Listy Biometryczne - Biometrical Letters
Vol. 36(1999), No. 1, 15-30

Some results on optimal allocation in two-step sequential design*

Roland Gautier and Luc Pronzato

Les Algorithmes – Bâtiment Euclide 2000, route des Lucioles,
Sophia Antipolis, F – 06410 BIOT, France

SUMMARY

In nonlinear situations, optimal experimental conditions generally depend upon unknown parameters to be estimated from the data collected during the experiments. A natural approach then consists in designing the experiments sequentially, that is, alternating estimation and design phases. We consider the determination of the optimal allocation of the resource between two design steps when the design strategies at each step are fixed.

KEY WORDS: optimal design, sequential design, approximate design.

1. Introduction

We consider a nonlinear regression model, with observations y_k given by

$$y_k = \eta(\bar{\boldsymbol{\theta}}, x_k) + \epsilon_k \,,$$

where $\bar{\boldsymbol{\theta}} \in \Theta$ is the true value of the model parameters, with $\Theta \subset I\!R^p$ a compact set, x_k denotes the experimental conditions for the k-th observation and is assumed to belong to a compact set, $\{\epsilon_k\}$ is an i.i.d. sequence of normal variables $\mathcal{N}(0, \sigma^2)$, with σ known, and $\eta(\boldsymbol{\theta}, x)$ is the model response for the value $\boldsymbol{\theta}$ of the model parameters and experimental conditions x. The response is assumed to be nonlinear in $\boldsymbol{\theta}$ and twice continuously differentiable with respect to $\boldsymbol{\theta}$.

Experimental design for parameter estimation is considered, and we wish to choose experimental conditions $\Xi_1^N = \{x_1, \ldots, x_N\}$, with $N \geq p$ fixed, that maxi-

*The paper was submitted on the occasion of 70-th birthday of Professor Tadeusz Caliński.

mize a function $\Phi(\cdot)$ (usually concave) of the Fisher information matrix

$$\mathrm{M}(\boldsymbol{\theta}, \Xi_1^N) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{\partial \eta(\boldsymbol{\theta}, x_i)}{\partial \boldsymbol{\theta}} \frac{\partial \eta(\boldsymbol{\theta}, x_i)}{\partial \boldsymbol{\theta}^T} .$$

For instance,

$$\Phi(\mathrm{M}) = \log\det[\mathrm{M}] \quad \text{or} \quad \Phi(\mathrm{M}) = -\mathrm{trace}[\mathrm{M}^{-1}] \tag{1}$$

are classical choices. We consider the situation where the experiments can be performed sequentially in *two steps*. First, a design $\Xi^{(1)} = \Xi_1^n$ with n observations is chosen. The values $Y_1^n = \{y_1, \ldots, y_n\}$ of these observations are then used to design the experiment $\Xi^{(2)} = \Xi_{n+1}^N$ for the second stage, with $N - n$ observations. Note that the total number of observations and the number of design stages are fixed, and no convergence issue is thus considered.

The optimal strategy corresponds to the solution of a stochastic dynamic programming problem (see Bellman, 1957). Denote $\mathcal{I}^0$ the prior information on the system, which we assume to be in the form of a normal prior distribution $\mathcal{N}(\hat{\boldsymbol{\theta}}^0, \Omega_0)$ for $\boldsymbol{\theta}$, and let $\mathcal{I}^n$ denote the information available after the observations Y_1^n: $\mathcal{I}^n = \{\mathcal{I}^0, \Xi_1^n, Y_1^n\}$. The problem to be solved is:

$$\max_{n \in \{0,\ldots,N\}} \left[\max_{\Xi_1^n} \left[\mathrm{E}_{Y_1^n} \left\{ \max_{\Xi_{n+1}^N} \left[\mathrm{E}_{\theta} \{ \Phi[\mathrm{M}(\boldsymbol{\theta}, \Xi_1^N)] | \mathcal{I}^n \} \right] | \mathcal{I}^0 \right\} \right] \right], \tag{2}$$

where $\mathrm{E}_{Y_1^n}\{\cdot\}$ and $\mathrm{E}_{\theta}\{\cdot\}$ respectively denote the expectations with respect to Y_1^n and $\boldsymbol{\theta}$. This problem is very difficult to solve, even in very simple situations, see, e.g., Zacks, 1977; Pronzato et al., 1993; Kulcsár et al., 1994. A simpler version of the problem is considered in this paper: we assume that the design strategies at the two design steps are fixed, $\Xi_1^n = D^1(\mathcal{I}^0)$ and $\Xi_{n+1}^N = D^2(\mathcal{I}^n)$ are two given functions of $\mathcal{I}^0$ and $\mathcal{I}^n$ respectively, and we restrict our attention to the allocation of the resource, that is, to the determination of n.

In Section 2 the problem is put in the framework of approximate design theory. The optimization with respect to $n \in \{0, \ldots, N\}$ is replaced by the determination of the optimal allocation of resource $\alpha \in [0, 1]$. Various approximations are considered in Section 4, to facilitate the solution of the problem. We show in particular how stochastic approximation algorithms can be used to get this solution. An example is presented in Section 4. Section 5 concludes.

2. Approximate design theory

When $\Xi_1^n = D^1(\mathcal{I}^0)$ and $\Xi_{n+1}^N = D^2(\mathcal{I}^n)$ are two fixed strategies, the problem to be solved becomes:

$$\max_{n\in\{0,\ldots,N\}} [\mathrm{E}_{Y_1^n}\{\mathrm{E}_\theta\{\Phi[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N(\mathcal{I}^0, \mathcal{I}^n))]|\mathcal{I}^n\}|\mathcal{I}^0\}],$$

where $\Xi_1^N(\mathcal{I}^0, \mathcal{I}^n) = \{D^1(\mathcal{I}^0), D^2(\mathcal{I}^n)\}$. Using simple open–loop strategies for D^1 and D^2 makes this problem much simpler than (2). For instance, one may use a Forced Certainty Equivalence strategy (FCE), see Pronzato et al. (1993), Runggaldier (1993), and determine:

$$\Xi_1^n = D^1(\mathcal{I}^0) = \Xi^{FCE}(n, \hat{\boldsymbol{\theta}}^0), \quad \Xi_{n+1}^N = D^2(\mathcal{I}^n) = \Xi^{FCE}(N-n, \hat{\boldsymbol{\theta}}^n), \tag{3}$$

where $\Xi^{FCE}(k, \boldsymbol{\theta}) = \arg\max_{\Xi_1^k} \Phi[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^k)]$ and

$$\hat{\boldsymbol{\theta}}^n = \arg\max_{\boldsymbol{\theta}\in\Theta} \pi(\boldsymbol{\theta}|\mathcal{I}^n),$$

with $\pi(\boldsymbol{\theta}|\mathcal{I}^n)$ the posterior density of $\boldsymbol{\theta}$. One can also use an Open–Loop–Feedback (OLF) strategy, and determine:

$$\Xi_1^n = D^1(\mathcal{I}^0) = \Xi^{OLF}(n, \mathcal{I}^0), \quad \Xi_{n+1}^N = D^2(\mathcal{I}^n) = \Xi^{OLF}(N-n, \mathcal{I}^n), \tag{4}$$

where $\Xi^{OLF}(k, \mathcal{I}) = \arg\max_{\Xi_1^k} \mathrm{E}_\theta\{\Phi[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^k)]|\mathcal{I}\}$. The strategies (3) and (4) respectively correspond to a myopic implementation of *local* and *average* (or *Bayesian*) optimal design.

For functions $\Phi(\cdot)$ such as those suggested in (1), the strategies (3) and (4) above can only be used for $p \le n \le N-p$. An alternative would be as follows. First one determines an optimal design (FCE or OLF) with N points at stage 1, and defines Ξ_1^n by selecting any n points among those N. Then, one determines Ξ_{n+1}^N at stage 2 as an optimal augmentation design of size $N-n$:

$$\Xi^{(2)} = \arg\max_{\Xi_{n+1}^N} \Phi[\mathbf{M}(\hat{\boldsymbol{\theta}}^n, \Xi_1^N)]$$

for FCE, or

$$\Xi^{(2)} = \arg\max_{\Xi_{n+1}^N} \mathrm{E}_\theta\{\Phi[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)]|\mathcal{I}^n\}$$

for OLF, with, in both cases, Ξ_1^n defined at stage 1.

In what follows we shall use instead approximate design theory, which permits to transform the discrete optimization problem above into a continuous one. Let

$\mathbf{I}(\boldsymbol{\theta}, \Xi_1^N)$ denote the information matrix *per sample*:

$$\mathbf{I}(\boldsymbol{\theta}, \Xi_1^N) = \frac{1}{N}\mathbf{M}(\boldsymbol{\theta}, \Xi_1^N)\,.$$

One has:

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}, \Xi_1^N) &= \frac{1}{N}\left[\mathbf{M}(\boldsymbol{\theta}, \Xi_1^n) + \mathbf{M}(\boldsymbol{\theta}, \Xi_{n+1}^N)\right] \\ &= \alpha\mathbf{I}(\boldsymbol{\theta}, \Xi_1^n) + (1-\alpha)\mathbf{I}(\boldsymbol{\theta}, \Xi_{n+1}^N)\,,\end{aligned}$$

with $\alpha = n/N$. We consider normalized design measures ξ ($\int \xi(dx) = 1$), and denote:

$$\mathbf{I}(\boldsymbol{\theta}, \xi) = \frac{1}{\sigma^2}\int \frac{\partial\eta(\boldsymbol{\theta}, x)}{\partial\boldsymbol{\theta}}\frac{\partial\eta(\boldsymbol{\theta}, x)}{\partial\boldsymbol{\theta}^T}\xi(dx)\,.$$

Let ξ^1 and ξ^2 be the design measures for stages 1 and 2 respectively, we define

$$\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2) = \alpha\mathbf{I}(\boldsymbol{\theta}, \xi^1) + (1-\alpha)\mathbf{I}(\boldsymbol{\theta}, \xi^2)\,,$$

with $\alpha \in [0, 1]$. Note the difference with Schwabe (1995), where α denotes the probability that additional resource will be available.

The information $\mathcal{I}^n$ was defined for an integer number n of observations, and an information $\mathcal{J}^\alpha$ with $\alpha \in [0, 1]$ is now required. Let m be the number[1] of support points x_i of ξ^1, and μ_i denote the associated weights. The observation y_i performed at x_i, $i = 1, ..., m$, thus receives the weight $\mu_i\alpha N$, which is equivalent to having pseudo–observations z_i defined by:

$$z_i = \eta(\boldsymbol{\theta}, x_i) + \epsilon_i'\,, \tag{5}$$

where the errors ϵ_i' are normally distributed $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$, with

$$\boldsymbol{\Sigma}_\alpha = (1/\alpha)\mathrm{diag}\{\sigma^2/(N\mu_i), i = 1, \ldots, m\}\,. \tag{6}$$

We thus define $\mathcal{J}^\alpha$ as $\mathcal{J}^\alpha = \{\mathcal{I}^0, \xi^1, Z_1^m\}$, with $Z_1^m = \{z_1, \ldots, z_m\}$.

The problem to be solved then becomes: $\max_{\alpha\in[0,1]} f(\alpha, \mathcal{I}^0)$, where

$$f(\alpha, \mathcal{I}^0) = E_{Z_1^m}\{E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2)]|\mathcal{J}^\alpha\}|\mathcal{I}^0\}\,,$$

with $\xi^1 = D^1(\mathcal{I}^0)$ and $\xi^2 = D^2(\mathcal{J}^\alpha)$ two given design strategies. This function satisfies the following properties.

An upper bound on this number exists when ξ^1 is a local optimal design (FCE strategy), see, e.g., Fedorov (1972), Silvey (1980). No such upper bound is available when average design (OLF strategy) is used, see, e.g., Chaloner and Verdinelli (1995), but a constraint on this number can be settled during the determination of the design.

THEOREM 1. *Assume that the same design strategy is used at stages 1 and 2, that is,* $\xi^1 = D(\mathcal{I}^0)$ *and* $\xi^2 = D(\mathcal{J}^\alpha)$. *Then* $f(0,\mathcal{I}^0) = f(1,\mathcal{I}^0)$.

Proof. When $\alpha = 0$, no observation is taken at stage 1 and $\mathcal{J}^\alpha = \mathcal{I}^0$ so that

$$f(0,\mathcal{I}^0) = E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, D^2(\mathcal{I}^0))]|\mathcal{I}^0\}\,.$$

When $\alpha = 1$, one gets

$$\begin{aligned} f(1,\mathcal{I}^0) &= E_{Z_1^m}\{E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, D^1(\mathcal{I}^0))]|\mathcal{J}^1\}|\mathcal{I}^0\} \\ &= E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, D^1(\mathcal{I}^0))]|\mathcal{I}^0\}\,. \end{aligned}$$

Choosing $D^1(\cdot) = D^2(\cdot)$ gives the result. □

THEOREM 2. *Assume that OLF is used at stage 2, that is,*

$$\xi^2 = \xi^{OLF}(\mathcal{J}^\alpha) = \arg\max_{\xi^2}[E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2)]|\mathcal{J}^\alpha\}]\,.$$

Then, $\forall \xi^1 = D^1(\mathcal{I}^0)$, $\forall \alpha \in [0,1)$, $f(\alpha,\mathcal{I}^0) \geq f(1,\mathcal{I}^0)$.

Proof. For any $\alpha \in [0,1)$, one has:

$$\begin{aligned} f(\alpha,\mathcal{I}^0) &\geq \max_{\xi^2}[E_{Z_1^m}\{E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, D^1(\mathcal{I}^0), \xi^2)]|\mathcal{J}^\alpha\}|\mathcal{I}^0\}] \\ &= \max_{\xi^2}[E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, D^1(\mathcal{I}^0), \xi^2)]|\mathcal{I}^0\}] \\ &\geq E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, D^1(\mathcal{I}^0))]|\mathcal{I}^0\} = f(1,\mathcal{I}^0). \end{aligned}$$

□

Remarks:

(i) When $\alpha > 0$, the inequality in Theorem 2 is strict if there exist two vectors Z_1^m and $Z_1'^{m'}$ such that

$$\max_{\xi^2}[E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2)]|\mathcal{J}^\alpha\}] \neq \max_{\xi^2}[E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2)]|\mathcal{J}'^\alpha\}]$$

where $\mathcal{J}'^\alpha = \{\mathcal{I}^0, \xi^1, Z_1'^{m'}\}$, which is the case except in pathological situations.

(ii) Theorem 2 simply expresses that any experiment, on the average, reduces the Bayesian risk, see Ivanenko and Labkovskii (1979). Such a property is not valid in general when FCE is used at stage 2, see Thau and Witsenhausen (1966) for a counter–example in the context of control theory. However, numerical simulations indicate that in general $f(\alpha,\mathcal{I}^0) > f(1,\mathcal{I}^0)$ for $\alpha \in (0,1)$ even if FCE is used.

(iii) Even when $\Phi(\cdot)$ is a concave function [(such as those proposed in (1)], concavity of $f(\alpha,\mathcal{I}^0)$ with respect to α is difficult to be proved due to the dependence of ξ^2 on $\mathcal{J}^\alpha$, and the dependence of the distribution of Z_1^m on α. This difficulty exists even if there is no explicit dependence of ξ^2 on α, as is the case for instance when $\xi^2 = \xi^{FCE}(\mathcal{J}^\alpha)$.

(iv) When σ^2 tends to 0, the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{J}^\alpha)$ tends to be concentrated around the estimator

$$\hat{\boldsymbol{\theta}}^\alpha = \arg\max_{\boldsymbol{\theta}\in\Theta} \pi(\boldsymbol{\theta}|\mathcal{J}^\alpha)\,, \tag{7}$$

and $E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta},\alpha,\xi^1,\xi^2)]|\mathcal{J}^\alpha\}$ tends to $\Phi[\mathbf{I}(\hat{\boldsymbol{\theta}}^\alpha,\alpha,\xi^1,\xi^2)]$. Reasonable choices for ξ^2 are such that $\xi^2 = D^2(\mathcal{J}^\alpha)$ tends to $\xi^{FCE}(\mathcal{J}^\alpha)$ when σ^2 tends to zero, so that the maximum of $\Phi[\mathbf{I}(\hat{\boldsymbol{\theta}}^\alpha,\alpha,\xi^1,\xi^2)]$ is reached for a value of α close to 0 when σ^2 tends to 0. This is in agreement with the approach suggested in Chapter 5 of Ermakov (1983), for which, when N tends to infinity, $n = \beta(N)$ observations should be taken at stage 1, with $\beta(N) \to \infty$ and $\beta(N) = o(N)$.

3. Approximations and algorithm

The evaluation of $f(\alpha,\mathcal{I}^0)$ requires the computation of two integrations, with respect to $\boldsymbol{\theta}$ and Z_1^m respectively. Different approaches can be used to approximate these integrals.

3.1. Approximating the posterior mean

We consider the evaluation of $E_\theta\{\Phi[\mathbf{I}(\boldsymbol{\theta},\alpha,\xi^1,\xi^2)]|\mathcal{J}^\alpha\}$, where $\xi^1 = D^1(\mathcal{I}^0)$ and $\xi^2 = D^2(\mathcal{J}^\alpha)$. We shall denote

$$g(\boldsymbol{\theta},\alpha) = \Phi[\mathbf{I}(\boldsymbol{\theta},\alpha,\xi^1,\xi^2)]\,,$$

and

$$P(\boldsymbol{\theta},\alpha) = -\log[\pi(Z_1^m|\boldsymbol{\theta})\pi(\boldsymbol{\theta})]\,, \tag{8}$$

where $\pi(Z_1^m|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ are respectively the likelihood of Z_1^m and the prior density of $\boldsymbol{\theta}$ (its support is assumed to be restricted to Θ). We thus have to evaluate

$$E_\theta\{g(\boldsymbol{\theta},\alpha)|\mathcal{J}^\alpha\} = \frac{\int g(\boldsymbol{\theta},\alpha)\exp[-P(\boldsymbol{\theta},\alpha)]d\boldsymbol{\theta}}{\int \exp[-P(\boldsymbol{\theta},\alpha)]d\boldsymbol{\theta}}\,. \tag{9}$$

A first approximation is obtained by replacing $g(\boldsymbol{\theta},\alpha)$ and $P(\boldsymbol{\theta},\alpha)$ by their second–order development in $\boldsymbol{\theta}$ around $\hat{\boldsymbol{\theta}}^\alpha$, and computing the corresponding two integrals in (9). This gives:

$$E_\theta\{g(\boldsymbol{\theta},\alpha)|\mathcal{J}^\alpha\} \approx g(\hat{\boldsymbol{\theta}}^\alpha,\alpha) + \frac{1}{2}\text{trace}\left[\frac{\partial^2 g(\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\hat{\boldsymbol{\theta}}^\alpha}\left(\frac{\partial^2 P(\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\hat{\boldsymbol{\theta}}^\alpha}\right)^{-1}\right]\,,$$

with $\hat{\boldsymbol{\theta}}^{\alpha}$ given by (7). Since

$$P(\boldsymbol{\theta},\alpha)=\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^{0})^{T}\boldsymbol{\Omega}_{0}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^{0})+\frac{N\alpha}{2\sigma^{2}}\sum_{i=1}^{m}\mu_{i}[z_{i}-\eta(\boldsymbol{\theta},x_{i})]^{2}\,,$$

one gets

$$\begin{aligned}\frac{\partial^{2}P(\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}_{|\hat{\boldsymbol{\theta}}^{\alpha}} &= \boldsymbol{\Omega}_{0}^{-1}+\frac{N\alpha}{\sigma^{2}}\sum_{i=1}^{m}\mu_{i}\frac{\partial\eta(\boldsymbol{\theta},x_{i})}{\partial\boldsymbol{\theta}}_{|\hat{\boldsymbol{\theta}}^{\alpha}}\frac{\partial\eta(\boldsymbol{\theta},x_{i})}{\partial\boldsymbol{\theta}^{T}}_{|\hat{\boldsymbol{\theta}}^{\alpha}}\\ &\quad+\frac{N\alpha}{\sigma^{2}}\sum_{i=1}^{m}\mu_{i}\frac{\partial^{2}\eta(\boldsymbol{\theta},x_{i})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}_{|\hat{\boldsymbol{\theta}}^{\alpha}}(\eta(\boldsymbol{\theta},x_{i})-z_{i})\,.\end{aligned}$$

Note that the second term on the right–hand side is equal to $N\alpha\mathbf{I}(\hat{\boldsymbol{\theta}}^{\alpha},\xi^{1})$, so that neglecting the errors $(\eta(\boldsymbol{\theta},x_{i})-z_{i})$, one obtains for $\partial^{2}P(\boldsymbol{\theta},\alpha)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}$ the approximate Hessian used in the Gauss–Newton algorithm for the computation of $\hat{\boldsymbol{\theta}}^{\alpha}$, which also coincides with the inverse of the covariance matrix for the usual normal approximation of the posterior density of $\boldsymbol{\theta}$.

A more precise approximation can be used when $g(\boldsymbol{\theta},\alpha)$ is positive (which can always be obtained provided that $g(\boldsymbol{\theta},\alpha)\geq M>-\infty$ by adding a suitable constant to $g(\boldsymbol{\theta},\alpha)$). In this case, we define

$$Q(\boldsymbol{\theta},\alpha)=P(\boldsymbol{\theta},\alpha)-\log[g(\boldsymbol{\theta},\alpha)]\,, \tag{10}$$

with $P(\boldsymbol{\theta},\alpha)$ still given by (8), and use Laplace approximation for computing the two integrals in (9): we replace $P(\boldsymbol{\theta},\alpha)$ and $Q(\boldsymbol{\theta},\alpha)$ by their second–order expansion in $\boldsymbol{\theta}$ around their minimum. One gets (Tierney and Kadane, 1986; Tanner, 1993):

$$E_{\theta}\{g(\boldsymbol{\theta},\alpha)|\mathcal{J}^{\alpha}\}\approx g(\tilde{\boldsymbol{\theta}}^{\alpha},\alpha)\frac{\pi(Z_{1}^{m}|\tilde{\boldsymbol{\theta}}^{\alpha})\pi(\tilde{\boldsymbol{\theta}}^{\alpha})\det^{-1/2}\left(\dfrac{\partial^{2}Q(\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}_{|\tilde{\boldsymbol{\theta}}^{\alpha}}\right)}{\pi(Z_{1}^{m}|\hat{\boldsymbol{\theta}}^{\alpha})\pi(\hat{\boldsymbol{\theta}}^{\alpha})\det^{-1/2}\left(\dfrac{\partial^{2}P(\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}_{|\hat{\boldsymbol{\theta}}^{\alpha}}\right)}\,,$$

where

$$\tilde{\boldsymbol{\theta}}^{\alpha}=\arg\min_{\boldsymbol{\theta}\in\Theta}Q(\boldsymbol{\theta},\alpha)\,, \tag{11}$$

and $\hat{\boldsymbol{\theta}}^{\alpha}$ is given by (7). When the normal approximation $\mathcal{N}(\hat{\boldsymbol{\theta}}^{\alpha},[\boldsymbol{\Omega}_{0}^{-1}+N\alpha\mathbf{I}(\hat{\boldsymbol{\theta}}^{0},\xi^{1})]^{-1})$ is used for the posterior distribution of $\boldsymbol{\theta}$, the Laplace approximation gives:

$$E_{\theta}\{g(\boldsymbol{\theta},\alpha)|\mathcal{J}^{\alpha}\}\approx g(\tilde{\boldsymbol{\theta}}^{\alpha},\alpha)\exp[-P(\tilde{\boldsymbol{\theta}}^{\alpha},\alpha)]\frac{\det^{-1/2}\left(\dfrac{\partial^{2}Q(\boldsymbol{\theta}\boldsymbol{\theta},\alpha)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}_{|\tilde{\boldsymbol{\theta}}^{\alpha}}\right)}{\det^{1/2}[\boldsymbol{\Omega}_{0}^{-1}+N\alpha\mathbf{I}(\hat{\boldsymbol{\theta}}^{0},\xi^{1})]^{-1}}\,,$$

where $Q(\boldsymbol{\theta}, \alpha)$ and $\tilde{\boldsymbol{\theta}}^{\alpha}$ are still given by (10) and (11), with now

$$P(\boldsymbol{\theta}, \alpha) = \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{\alpha})^T [\boldsymbol{\Omega}_0^{-1} + N\alpha \mathbf{I}(\hat{\boldsymbol{\theta}}^0, \xi^1)](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{\alpha}) .$$

The approximations above permits to express $E_{\theta}\{g(\boldsymbol{\theta}, \alpha) | \mathcal{J}^{\alpha}\}$ as a function of $\mathcal{J}^{\alpha}$, the expectation of which with respect to Z_1^m then needs to be evaluated. The distribution of Z_1^m can be approximated by a normal distribution:

$$Z_1^m \sim \mathcal{N}\left(\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}^0, \xi^1), \boldsymbol{\Sigma}_{\alpha} + \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta}, \xi^1)}{\partial \boldsymbol{\theta}^T}_{|\hat{\boldsymbol{\theta}}^0} \boldsymbol{\Omega}_0 \frac{\partial \boldsymbol{\eta}^T(\boldsymbol{\theta}, \xi^1)}{\partial \boldsymbol{\theta}}_{|\hat{\boldsymbol{\theta}}^0}\right) ,$$

where $\boldsymbol{\eta}(\boldsymbol{\theta}, \xi^1) = [\eta(\boldsymbol{\theta}, x_1), \ldots, \eta(\boldsymbol{\theta}, x_m)]^T$.

When $\xi^2 = D^2(\mathcal{J}^{\alpha})$ depends only on $\hat{\boldsymbol{\theta}}^{\alpha}$, the dependence of $E_{\theta}\{g(\boldsymbol{\theta}, \alpha) | \mathcal{J}^{\alpha}\}$ on $\mathcal{J}^{\alpha}$ is only through $\hat{\boldsymbol{\theta}}^{\alpha}$, and a normal approximation can be used for the predictive distribution of $\hat{\boldsymbol{\theta}}^{\alpha}$:

$$\hat{\boldsymbol{\theta}}^{\alpha} \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}^0, \boldsymbol{\Omega}_0 - [\boldsymbol{\Omega}_0^{-1} + N\alpha \mathbf{I}(\hat{\boldsymbol{\theta}}^0, \xi^1)]^{-1}\right) .$$

3.2. Approximating $f(\alpha, \mathcal{I}^0)$

Define

$$G(\boldsymbol{\theta}, Z_1^m, \alpha) = \Phi[\mathbf{I}(\boldsymbol{\theta}, \alpha, \xi^1, \xi^2)] .$$

One has to evaluate

$$\begin{aligned} f(\alpha, \mathcal{I}^0) &= \int G(\boldsymbol{\theta}, Z_1^m, \alpha) \pi(\boldsymbol{\theta} | \mathcal{J}^{\alpha}) \pi(Z_1^m) d\boldsymbol{\theta} dZ_1^m \\ &= \int G(\boldsymbol{\theta}, Z_1^m, \alpha) \pi(Z_1^m | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} dZ_1^m . \end{aligned} \tag{12}$$

As in Section 3.1, one can use a second–order expansion of $G(\boldsymbol{\theta}, Z_1^m, \alpha)$ and

$$P(\boldsymbol{\theta}, Z_1^m, \alpha) = -\log[\pi(Z_1^m | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})]$$

around $(\hat{\boldsymbol{\theta}}, \hat{Z}_1^m) = \arg\min_{\boldsymbol{\theta} \in \Theta, Z_1^m} P(\boldsymbol{\theta}, Z_1^m, \alpha)$. One gets $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^0$, $\hat{z}_i = \eta(\hat{\boldsymbol{\theta}}^0, x_i)$, and

$$\begin{aligned} f(\alpha, \mathcal{I}^0) \approx \pi(\hat{Z}_1^m | \hat{\boldsymbol{\theta}}^0) \pi(\hat{\boldsymbol{\theta}}^0) (2\pi)^{(m+p)/2} \det{}^{-\frac{1}{2}} \left[\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m}\right] \\ \times \left\{ G(\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m, \alpha) + \tfrac{1}{2} \mathrm{trace} \left[\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m} \left(\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m}\right)^{-1}\right] \right\} , \end{aligned}$$

where

$$\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha) = \begin{pmatrix} \frac{\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} & \frac{\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial \boldsymbol{\theta} \partial Z_1^{mT}} \\ \frac{\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial Z_1^m \partial \boldsymbol{\theta}^T} & \frac{\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial Z_1^m \partial Z_1^{mT}} \end{pmatrix},$$

and

$$\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha) = \begin{pmatrix} \frac{\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} & \frac{\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial \boldsymbol{\theta} \partial Z_1^{mT}} \\ \frac{\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial Z_1^m \partial \boldsymbol{\theta}^T} & \frac{\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)}{\partial Z_1^m \partial Z_1^{mT}} \end{pmatrix}.$$

This gives

$$\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m} = \begin{pmatrix} \boldsymbol{\Omega}_0^{-1} + N\alpha \mathbf{I}(\hat{\boldsymbol{\theta}}^0, \xi^1) & \mathbf{H}_\alpha(\hat{\boldsymbol{\theta}}^0) \\ \mathbf{H}_\alpha^T(\hat{\boldsymbol{\theta}}^0) & \boldsymbol{\Sigma}_\alpha^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_\alpha$ is given by (6) and

$$[\mathbf{H}_\alpha(\boldsymbol{\theta})]_{i,j} = -\frac{N\alpha}{\sigma^2} \mu_j \frac{\partial \eta(\boldsymbol{\theta}, x_j)}{\partial \theta_i},$$

so that

$$\det\left[\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m}\right] = \det[\boldsymbol{\Sigma}_\alpha^{-1}] \det[\boldsymbol{\Omega}_0^{-1}].$$

One finally gets:

$$f(\alpha, \mathcal{I}^0) \approx G(\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m, \alpha) + \frac{1}{2} \text{trace}\left[\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m} \left(\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m}\right)^{-1}\right], \quad (13)$$

with

$$\left(\partial^2 P(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\hat{\boldsymbol{\theta}}^0, \hat{Z}_1^m}\right)^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_0 & -\boldsymbol{\Omega}_0 \mathbf{H}_\alpha(\hat{\boldsymbol{\theta}}^0) \boldsymbol{\Sigma}_\alpha \\ -\boldsymbol{\Sigma}_\alpha \mathbf{H}_\alpha^T(\hat{\boldsymbol{\theta}}^0) \boldsymbol{\Omega}_0 & \boldsymbol{\Sigma}_\alpha + \boldsymbol{\Sigma}_\alpha \mathbf{H}_\alpha^T(\hat{\boldsymbol{\theta}}^0) \boldsymbol{\Omega}_0 \mathbf{H}_\alpha(\hat{\boldsymbol{\theta}}^0) \boldsymbol{\Sigma}_\alpha \end{pmatrix}.$$

Again, a more precise approximation can be obtained when $G(\boldsymbol{\theta}, Z_1^m, \alpha) > 0$. Define

$$Q(\boldsymbol{\theta}, Z_1^m, \alpha) = P(\boldsymbol{\theta}, Z_1^m, \alpha) - \log[G(\boldsymbol{\theta}, Z_1^m, \alpha)],$$

and $(\tilde{\boldsymbol{\theta}}, \tilde{Z}_1^m) = \arg\min_{\boldsymbol{\theta} \in \Theta, Z_1^m} Q(\boldsymbol{\theta}, Z_1^m, \alpha)$. One has, see Tierney and Kadane (1986), Tanner (1993),

$$f(\alpha, \mathcal{I}^0) \approx G(\tilde{\boldsymbol{\theta}}, \tilde{Z}_1^m, \alpha) \pi(\tilde{Z}_1^m | \tilde{\boldsymbol{\theta}}) \pi(\tilde{\boldsymbol{\theta}}) (2\pi)^{(m+p)/2} \det\left[\partial^2 Q(\boldsymbol{\theta}, Z_1^m, \alpha)_{|\tilde{\boldsymbol{\theta}}, \tilde{Z}_1^m}\right]^{-1/2}. \quad (14)$$

The approximations (13) and (14) require the computation of the first and second–order derivatives $\partial G(\boldsymbol{\theta}, Z_1^m, \alpha)/\partial Z_1^m$ and $\partial^2 G(\boldsymbol{\theta}, Z_1^m, \alpha)/\partial Z_1^m \partial Z_1^{mT}$. When ξ^2 only depends on $\mathcal{J}^\alpha$ through $\hat{\boldsymbol{\theta}}^\alpha$, $G(\boldsymbol{\theta}, Z_1^m, \alpha)$ can be written as $G'[\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(Z_1^m), \alpha]$, and the first and second–order derivatives of $G'(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}, \alpha)$ with respect to $\hat{\boldsymbol{\theta}}$ and of $\hat{\boldsymbol{\theta}}(Z_1^m)$ with respect to Z_1^m are thus required. The implicit–function theorem can then be used to compute $\partial G'(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}, \alpha)/\partial \hat{\boldsymbol{\theta}}$ and $\partial^2 G'(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}, \alpha)/\partial \hat{\boldsymbol{\theta}} \partial \hat{\boldsymbol{\theta}}^T$, see Pronzato and Pázman (1994), Pázman, (1993).

3.3. Stochastic approximation

Stochastic approximation techniques (see, e.g., Kushner and Yin, 1997) can be used to maximize the expression of $f(\alpha, \mathcal{I}^0)$ given by (12) with respect to α, without requiring any evaluation or approximation of $f(\alpha, \mathcal{I}^0)$.

First note that, conditionally on $\boldsymbol{\theta}$, Z_1^m is generated from m i.i.d. variables $\mathcal{N}(0,1)$, according to:

$$z_i = \eta(\boldsymbol{\theta}, x_i) + \frac{\sigma}{N\alpha\mu_i} e_i \,,$$

see (5,6). We can thus write $Z_1^m = \zeta(\alpha, \boldsymbol{\theta}, \mathbf{e}_1^m)$, with $\mathbf{e}_1^m = (e_1, \ldots, e_m)$, and $G(\boldsymbol{\theta}, Z_1^m, \alpha) = \Gamma(\boldsymbol{\theta}, \mathbf{e}_1^m, \alpha)$.

The simplest version of the algorithm is as follows:

$$\alpha_{k+1} = \alpha_k + a_k \frac{d\Gamma(\boldsymbol{\theta}^{(k)}, \mathbf{e}_1^{m(k)}, \alpha)}{d\alpha}_{|\alpha_k} \,, \tag{15}$$

where, at iteration k, $\boldsymbol{\theta}^{(k)}$ and $\mathbf{e}_1^{m(k)}$ are independently distributed, respectively $\pi(\boldsymbol{\theta})$ and $\mathcal{N}(0,1)$. The sequence a_k satisfies

$$a_k > 0 \,, \quad \sum_{k=1}^{\infty} a_k = \infty \,, \quad \sum_{k=1}^{\infty} (a_k)^2 < \infty \,.$$

A typical choice is $a_k = A/k$, with A a positive constant. Note that taking $a_k = A/k^\beta$, with $\beta \in (0,1)$, and averaging the iterates, that is taking

$$\hat{\alpha}_k = \frac{1}{k} \sum_{i=1}^{k} \alpha_k \,, \tag{16}$$

is known to yield a faster convergence. In practice the derivative $d\Gamma(\boldsymbol{\theta}^{(k)}, \mathbf{e}_1^{m(k)}, \alpha)/d\alpha_{|\alpha_k}$ is evaluated by finite differences,

$$\frac{d\Gamma(\boldsymbol{\theta}^{(k)}, \mathbf{e}_1^{m(k)}, \alpha)}{d\alpha}_{|\alpha_k} = \frac{\Gamma(\boldsymbol{\theta}^{(k)}, \mathbf{e}_1^{m(k)}, \alpha_k + \delta) - \Gamma(\boldsymbol{\theta}^{(k)}, \mathbf{e}_1^{m(k)}, \alpha_k)}{\delta} \tag{17}$$

with δ small enough.

Remark:
When the number m of support points of ξ^1 is larger than the dimension p of $\boldsymbol{\theta}$, and when the dependence of $G(\boldsymbol{\theta}, Z_1^m, \alpha)$ in Z_1^m is only through $\hat{\boldsymbol{\theta}}^\alpha$, it might be advantageous to directly generate vectors of estimates in the stochastic approximation algorithm. The approximation given in Pázman and Pronzato (1992) and Pázman (1993) would then be useful. It could also be used to construct an approximation of $f(\alpha, \mathcal{I}^0)$, as done in Section 3.2.

4. Example

Consider the model of exponential decay $\eta(\theta, x) = \exp(-\theta x)$, with θ scalar and $I(\theta, \xi) = (1/\sigma^2) \int x^2 \exp(-2\theta x) \xi(dx)$ the criterion to be maximized. The prior density for θ is normal $\mathcal{N}(1, 0.01)$.

Assume that FCE is used at both design stages, so that ξ^1 and ξ^2 have one support point, respectively at $1/\hat{\theta}^0$ and $1/\hat{\theta}^\alpha$.

Fig. 1 presents $f(\alpha, \mathcal{I}^0)$ obtained by different methods, when $\sigma^2 = 0.01$ and $N = 100$. Note that $f(0, \mathcal{I}^0) = f(1, \mathcal{I}^0)$, in agreement with Theorem 1, and that $f(\alpha, \mathcal{I}^0)$ is a concave function of α.

Numerical integration is used for the curve in full line, with respect to θ, with the prior distribution $\pi(\theta)$, and for each θ with respect to Z_1^1 with the conditional distribution $\pi(Z_1^1|\theta)$. For the curve in dashed line, normal approximations are used for the posterior $\pi(\theta|Z_1^1)$ and the marginal $\pi(Z_1^1)$, see Section 3.1. The expectation with respect to θ can then be calculated analytically, and a numerical integration is used for the expectation with respect to Z_1^1. The curves in dash–dotted line and in dotted line respectively correspond to the approximations (13) and (14). Note the good agreement of the approximation (14) with the function obtained by numerical integration (full line).

Fig. 2 gives the evolution of α_k given by (15) (respectively of $\hat{\alpha}_k$ given (16)) generated by the stochastic approximation method of Section 3.3, with $a_k = 100/k$ (respectively $a_k = 100/k^{0.5}$).

Consider now the case where the parameters are estimated by least–squares at the second design stage and FCE is used. The two design stages thus differ (FCE with Bayesian estimation is used at stage 1, FCE with LS at stage 2), and Theorem 1 no longer applies: when α decreases, the information collected at stage 1 decreases too, and the experiment designed at stage 2 becomes very poor. Fig. 3 presents $f(\alpha, \mathcal{I}^0)$ obtained when $\sigma^2 = 0.01$ and $N = 100$. The curve in full line is obtained by numerical integration, the curve in dash–dotted line corresponds to the approximation (13). The curve corresponding to the approximation (14) is not distinguishable from the curve in full line.

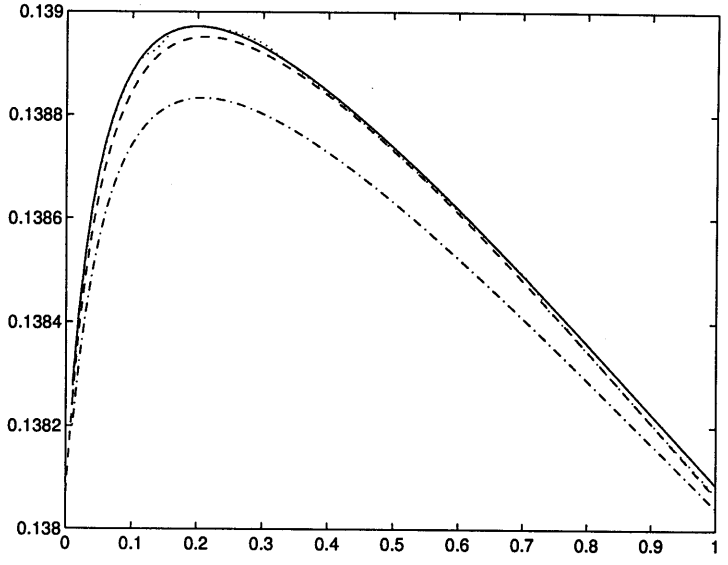

Figure 1. $f(\alpha, \mathcal{I}^0)$ as a function of α (FCE with Bayesian estimation is used at stage 2). Full line: numerical integration, dashed line: normal approximation and numerical integration, dash-dotted line: approximation (13), dotted line: approximation (14).

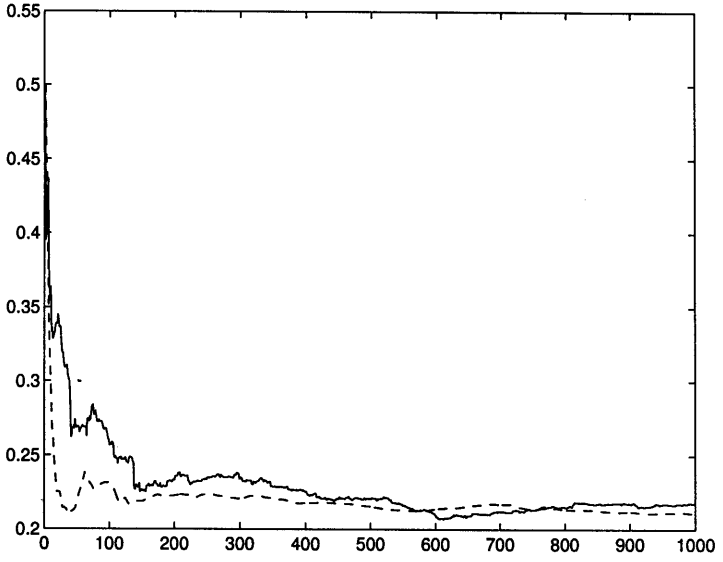

Figure 2. Evolution of α_k (full line) and $\hat{\alpha}_k$ (dashed line) as functions of k

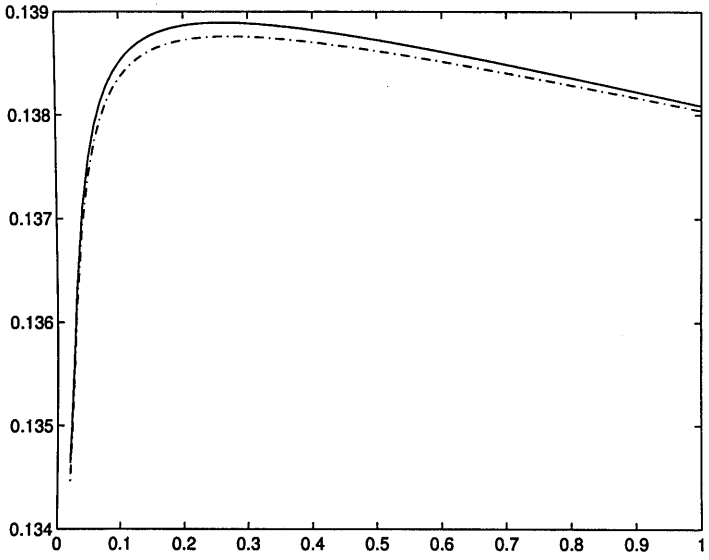

Figure 3. $f(\alpha, \mathcal{I}^0)$ as a function of α (FCE with LS estimation is used at stage 2). Full line: numerical integration, dash-dotted line: approximation (13).

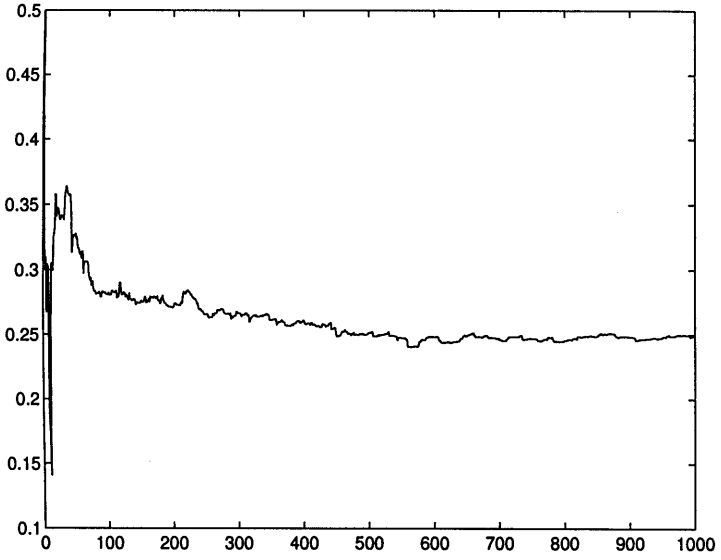

Figure 4. Evolution of α_k as a function of k

Fig. 4 gives the evolution of α_k given by (15) generated by the stochastic approximation method of Section 3.3, with $a_k = 100/k$.

5. Conclusions and further developments

Different objectives could be considered at the two design stages. In particular, one may wish to discriminate between model structures at the first stage, and then estimate the parameters of the structure retained in the second stage. The evaluation of the performance of some two–stage designs is considered in Montepiedra and Yeh (1998) in the case of linear models, for which the optimal designs do not depend on the value of the parameters of the models. It would be of interest to investigate how the results presented here could be extended to this discrimination/estimation problem for nonlinear models. Note that classical methods (Hill et al., 1968; Borth, 1975; Huang, 1991) are fully sequential: a new support point is chosen after each observation, so that this problem of resource allocation seems widely open.

A straightforward extension would consist in increasing the number of design stages considered. Let M be this number, with α_i, $i = 1, \ldots, M$, the resources allocated to these stages. The problem to be solved would then be (with notations similar to those of Section 2) $\max_{\boldsymbol{\alpha} \in \mathcal{S}^M} f(\boldsymbol{\alpha}, \mathcal{I}^0)$, where

$$f(\boldsymbol{\alpha}, \mathcal{I}^0) = E_{Z_1^{m_1+\cdots+m_{M-1}}} \{ E_\theta \{ \Phi[\mathbf{I}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \xi^1, \ldots, \xi^M)] | \mathcal{J}^{\alpha_{M-1}} \} | \mathcal{I}^0 \} ,$$

where $\xi^1 = D^1(\mathcal{I}^0)$, $\xi^i = D^i(\mathcal{J}^{\alpha_{i-1}})$, $i = 2, \ldots, M$, and $\mathcal{J}^{\alpha_i} = \{\mathcal{I}^0, \xi^1, \ldots, \xi^i, Z_1^{m_1+\cdots+m_i}\}$, with m_i the number of support points of ξ^i, and where $\mathcal{S}^M$ denotes the M-dimensional canonical simplex:

$$\mathcal{S}^M = \{ \boldsymbol{\alpha} \in \mathbb{R}^M \mid \alpha_i \geq 0 , \sum_{i=1}^{M} \alpha_i = 1 \} .$$

References

Bellman R. (1957). *Dynamic Programming.* Princeton University Press, Princeton, N.J.

Borth D. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. *Journal of Royal Statistical Society* **B37**, 77–87.

Chaloner K. and Verdinelli I. (1995). Bayesian experimental design: a review. *Statistical Science* **10**, 273–304.

Ermakov C. (1983). *Mathematical Theory of Experimental Design.* Nauka, Moscow (in Russian).

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

Hill W., Hunter W. and Wichern D. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics* **10**, 145–160.

Huang C.-Y. (1991). Planification d'expériences pour la discrimination entre structures de modèles. Thèse de Doctorat, Université Paris XI Orsay.

Ivanenko V. and Labkovskii V. (1979). Uncertainty function of Bayes systems. *Sov. Phys. Dokl.*, **24**, 703–704.

Kulcsár C., Pronzato L. and Walter E. (1994). Optimal experimental design and therapeutic drug monitoring. *Int. Journal of Biomedical Computing* **36**, 95–101.

Kushner H. and Yin G. (1997). *Stochastic Approximation Algorithms and Applications.* Springer, Heidelberg.

Montepiedra G. and Yeh A. (1998). Two–stage designs for model discrimination and parameter estimation. In: A. Atkinson, L. Pronzato and H. Wynn (Eds.), *Advances in Model–Oriented Data Analysis and Experimental Design, Proceedings of MODA'5, Marseilles, June 22–26, 1998*, 195–203. Physica Verlag, Heidelberg.

Pázman A. (1993). *Nonlinear Statistical Models.* Kluwer, Dordrecht.

Pázman A. and Pronzato L. (1992). Nonlinear experimental design based on the distribution of estimators. *Journal of Statistical Planning and Inference* **33**, 385–402.

Pronzato L. and Pázman A. (1994). Second-order approximation of the entropy in nonlinear least-squares estimation. *Kybernetika* **30**, 187–198. (Erratum, 32(1), 104, 1996).

Pronzato L., Walter E. and Kulcsár C. (1993). A dynamical-system approach to sequential design. In: W. Müller, H. Wynn and A. Zhigljavsky (Eds.), *Model-Oriented Data Analysis III, Proceedings MODA3, St Petersburg, May 1992*, 11–24. Physica Verlag, Heidelberg.

Runggaldier W. (1993). Concepts of optimality in stochastic control. In: R. Barlow *et al.*, (Eds.), *Reliability and Decision Making*, 101–114. Elsevier, Amsterdam.

Schwabe R. (1995). *Uncertain resources and designing for additional information.* Technical Report A-17, Mathematisches Insitüt, Freie Universität Berlin.

Silvey S. (1980). *Optimal Design.* Chapman & Hall, London.

Tanner M. (1993). *Tools for Statistical Inference. Methods for Exploration of Posterior Distributions and Likelihood Functions.* Springer, Heidelberg.

Thau F. and Witsenhausen H. (1966). A comparison of closed-loop and open-loop optimum systems. *IEEE Transactions on Automatic Control* **11**, 619–621.

Tierney L. and Kadane J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86.

Zacks S. (1977). Problems and approaches in design of experiments for estimation and testing in nonlinear models. In: P. Krishnaiah (Ed.), *Multivariate Analysis IV*, 209–223. North Holland, Amsterdam.

Received 10 August 1998; revised 5 January 1999

O optymalnym przydziale zasobów w dwufazowych układach sekwencyjnych

STRESZCZENIE

W sytuacjach nieliniowych optymalny plan doświadczenia zależy od nieznanych parametrów, które mogą być estymowane na podstawie danych zebranych w tymże doświadczeniu. Naturalnym podejściem jest wtedy zakładanie doświadczenia w sposób sekwencyjny, powtarzając kolejno fazy estymacji i planowania. W pracy rozważany jest problem optymalnego przydziału zasobów doświadczalnych do wymienionych dwu faz przy założeniu ustalonej strategii w ramach każdej z nich.

SŁOWA KLUCZOWE: układ optymalny, układ sekwencyjny, układ przybliżony.