \geq q > 0$ on $\mathcal{X}$.

The difficulty for studying the convergence of the strategy $S_\alpha(\mu)$ is that $\Phi[\mathbf{M}(\xi_k)]$ is not monotonically increasing, due to (i) the predetermined step length $1/(1+a_k)$ and (ii) the stochastic character of X_k. While (i) is standard in the construction of optimum designs, see [10], (ii) is less usual and forms a specific feature of the context considered here. Next theorem shows that, under H1 and H2, $\Phi[\mathbf{M}(\xi_k)]$ converges μ-a.s. to $\Phi[\mathbf{M}(\xi_\alpha^*)]$ when k tends to infinity. Notice that H2 eliminates the unboundedness case encountered in the dichotomous theorem of [10].

Theorem 4. *Consider the strategy $S_\alpha(\mu)$ defined by (0)-(ii) above, with a_k the number of points selected from k points sampled, see (4), and ξ_k the empirical measure of the selected points. Under H1 and H2, $S_\alpha(\mu)$ satisfies*

$$\lim_{k\to\infty} a_k/k = \alpha\,,\ \mu\text{-a.s.}, \tag{12}$$

and

$$\lim_{k\to\infty} \Phi[\mathbf{M}(\xi_k)] = \Phi[\mathbf{M}(\xi^*_\alpha)]\,,\ \mu\text{-a.s.}, \tag{13}$$

with $\xi^*_\alpha \leq \mu/\alpha$ an optimal constrained design measure.

Proof. Define $\mathcal{X}^*_{\alpha,k} = \{x \ / \ P_k(x) < \alpha\}$. One has $\mu(\mathcal{X}^*_{\alpha,k}) = \alpha$ which proves (12).

Define $\Phi^* = \Phi(\xi^*_\alpha)$, $\Phi_k = \Phi[\mathbf{M}(\xi_k)]$ and $T_A = \inf\{k \ / \ \Phi_k \leq A\}$.

Consider first an iteration when ξ_k is updated, which occurs when $X_k \in \mathcal{X}^*_{\alpha,k}$. Denote A_k this event, which has probability α. We have

$$\Phi_{k+1} = \Phi_k + \frac{1}{1+a_k}\, F_\Phi(\xi_k, X_k) + \frac{1}{2(1+a_k)^2}\, \nabla^2_\Phi(\xi_k, X_k, \gamma)$$

with

$$\nabla^2_\Phi(\xi_k, X_k, \gamma) = \nabla^2_\Phi[(1-\gamma)\mathbf{M}(\xi_k) + \gamma f(X_k)f^\top(X_k), f(X_k)f^\top(X_k)]\,,$$

for some $\gamma \in [0, 1/(1+a_k)]$. Concavity of $\Phi(\cdot)$ implies

$$F_\Phi(\xi_k; \xi^*_\alpha) \geq \Phi^* - \Phi_k$$

that is,

$$\int_{\mathcal{X}} F_\Phi(\xi_k, x)\xi^*_\alpha(dx) \geq \Phi^* - \Phi_k\,,$$

or equivalently, since $\xi^*_\alpha = \mu/\alpha$ on $\mathcal{X}^*_\alpha$ and 0 on $\mathcal{X}\backslash\mathcal{X}^*_\alpha$,

$$\int_{\mathcal{X}^*_\alpha} F_\Phi(\xi_k, x)\mu(dx)/\alpha \geq \Phi^* - \Phi_k\,.$$

Therefore, from the definition of $\mathcal{X}^*_{\alpha,k}$,

$$\mathtt{E}_{X_k}\{F_\Phi(\xi_k, X_k)|A_k\} = \int_{\mathcal{X}^*_{\alpha,k}} F_\Phi(\xi_k, x)\mu(dx)/\alpha \geq \Phi^* - \Phi_k\,.$$

Conditionally on $(T_A = \infty)$, H1 implies

$$\mathtt{E}\{\Phi_{k+1}|\mathcal{F}_{k-1}, A_k\} = \mathtt{E}_{X_k}\{\Phi_{k+1}|\xi_k, a_k, A_k\} > \Phi_k + \frac{1}{1+a_k}\,(\Phi^* - \Phi_k) + \frac{\Delta_\alpha(A)}{2\alpha(1+a_k)^2}\,. \tag{14}$$

On $\bar{A}_k$, that is, when ξ_k is not updated, $\Phi_{k+1} = \Phi_k$. Together with (14), this gives on $(T_A = \infty)$ for any iteration k

$$\mathtt{E}\{\Phi_{k+1} - \Phi_k|\mathcal{F}_{k-1}\} > \frac{\alpha}{1+a_k}\,(\Phi^* - \Phi_k) + \frac{\Delta_\alpha(A)}{2(1+a_k)^2}\,, \tag{15}$$

which implies $\limsup_{k\to\infty} \Phi_k \geq \Phi^*$ μ-a.s., and thus from Lemma 1 $\limsup_{k\to\infty} \Phi_k = \Phi^*$ μ-a.s., on $(T_A = \infty)$.

Assume now that $\limsup_{k\to\infty} \Phi_k - \liminf_{k\to\infty} \Phi_k > \epsilon > 0$. This would imply that (Φ_k) crosses the interval $[\Phi^* - \epsilon, \Phi^* - \epsilon/6]$ infinitely often. We prove that it is impossible, using an approach similar to the proof of Doob's upcrossing Lemma, see [11], p. 106.

First, we construct a previsible process (C_k) from (Φ_k) as follows. We wait until Φ_k gets above $\Phi^* - \epsilon/6$ and set C to one until Φ_k gets below $\Phi^* - \epsilon$. Formally, this is written as follows:

$$C_2 = \mathcal{I}_{\{\Phi_1 \geq \Phi^* - \epsilon/6\}}\,,\ C_{k+1} = C_k\mathcal{I}_{\{\Phi_k \geq \Phi^* - \epsilon\}} + (1 - C_k)\mathcal{I}_{\{\Phi_k \geq \Phi^* - \epsilon/6\}}\,,\ k \geq 2\,.$$

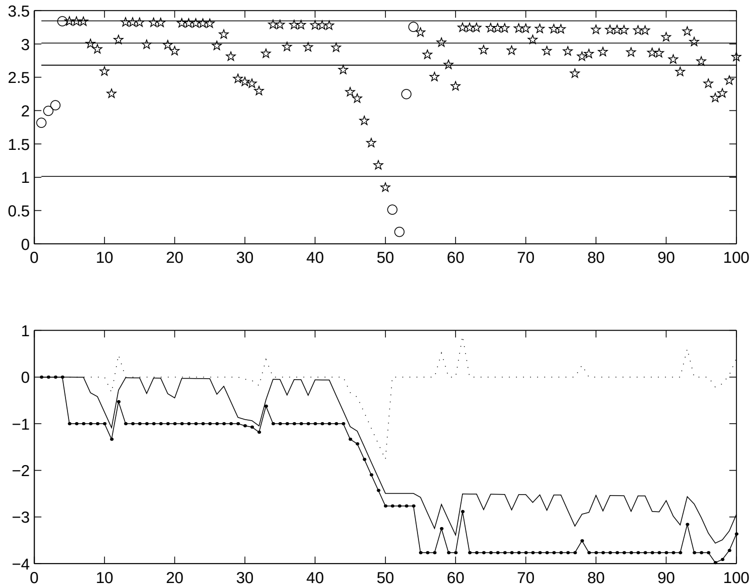

FIGURE 1. Top: sequence Φ_k (stars when $C_k = 1$, circles when $C_k = 0$); bottom : V_k (full line), Z_k (dotted line) and W_k (dotted full line).

We then define the martingale transform (V_k) of (Φ_k) by

$$V_k = \sum_{i=2}^{k} C_i(\Phi_i - \Phi_{i-1})\,.$$

One has $V_k < -(5\epsilon/6)U_k + [\Phi_k - (\Phi^* - \epsilon/6)]^+$, with U_k the number of downcrossings of $[\Phi^* - \epsilon, \Phi^* - \epsilon/6]$ made by the process $i \mapsto \Phi_i$ by time k. Therefore,

$$\mathtt{E}\{V_k\} < -(5\epsilon/6)\mathtt{E}\{U_k\} + \mathtt{E}\{[\Phi_k - (\Phi^* - \epsilon/6)]^+\}\,. \tag{16}$$

Next, we construct another martingale transform (W_k) of (Φ_k), that we shall use as a lower bound on V_k for k large enough, so that $\mathtt{E}\{W_k\}$ will give a lower bound on $\mathtt{E}\{V_k\}$. The difficulty is that the bound given by (15) can be used only when $\Phi_k < \Phi^*$. For that reason we shall use again the threshold $\Phi^* - \epsilon/6$.

We are only interested in time intervals where $C_k = 1$. First we construct a process Z_k that will mimic V_k when Φ_k leaves the region above $\Phi^* - \epsilon/6$:

$$Z_1 = 0\,,\ Z_{k+1} = C_{k+1}\mathcal{I}_{\{\Phi_k < \Phi^* - \epsilon/6\}}(Z_k + \Phi_{k+1} - \Phi_k)\,,\ k \geq 2\,.$$

Then we consider the initial part of the time interval when $C_k = 1$ and Φ_k fluctuates around Φ^*. We construct

$$Q_1 = 0\,,\ Q_{k+1} = Q_k - (\epsilon/2)\max(C_{k+1} - C_k, 0) + Z_k \max(C_k - C_{k+1}, 0)\,,\ k \geq 2$$

and define W_k as $W_k = Q_k + Z_k$. This construction guarantees $V_k - V_{k_0} \geq W_k - W_{k_0}$ for all $k > k_0$ provided that $\Phi_{k+1} - \Phi_k > -\epsilon/6$ and $\Phi_k < \Phi^* + \epsilon/6$ when $k \geq k_0$.

Figure 1 illustrates the construction. The upper part of the figure gives Φ_k as a function of k. Stars are for time instants when $C_k = 1$, circles for $C_k = 0$. The horizontal lines, from top to bottom, correspond to the values $\Phi^* + \epsilon/6$, Φ^*, $\Phi^* - \epsilon/6$, $\Phi^* - \epsilon$. The lower part of the figure presents the sequences (V_k) (full line), (Z_k) (dotted line) and (W_k) (dotted full line).

Define the following quantities

$$\begin{aligned} K_1 &= \sup\{k \;/\; \Phi_k > \Phi^* + \epsilon/6\}\,, \\ K_2 &= \sup\left\{k \;/\; \frac{\epsilon}{6}\,\frac{\alpha}{1+a_k} + \frac{\Delta_\alpha(A)}{2(1+a_k)^2} < 0\right\}\,, \\ K_3 &= \sup\left\{k \;/\; \Phi_{k+1} - \Phi_k < -\epsilon/6\right\}\,, \end{aligned}$$

and $K^* = \max\{K_1, K_2, K_3\} \leq \infty$. On $(T_A = \infty) \cap (K^* < \infty)$ we have, for any $k > K^*$,

$$V_k \geq V_{K^*} + W_k - W_{K^*}\,,$$

and, by construction, using (15), $\mathtt{E}\{W_k - W_{K^*}\} \geq -(\epsilon/2)\mathtt{E}\{U_k - U_{K^*}\}$. Together with (16), this gives, for any $k > K^*$,

$$\mathtt{E}\{U_k\} < \frac{3}{\epsilon}\left(\mathtt{E}\{[\Phi_k - (\Phi^* - \epsilon/6)]^+\} - \mathtt{E}\{V_{K^*}\}\right) < \infty\,,$$

and thus $\mathrm{Prob}(U_\infty = \infty) = 0$. Now, $\limsup_{k\to\infty} \Phi_k = \Phi^*$, μ-a.s., and (12) imply $\mathrm{Prob}(K_1 < \infty) = \mathrm{Prob}(K_2 < \infty) = 1$. Also, on $(T_A = \infty)$ H1 implies

$$\Phi_{k+1} > \Phi_k - \frac{h_\alpha(A)}{1+a_k}$$

for some $h_\alpha(A) > 0$, so that $\mathrm{Prob}(K_3 < \infty) = 1$. Therefore, on $(T_A = \infty)$, $\mathrm{Prob}(U_\infty < \infty) = 1$ and $\liminf_{k\to\infty} \Phi_k = \limsup_{k\to\infty} \Phi_k = \Phi^*$, μ-a.s.

Finally, (12) and H2 imply $\mathrm{Prob}\{\cup_{A=-1}^{-\infty}(T_A = \infty)\} = 1$, which completes the proof. □

When only n points can be accepted among N, which corresponds to the original problem of Section 1, we can adapt α and take $\alpha = (n - a_k)/(N - k)$ at step k. This corresponds to the strategy

$$S_{N,n} : \begin{cases} \text{accept } X_k \text{ if } P_k < \frac{n-a_k}{N-k} \\ \text{reject } X_k \text{ otherwise} \end{cases} \tag{17}$$

which coincides with the one-step-ahead rule suggested in [3]. Theorem 4 shows that it is asymptotically optimal when $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \to \infty$.

When μ is unknown, the probability P_k in (17) or step (ii) of $S_\alpha(\mu)$ can be replaced by $\hat{P}_k$ evaluated from the empirical measure of the X_i's, say $\hat{\mu}_k$ at step k, with the initial measure $\hat{\mu}_1$ obtained from n_0 initial samples. Since the random variables X_k are observed whatever the strategy is, $\hat{P}_k$ converges a.s. to P_k, which permits to asymptotically sample from ξ^*_α, without knowing μ. Illustrative examples are presented in the next section. Kernel estimation of μ, or a parametric representation μ_β, with β to be estimated from the sequence (X_k), could also be considered.

Finally note that if θ can be estimated on line, then $\hat{\theta}^0$ can be replaced at step k by $\hat{\theta}^k$ estimated by least squares from the observations collected so far. If the criterion $\Phi(\cdot)$ is such that $\Phi(\mathbf{M}) > A$ implies $\lambda_{\min}(\mathbf{M}) > \gamma > 0$, consistency and asymptotic normality of $\hat{\theta}^k$ will hold under H2 (and additional conditions on higher order derivatives of $\eta(\theta, x)$ with respect to θ and their tail cross product, see [12]) and the strategy $S_\alpha(\mu)$ is such that $\Phi[\mathbf{M}(\xi_k)]$ converges μ-a.s. to $\Phi\{\mathbf{M}[\xi^*_\alpha(\bar{\theta})]\}$, with $\xi^*_\alpha(\bar{\theta}) \leq \mu/\alpha$ the optimal constrained design measure for the true value $\bar{\theta}$ of the model parameters in (1).

5. Examples

We consider the quadratic regression model $\eta(\theta, x) = \theta_0 + \theta_1 x + \theta_2 x^2$, with the design criterion $\Phi(\cdot) = \log\det(\cdot)$.

5.1. **Example 1.** The experimental variables X_k are normally distributed $\mathcal{N}(0, 1)$ and $\alpha = 0.5$. Easy calculations show that the optimal constrained measure $\xi^*_\alpha \leq \mu/\alpha$ is equal to μ/α on $\mathcal{X}^*_\alpha = (-\infty, -a] \cup [-b, b] \cup [a, \infty)$, with $a \simeq 1.028$, $b \simeq 0.2482$. Figure 2 presents a plot of the sensitivity function $d(\xi^*_\alpha, x) = f^\top(x)\mathbf{M}^{-1}(\xi^*_\alpha)f(x)$ and illustrates Theorem 2. An histogram of the 10,000 first samples X_k accepted by the strategy $S_\alpha(\mu)$ of Section 3, with μ replaced by the empirical measure $\hat{\mu}_k$ ($\hat{\mu}_1$ is constructed from $n_0 = 3$ initial samples), is presented in Figure 3. Figure 4 gives Φ_k as a function of k, the value $\Phi[\mathbf{M}(\xi^*_\alpha)]$ is indicated by a dashed line.

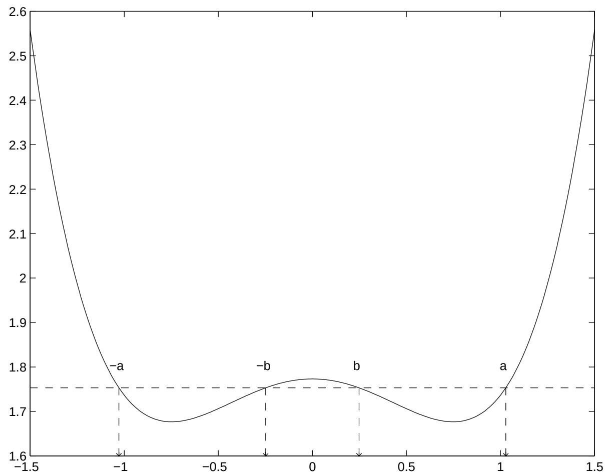

FIGURE 2. Sensitivity function $d(\xi_\alpha^*, x)$ for ξ_α^* in Example 1.

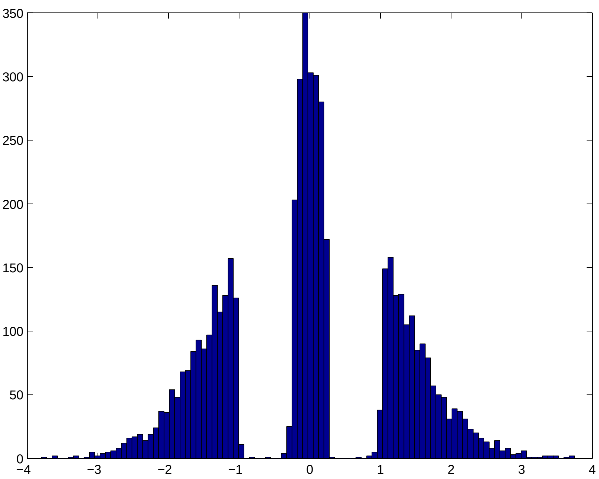

FIGURE 3. Histogram of the $10,000$ first samples X_k accepted by $S_\alpha(\mu)$ in Example 1.

5.2. **Example 2.** Assume now that the experimental variables X_k are uniformly distributed in $[-1, 1]$, again with $\alpha = 0.5$. It is shown in [7] that the optimal constrained measure ξ_α^* is supported on $\mathcal{X}_\alpha^* = [-1, -a] \cup [-b, b] \cup [a, 1]$, with $b = a - 1/2$, $a \simeq 0.7098$. Figure 5 presents an histogram of the $10,000$ first samples X_k accepted by the strategy $S_\alpha(\mu)$ of Section 3, with μ replaced by the empirical measure $\hat{\mu}_k$ ($\hat{\mu}_1$ is constructed from $n_0 = 3$ initial samples).

5.3. **Example 3.** This example corresponds to a situation where the assumptions used in Theorem 4 do not hold: μ is a mixture of the normal measure μ_n for $\mathcal{N}(0, 1)$ and the discrete measure μ_d supported at $\{-1, -1/2, 0, 1/2, 1\}$ with respective weights $(1/8, 1/4, 1/4, 1/4, 1/8)$, $\mu = 0.5\mu_n + 0.5\mu_d$. We take $\alpha = 0.1$. Easy calculations then show that the optimal 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 at zero. Figure 6 presents a plot of the sensitivity function $d(\xi_\alpha^*, x) = f^\top(x)\mathbf{M}^{-1}(\xi_\alpha^*)f(x)$ and illustrates the optimality of ξ_α^*, see Corollary 1 in [8]. We modify step (ii) of the strategy $S_\alpha(\mu)$ and accept X_k only when $P_k < \alpha$ *and* $a_k/k < \alpha$. Figure 7, left, gives an histogram of the first 10,000 samples accepted (μ is replaced by the empirical measure $\hat{\mu}_k$ and $\hat{\mu}_1$ is constructed from $n_0 = 3$ initial samples). The right part of the figure

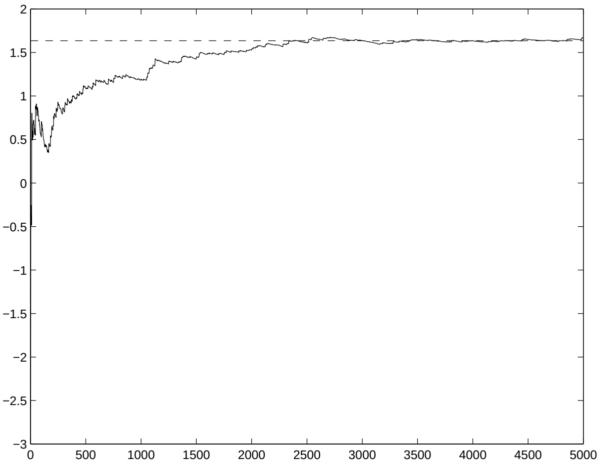

FIGURE 4. Φ_k generated by $S_\alpha(\mu)$ as a function of k in Example 1; $\Phi[\mathbf{M}(\xi_\alpha^*)]$ corresponds to the dashed line.

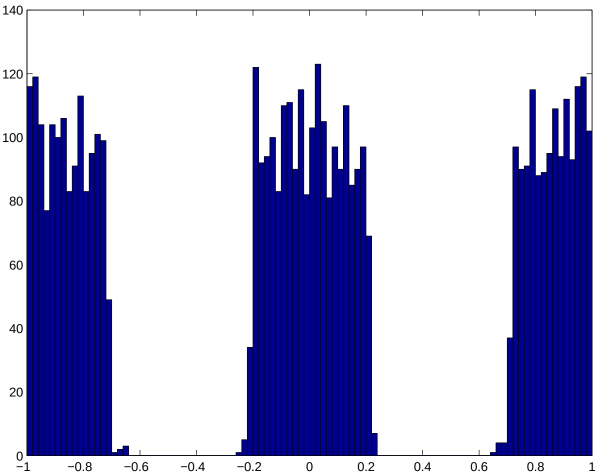

FIGURE 5. Histogram of the 10,000 first samples X_k accepted by $S_\alpha(\mu)$ in Example 2.

presents Φ_k as a function of k, with the optimal value $\Phi[\mathbf{M}(\xi_\alpha^*)]$ indicated by the dashed line. Although we do not have a proof for convergence, simulations indicate that $\Phi_k \to \Phi^*$ as $k \to \infty$.

6. Conclusions and further developments

Two simple open-loop strategies for the sequential selection of n experiments $X_{k_1}, \ldots, X_{k_n}$ among N i.i.d. X_i's have been shown to be asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $\alpha \in (0,1)$, $N \to \infty$, and rather general design criterion Φ. The first one is based on the construction of an optimum constrained design measure $\xi_\alpha^* \leq \mu/\alpha$, with μ the probability measure for X_1. The second does not rely on the prior construction of ξ_α^*, but asymptotically samples from it, that is, $\Phi[\mathbf{M}(\xi_k)] \to \Phi[\mathbf{M}(\xi_\alpha^*)]$, μ-a.s., $k \to \infty$, where ξ_k denotes the empirical measure of the points accepted by the strategy. In the one dimensional case ($\dim(\theta) = 1$), asymptotic optimality of a similar strategy is proved in [1] for $N \to \infty$ with n fixed, provided the tail of μ is thin enough (von Mises distribution). In the case $\dim(\theta) > 1$, extending Theorem 3 to the situation where n is fixed and $N \to \infty$ remains an open issue.

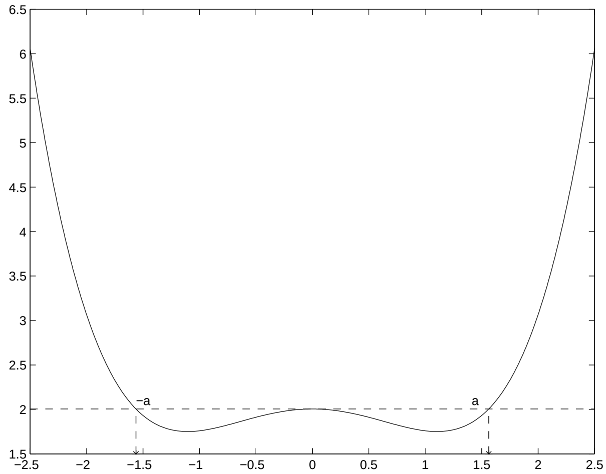

FIGURE 6. Sensitivity function $d(\xi_\alpha^*, x)$ for the optimal constrained measure ξ_α^* in Example 3.

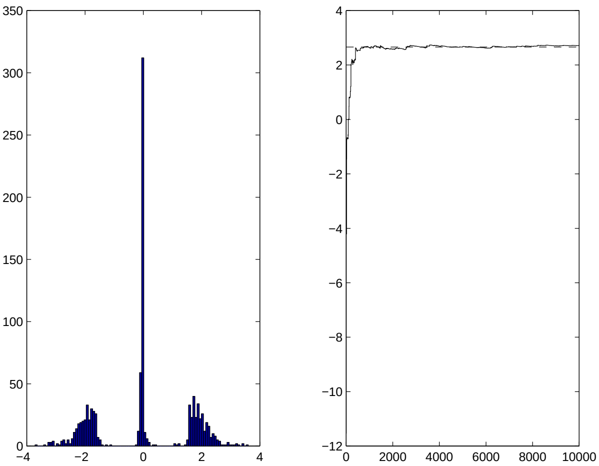

FIGURE 7. Left: histogram of the 10, 000 first samples X_k accepted by $S_\alpha(\mu)$ in Example 3. Right: Φ_k as a function of k; $\Phi[\mathbf{M}(\xi_\alpha^*)]$ corresponds to the dashed line.

We assumed throughout the paper that μ was atomless. However, as illustrated by Example 3, this condition does not seem essential. A natural extension of Theorem 4 would concern the case of measures with atoms, with an atomic part small enough for H2 to hold.

Important 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, with $(\mathcal{Z}_n)$ the family of σ-algebra generated by (Z_k), $0 \le k \le n$, and the conditional probability measure $\mu(\cdot|\mathcal{Z}_k)$ for the experimental conditions X_k is known for any k. Addressing this situation would be of interest e.g. in survey sampling, for the sequential selection of participants.

Second, applications to parameter estimation in dynamical systems 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.

Acknowledgments. This work originated from a comment by V.V. Fedorov at a meeting on experimental design held in Cardiff in April 2000, see [3]; he is gratefully acknowledged for pointing out the connection between the results in [3] and optimum design with constraints. Also, stimulating discussion with H.P. Wynn and E. Thierry at different occasions proved extremely useful.

References

[1] L. Pronzato. Optimal and asymptotically optimal decision rules for sequential screening and resource allocation. *IEEE Transactions on Automatic Control*, 46(5):687–697, 2001.

[2] L. Pronzato. Sequential selection of observations in randomly generated experiments. In *Proc. ProbaStat'98, Smolenice (Slovaquie), Feb. 98,* Tatra Mountains Mathematical Publications, volume 17, pages 167–175, 1999.

[3] L. Pronzato. Sequential construction of an experiment design from and i.i.d. sequence of experiments without replacement. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, chapter 11, pages 113–122. Kluwer, Dordrecht, 2001.

[4] S.C. Albright and C. Derman. Asymptotically optimal policies for the stochastic sequential assignment problem. *Management Science*, 19(1):46–51, 1972.

[5] H.P. Wynn. Optimum submeasures with applications to finite population sampling. In S.S. Gupta and J.O. Berger, editors, *Statistical Decision Theory and Related Topics III. Proc. 3rd Perdue Symp., vol. 2*, pages 485–495. Academic Press, New York, 1982.

[6] V.V. Fedorov. Optimal design with bounded density: optimization algorithms of the exchange type. *J. Statist. Planning and Inference*, 22:1–13, 1989.

[7] V.V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments.* Springer, Berlin, 1997.

[8] M. Sahm and R. Schwabe. A note on optimal bounded designs. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, chapter 13, pages 131–140. Kluwer, Dordrecht, 2001.

[9] I. Molchanov and S. Zuyev. Variational calculus in the space of measures and optimal design. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, chapter 8, pages 79–90. Kluwer, Dordrecht, 2001.

[10] C.F.J. Wu and H.P. Wynn. The convergence of general step–length algorithms for regular optimum design criteria. *Annals of Statistics*, 6(6):1273–1285, 1978.

[11] D. Williams. *Probability with Martingales.* Cambridge University Press, Cambridge, 1991.

[12] R.L. Jennrich. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.*, 40:633–643, 1969.

Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, bât. Euclide, Les Algorithmes, 2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France, Tel: 33 (0)4 92 94 27 08; Fax: 33 (0)4 92 94 28 98

E-mail address: pronzato@i3s.unice.fr