Probability and Statistics 20 (2000) 97–114

ADAPTIVE CONTROL FOR SEQUENTIAL DESIGN

Roland Gautier and Luc Pronzato
CNRS/Université de Nice Sophia Antipolis, France

Abstract

The optimal experiment for estimating the parameters of a nonlinear regression model usually depends on the value of these parameters, hence the problem of designing experiments that are *robust* with respect to parameter uncertainty. Sequential design permits to adapt the experiment to the value of the parameters, and can thus be considered as a robust design procedure. By designing the experiments sequentially, one introduces a feedback of information, and thus dynamics, into the design procedure. Several sequential schemes, corresponding to different control policies, are considered. The optimal one corresponds to closed-loop control, and is solution of a stochastic dynamic-programming problem, which is extremely difficult to solve. A suboptimal strategy is proposed, which relies on a normal approximation of the future posterior of $\boldsymbol{\theta}$, independent of future observations. The design criterion obtained involves several mathematical expectations, which are approximated by Laplace method. Finally, stochastic approximation algorithms are also suggested to determine (sub)optimal sequential experiments without having to compute expectations.

Keywords and phrases: active control, adaptive control, certainty equivalence, closed-loop control, dynamic programming, open-loop feedback, optimal design, sequential design.

1991 Mathematics Subject Classifications: Primary 62L05, 62K05; Secondary 49L20, 49N35.

1. Introduction

The optimal experimental conditions for estimating the parameters of a nonlinear regression model generally depend upon the value of the parameters to be estimated. Since this value is unknown, designing an experiment optimally for a badly chosen nominal value of the parameters may lead to poor results. A natural approach to face this robustness problem then consists in designing the experiments sequentially, that is alternating estimation and design phases. Each design phase can be considered as a control action applied on the system, so that sequential design corresponds to adaptive control, see [10]. For instance, usual non-Bayesian miopic sequential design can be considered as *Forced Certainty Equivalence* control, see [11]. In this case, at each design phase the unknown value $\bar{\boldsymbol{\theta}}$ of the model parameters is replaced by its current estimate $\hat{\boldsymbol{\theta}}^k$, and the experiment is designed optimally for this estimate. However, this policy has two drawbacks: (i) it does not take into account uncertainty on the value of $\boldsymbol{\theta}$, (ii) it is passive, that is, each design phase is considered as the last one to be performed. In Section 2, we express optimal sequential design as a stochastic-control problem. The optimal strategy corresponds to closed-loop control, see [1], and is extremely difficult to solve, even in very simple situations, see [14, 10, 4]. A suboptimal active strategy is presented in Section 3. It relies on a normal approximation of the future posterior distribution of $\boldsymbol{\theta}$, this approximation being independent of future observations. The corresponding design criterion involves several mathematical expectations. Laplace method is used in Section 4 to approximate these expectations. In Section 5, we suggest to use a stochastic approximation algorithm to optimize the design criterion without having to compute posterior expectations. Some comparative results between different strategies are given in Section 6. Finally, Section 7 concludes.

2. Optimal sequential design and stochastic control

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

$$y_k = \eta(\bar{\boldsymbol{\theta}}, \xi^k) + \epsilon_k,$$

where $\bar{\boldsymbol{\theta}} \in \Theta$ is the true value of the model parameters, with $\Theta \subset \mathbb{R}^p$ a compact set, ξ^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 $\eta(\boldsymbol{\theta}, \xi)$ is the model response for the value $\boldsymbol{\theta}$ of the model parameters and experimental conditions ξ. The response is assumed to be nonlinear in $\boldsymbol{\theta}$ and twice continuously differentiable with respect to $\boldsymbol{\theta}$.

We consider experimental design for parameter estimation and wish to choose experimental conditions $\boldsymbol{\xi}_1^N = \{\xi^1, \ldots, \xi^N\}$, with $N \geq p$ fixed, that maximize a function $\Phi(\cdot)$ (usually concave) of the Fisher information matrix

$$\mathbf{M}\left(\boldsymbol{\theta}, \boldsymbol{\xi}_1^N\right) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{\partial \eta\left(\boldsymbol{\theta}, \xi^i\right)}{\partial \boldsymbol{\theta}} \frac{\partial \eta\left(\boldsymbol{\theta}, \xi^i\right)}{\partial \boldsymbol{\theta}^T} .$$

For instance, $\Phi(\mathbf{M}) = \log \det[\mathbf{M}]$, or $\Phi(\mathbf{M}) = -\mathrm{trace}[\mathbf{M}^{-1}]$ are classical choices. We adopt a Bayesian approach (for design, not for estimation), and wish to maximize $\mathrm{E}\{\Phi[\mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^N)] | \mathcal{I}^0\}$ with respect to $\boldsymbol{\xi}_1^N$, with $\mathcal{I}^0$ the prior information (prior distribution $\pi(\cdot)$ of $\boldsymbol{\theta}$ and distribution $\pi_\epsilon(\cdot)$ of $\{\epsilon_k\}$). Note that ξ^k may depend on $\mathcal{I}^{k-1} = \{\mathcal{I}^0, \boldsymbol{\xi}_1^{k-1}, \mathbf{y}_1^{k-1}\}$, which corresponds to the information available at step k, with $\mathbf{y}_1^k = \{y_1, \ldots, y_k\}$. The expectation $\mathrm{E}\{\cdot\}$ is with respect to all random variables involved, that is, $\boldsymbol{\theta}$ and $\{\epsilon_k\}$. The optimal strategy at step j, when $\mathcal{I}^{j-1}$ is available, is then solution of the following stochastic dynamic-programming problem:

$$\begin{aligned} (1) \qquad & \max_{\xi^j} [\ \mathrm{E}_{y_j} \{ \max_{\xi^{j+1}} [\ \mathrm{E}_{y_{j+1}} \{ \ldots \max_{\xi^{N-1}} [\ \mathrm{E}_{y_{N-1}} \{ \max_{\xi^N} [\ \mathrm{E}_{\boldsymbol{\theta}} \{ \Phi[\mathbf{M}(\boldsymbol{\theta}, \xi_1^N)] \\ & |\mathcal{I}^{N-1}\}] | \mathcal{I}^{N-2}\}] \ldots | \mathcal{I}^j\}] | \mathcal{I}^{j-1}\}] \end{aligned}$$

Once the solution is found, we apply ξ_j, increment j by 1 and repeat. However, the problem (1) can be solved analytically in very special situations only, and its numerical solution is feasible only for N very small ($N = 2$ in [14, 10, 4]). Suboptimal solutions must therefore be considered. They usually rely on an approximation of future posterior densities $\pi(\boldsymbol{\theta} | \mathcal{I}^k)$.

FORCED CERTAINTY EQUIVALENCE (*FCE*) control: it corresponds to *sequential local design*, and amounts to substituting the Dirac function $\delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{j-1})$ for $\pi(\boldsymbol{\theta} | \mathcal{I}^{N-1})$ in (1), with $\boldsymbol{\theta}^{j-1}$ some estimator of $\boldsymbol{\theta}$, e.g.

$$(2) \qquad \tilde{\boldsymbol{\theta}}^{j-1} = \arg\max_{\boldsymbol{\theta}} \pi\left(\boldsymbol{\theta} | \mathcal{I}^{j-1}\right) .$$

We then maximize $\Phi[\mathbf{M}(\tilde{\boldsymbol{\theta}}^{j-1}, [\boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N])]$ with respect to $\boldsymbol{\xi}_j^N$, with $\boldsymbol{\xi}_1^{j-1}$ fixed, apply ξ_j, and repeat.

OPEN-LOOP FEEDBACK (*OLF*) control: it corresponds to *average-optimal design*, and amounts to substituting the current posterior density $\pi(\boldsymbol{\theta}|\mathcal{I}^{j-1})$ for $\pi(\boldsymbol{\theta}|\mathcal{I}^{N-1})$ in (1). We then maximize $\hat{\mathrm{E}}^{j}_{\boldsymbol{\theta}}\{\Phi[\mathbf{M}(\boldsymbol{\theta},[\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N])]\}$ with respect to $\boldsymbol{\xi}_j^N$, where $\boldsymbol{\xi}_1^{j-1}$ is fixed and $\hat{\mathrm{E}}^{j}_{\boldsymbol{\theta}}\{\cdot\}$ denotes the expectation for $\pi(\boldsymbol{\theta}|\mathcal{I}^{j-1})$, apply ξ^j, and repeat. When $\pi(\cdot)$ is normal $\mathcal{N}(\hat{\boldsymbol{\theta}}^0,\boldsymbol{\Sigma}_0)$ and $\pi_\epsilon(\cdot)$ is normal $\mathcal{N}(0,\sigma^2)$, $\pi(\boldsymbol{\theta}|\mathcal{I}^{j-1})$ can be approximated by $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1},\boldsymbol{\Sigma}_{j-1}(\tilde{\boldsymbol{\theta}}^{j-1}))$,

$$\boldsymbol{\Sigma}_k(\boldsymbol{\theta})=\left[\boldsymbol{\Sigma}_0^{-1}+\mathbf{M}\left(\boldsymbol{\theta},\boldsymbol{\xi}_1^k\right)\right]^{-1}. \tag{3}$$

Note that *FCE* is independent of σ^2 and $\boldsymbol{\Sigma}_0$. *OLF* with the approximation above depends on $\boldsymbol{\Sigma}_0/\sigma^2$. In *FCE* and *OLF*, all support points in $\boldsymbol{\xi}_j^N$ at step j play the same role. Therefore, the sequential character of the experiment is not taken into account. These strategies use feedback but are *passively adaptive*: decisions are taken as if no other observation would take place in the future. On the other hand, *active* control aims at taking into account the influence of present action on future uncertainty. A sequential design strategy based on active control is presented in the next section.

3. ACTIVE SEQUENTIAL DESING

First notice that the terminal cost $\Phi[\mathbf{M}(\boldsymbol{\theta},\boldsymbol{\xi}_1^N)]$ can usually be decomposed into a sum of terms. For instance, when $\Phi(\cdot)=-\mathrm{trace}[(\cdot)^{-1}]$, we can write

$$-\mathrm{trace}\left[\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^N)\right]=$$

$$=-\mathrm{trace}\left[\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\right]+\frac{\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}{\sigma^2+\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}$$

$$+\ldots+\frac{\mathbf{r}^T(\xi_N)\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}{\sigma^2+\mathbf{r}^T(\xi_N)\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}$$

where $\mathbf{r}(\xi_j)=\frac{\partial\eta(\boldsymbol{\theta},\xi_j)}{\partial\boldsymbol{\theta}}$. The dynamic-programming formulation of the design problem then becomes:

$$\begin{aligned}
&\max_{\xi^1}\Bigg[\mathrm{E}_{y_1}\Bigg\{\max_{\xi^2}\Bigg[\mathrm{E}_{y_2}\Bigg\{\dots\max_{\xi^p}\Bigg[\mathrm{E}_{y_p}\Bigg\{\ \mathrm{E}_{\boldsymbol{\theta}}\Big\{-\mathrm{trace}\,\Big[\mathbf{M}^{-1}(\boldsymbol{\theta},\xi_1^p)\Big]\,|\mathcal{I}^p\Big\}+\\
&\max_{\xi^{p+1}}\Bigg[\ \mathrm{E}_{y_{p+1}}\Bigg\{\ \mathrm{E}_{\boldsymbol{\theta}}\Bigg\{\frac{\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}{\sigma^2+\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}|\mathcal{I}^{p+1}\Bigg\}+\dots+\\
&\max_{\xi^{N-1}}\Bigg[\mathrm{E}_{y_{N-1}}\Bigg\{\ \mathrm{E}_{\boldsymbol{\theta}}\Bigg\{\frac{\mathbf{r}^T(\xi_{N-1})\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-2})\mathbf{r}(\xi_{N-1})}{\sigma^2+\mathbf{r}^T(\xi_{N-1})\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-2})\mathbf{r}(\xi_{N+1})}|\mathcal{I}^{N-1}\Bigg\}+\\
&\max_{\xi^N}\Bigg[\mathrm{E}_{\boldsymbol{\theta}}\Bigg\{\frac{\mathbf{r}^T(\xi_N)\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}{\sigma^2+\mathbf{r}^T(\xi_N)\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}|\mathcal{I}^{N-1}\Bigg\}\Bigg]\ |\mathcal{I}^{N-2}\Bigg\}\Bigg]\\
&\dots\quad|\mathcal{I}^p\}]\,|\mathcal{I}^{p-1}\Big\}\Big]\dots|\mathcal{I}^1\Big\}\Big]\,|\mathcal{I}^0\Big\}\Big]
\end{aligned}\tag{4}$$

Approximating at step j the future posterior densities $\pi(\boldsymbol{\theta}|\mathcal{I}^k)$, $k \geq j$, by $\delta(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}^{j-1})$ (resp. $\pi(\boldsymbol{\theta}|\mathcal{I}^{j-1})$), one gets again the passive *FCE* (resp. *OLF*) strategies. To make the strategy *active*, we shall approximate $\pi(\boldsymbol{\theta}|\mathcal{I}^k)$ by the density of the normal distribution $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1},\boldsymbol{\Sigma}_k(\tilde{\boldsymbol{\theta}}^{j-1}))$ with $\boldsymbol{\Sigma}_k(\boldsymbol{\theta})$ given by (3). This density depends on $\xi^j,\xi^{j+1},\dots,\xi^k$, which makes it active, but not on $y^j,\dots,y^k$, which makes the computations much easier than in (4). Since

$$\mathrm{E}_{y_k}\Big\{\ \mathrm{E}_{\boldsymbol{\theta}}\Big\{f(\boldsymbol{\theta},\mathcal{I}^{k-1})|\mathcal{I}^k\Big\}\,|\mathcal{I}^{k-1}\Big\} = \ \mathrm{E}_{\boldsymbol{\theta}}\Big\{f(\boldsymbol{\theta},\mathcal{I}^{k-1})|\mathcal{I}^{k-1}\Big\}\ ,$$

we simply need to maximize

$$\begin{aligned}
&\hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,p-1}\Big\{-\mathrm{trace}\,\Big[\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\Big]\Big\}+\ \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,p}\Bigg\{\frac{\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}{\sigma^2+\mathbf{r}^T(\xi_{p+1})\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^p)\mathbf{r}(\xi_{p+1})}\Bigg\}\\
&+\dots+\ \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,N-2}\Bigg\{\frac{\mathbf{r}^T(\xi_{N-1})\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-2})\mathbf{r}(\xi_{N-1})}{\sigma^2+\mathbf{r}^T(\xi_{N-1})\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-2})\mathbf{r}(\xi_{N+1})}\Bigg\}\\
&+\ \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,N-1}\Bigg\{\frac{\mathbf{r}^T(\xi_N)\mathbf{M}^{-2}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}{\sigma^2+\mathbf{r}^T(\xi_N)\mathbf{M}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}_1^{N-1})\mathbf{r}(\xi_N)}\Bigg\}\ ,
\end{aligned}\tag{5}$$

where $\boldsymbol{\xi}_1^{j-1}$ is fixed and $\hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,k}\{\cdot\}$ denotes the expectation for $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1}, \boldsymbol{\Sigma}_k(\tilde{\boldsymbol{\theta}}^{j-1}))$. Note that the dependence in $\boldsymbol{\Sigma}_0$ and σ^2 is only through $\boldsymbol{\Sigma}_0/\sigma^2$.

The matrix $\mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^p)$ in (5) may be singular if repetitions at the same support points are present in the p first observations. For that reason, we shall consider the criterion $\Phi(\mathbf{M}) = -\mathrm{trace}[(\boldsymbol{\Sigma}^{-1} + \mathbf{M})^{-1}]$. When $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, it correspond to the usual Bayesian A-optimality criterion, see [2, 7]. We use below $\boldsymbol{\Sigma}^{-1} = \lambda \mathbf{I}$ with λ small, which corresponds to a non informative prior. The criterion (5) then becomes

$$
\begin{aligned}
& -\mathrm{trace}\,[\boldsymbol{\Sigma}] + \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,0}\left\{\frac{\mathbf{r}^T(\xi_1)\boldsymbol{\Sigma}^2\mathbf{r}(\xi_1)}{\sigma^2 + \mathbf{r}^T(\xi_1)\boldsymbol{\Sigma}\mathbf{r}(\xi_1)}\right\} \\
& + \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,1}\left\{\frac{\mathbf{r}^T(\xi_2)\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \xi_1)\right]^{-2}\mathbf{r}(\xi_2)}{\sigma^2 + \mathbf{r}^T(\xi_2)\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \xi_1)\right]^{-1}\mathbf{r}(\xi_2)}\right\} \\
& + \cdots + \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,p}\left\{\frac{\mathbf{r}^T(\xi_{p+1})\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^p)\right]^{-2}\mathbf{r}(\xi_{p+1})}{\sigma^2 + \mathbf{r}^T(\xi_{p+1})\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^p)\right]^{-1}\mathbf{r}(\xi_{p+1})}\right\} \\
& + \cdots + \hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,N-1}\left\{\frac{\mathbf{r}^T(\xi_N)\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^{N-1})\right]^{-2}\mathbf{r}(\xi_N)}{\sigma^2 + \mathbf{r}^T(\xi_N)\left[\boldsymbol{\Sigma}^{-1} + \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^{N-1})\right]^{-1}\mathbf{r}(\xi_N)}\right\}
\end{aligned}
\tag{6}
$$

Note that at step j we only have to maximize the sum of the last $N - j$ terms. This idea has been successfully applied to usual control problems, such as regulation, trajectory following, target attainment, for which better performances than *FCE* or *OLF* were obtained, see [5, 8]. In [3] the method is applied to one-dimensional sequential design problems ($\dim \boldsymbol{\theta} = 1$), and its superiority over the passive *FCE* and *OLF* policies is evidenced. However, the design criterion to be optimized requires the computation of several mathematical expectations, which makes the problem cumbersome when $\dim \boldsymbol{\theta} > 1$. For that reason, Laplace method is used in the next section to approximate the corresponding integrals.

4. LAPLACE APPROXIMATION

OLF and the active strategy above require the computation of integrals, which can be approximated in different ways. Consider first the evaluation of $\mathrm{E}_{\boldsymbol{\theta}}\{\Phi[\mathbf{M}(\boldsymbol{\theta},[\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N])]|\mathcal{I}^{j-1}\}$, which is the design criterion at step j for *OLF*. Define

$$g(\boldsymbol{\theta}) = \Phi\left[\mathbf{M}\left(\boldsymbol{\theta},\left[\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N\right]\right)\right],$$

and

$$P(\boldsymbol{\theta}) = -\log\left[\pi\left(\mathbf{y}_1^{j-1}|\boldsymbol{\theta}\right)\pi(\boldsymbol{\theta})\right], \tag{7}$$

where $\pi(\mathbf{y}_1^{j-1}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ are respectively the likelihood of $\mathbf{y}_1^{j-1}$ and the prior density of $\boldsymbol{\theta}$. The criterion to be maximized with respect to $\boldsymbol{\xi}_j^N$ is

$$\mathrm{E}_{\boldsymbol{\theta}}\left\{g(\boldsymbol{\theta})|\mathcal{I}^{j-1}\right\} = \frac{\int g(\boldsymbol{\theta})\exp\left[-P(\boldsymbol{\theta})\right]d\boldsymbol{\theta}}{\int \exp\left[-P(\boldsymbol{\theta})\right]d\boldsymbol{\theta}}. \tag{8}$$

A first approximation is obtained by replacing $g(\boldsymbol{\theta})$ and $P(\boldsymbol{\theta})$ by their second-order development in $\boldsymbol{\theta}$ around $\tilde{\boldsymbol{\theta}}^{j-1}$ given by (2), and then by computing the corresponding two integrals in (8). This gives

$$\mathrm{E}_{\boldsymbol{\theta}}\left\{g(\boldsymbol{\theta})|\mathcal{I}^{j-1}\right\} \approx g\left(\tilde{\boldsymbol{\theta}}^{j-1}\right) + \frac{1}{2}\mathrm{trace}\left[\frac{\partial^2 g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}}\left(\frac{\partial^2 P(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}}\right)^{-1}\right]. \tag{9}$$

Since $\pi(\boldsymbol{\theta})$ and $\pi(\mathbf{y}_1^{j-1}|\boldsymbol{\theta})$ are assumed to be normal, we get $P(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^0)^T \boldsymbol{\Sigma}_0{}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^0) + \frac{1}{2\sigma^2}\sum_{i=1}^{j-1}\left[y_i-\eta(\boldsymbol{\theta},\xi^i)\right]^2 + C$, with C an additive constant, which gives

$$\begin{aligned}\frac{\partial^2 P(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}} = {} & \boldsymbol{\Sigma}_0{}^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^{j-1}\frac{\partial\eta(\boldsymbol{\theta},\xi^j)}{\partial\boldsymbol{\theta}}_{|\tilde{\boldsymbol{\theta}}^{j-1}}\frac{\partial\eta(\boldsymbol{\theta},\xi^j)}{\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}} \\ & + \frac{1}{\sigma^2}\sum_{i=1}^{j-1}\frac{\partial\eta^2(\boldsymbol{\theta},\xi^j)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}}\left(\eta\left(\tilde{\boldsymbol{\theta}}^{j-1},\xi_i\right)-y_i\right).\end{aligned}$$

Note that the second term on the right-hand side is equal to $\mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}_1^{j-1})$, so that neglecting the errors $(\eta(\boldsymbol{\theta}, \xi_i) - y_i)$, one obtains for $\partial^2 P(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ the approximate Hessian used in the Gauss-Newton algorithm for the computation of $\tilde{\boldsymbol{\theta}}^{j-1}$, which also coincides with the inverse of the covariance matrix $\boldsymbol{\Sigma}_{j-1}(\tilde{\boldsymbol{\theta}}^{j-1})$ for the classical normal approximation of the posterior density of $\boldsymbol{\theta}$, see (3). This makes (9) correspond to an L-optimal design criterion.

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

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

with $P(\boldsymbol{\theta})$ still given by (7), and we use Laplace approximation for computing the two integrals in (8): we replace $P(\boldsymbol{\theta})$ and $Q(\boldsymbol{\theta})$ by their second-order development in $\boldsymbol{\theta}$ around their minimum. One gets, see [13, 12],

$$\mathrm{E}_{\boldsymbol{\theta}}\left\{g\left(\boldsymbol{\theta}\right)|\mathcal{I}^{j-1}\right\} \approx g\left(\check{\boldsymbol{\theta}}^{j-1}\right) \frac{\pi\left(\mathbf{y}_1^{j-1}|\check{\boldsymbol{\theta}}^{j-1}\right)\pi\left(\check{\boldsymbol{\theta}}^{j-1}\right)\det^{-\frac{1}{2}}\left(\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\check{\boldsymbol{\theta}}^{j-1}}\right)}{\pi\left(\mathbf{y}_1^{j-1}|\tilde{\boldsymbol{\theta}}^{j-1}\right)\pi\left(\tilde{\boldsymbol{\theta}}^{j-1}\right)\det^{-\frac{1}{2}}\left(\frac{\partial^2 P(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\tilde{\boldsymbol{\theta}}^{j-1}}\right)}, \tag{11}$$

where $\tilde{\boldsymbol{\theta}}^{j-1}$ is given by (2) and

$$\check{\boldsymbol{\theta}}^{j-1} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\theta}} Q(\boldsymbol{\theta}). \tag{12}$$

When the posterior distribution of $\boldsymbol{\theta}$ is approximated by $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1}$, $\boldsymbol{\Sigma}_{j-1}(\tilde{\boldsymbol{\theta}}^{j-1}))$, the Laplace approximation above gives

$$\mathrm{E}_{\boldsymbol{\theta}}\left\{g\left(\boldsymbol{\theta}\right)|\mathcal{I}^{j-1}\right\} \approx g\left(\check{\boldsymbol{\theta}}^{j-1}\right)\exp\left[-P\left(\check{\boldsymbol{\theta}}^{j-1}\right)\right] \frac{\det^{-\frac{1}{2}}\left(\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}_{|\check{\boldsymbol{\theta}}^{j-1}}\right)}{\det^{\frac{1}{2}}\boldsymbol{\Sigma}_{j-1}\left(\tilde{\boldsymbol{\theta}}^{j-1}\right)}, \tag{13}$$

with

$$P\left(\boldsymbol{\theta}\right) = \frac{1}{2}\left(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}^{j-1}\right)^T \boldsymbol{\Sigma}_{j-1}^{-1}\left(\tilde{\boldsymbol{\theta}}^{j-1}\right)\left(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}^{j-1}\right),$$

and $Q(\boldsymbol{\theta})$ and $\check{\boldsymbol{\theta}}^{j-1}$ still given by (10) and (12). Approximation (9) and (11) will be compared in Section 6, Example 1.

In the case of the active strategy, a normal approximation is used for the posterior distribution, and we shall only consider the approximation (13). We have to evaluate expectations of the form $\hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,k}\{g_k(\boldsymbol{\theta})\}$ with $g_k(\boldsymbol{\theta})$ a function of $\boldsymbol{\theta}$, see (6). Laplace approximation then gives

$$\hat{\mathrm{E}}_{\boldsymbol{\theta}}^{j,k}\left\{g_k\left(\boldsymbol{\theta}\right)\right\} \approx g_k\left(\check{\boldsymbol{\theta}}^{j-1}\right)\exp\left[-P\left(\check{\boldsymbol{\theta}}^{j-1}\right)\right]\frac{\det^{-\frac{1}{2}}\left(\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}|\check{\boldsymbol{\theta}}^{j-1}\right)}{\det^{\frac{1}{2}}\boldsymbol{\Sigma}_k\left(\tilde{\boldsymbol{\theta}}^{j-1}\right)},$$

with

$$P\left(\boldsymbol{\theta}\right) = \frac{1}{2}\left(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}^{j-1}\right)^T\boldsymbol{\Sigma}_k^{-1}\left(\tilde{\boldsymbol{\theta}}^{j-1}\right)\left(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}^{j-1}\right).$$

and $Q(\boldsymbol{\theta})$ and $\check{\boldsymbol{\theta}}^{j-1}$ given by (10) and (12).

5. STOCHASTIC APPROXIMATION

Stochastic approximation techniques (see, e.g., [6]) can be used to optimize criteria based on mathematical expectations without requiring the computation (or approximation) of the expectations. Their application to average optimal design is considered, e.g., in [9]. We show in this section how such techniques can be used for *OLF* and active sequential design. Consider first the maximization of $\mathrm{E}_{\boldsymbol{\theta}}\{\Phi[\mathbf{M}(\boldsymbol{\theta},[\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N])]|\mathcal{I}^{j-1}\}$ with respect to $\boldsymbol{\xi}_j^N$ for *OLF*, and define

$$g\left(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N\right) = \Phi\left[\mathbf{M}\left(\boldsymbol{\theta},\left[\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N\right]\right)\right],$$

The criterion to be maximized with respect to $\boldsymbol{\xi}_j^N$ is

$$\begin{aligned} f\left(\boldsymbol{\xi}_j^N\right) &= \mathrm{E}_{\boldsymbol{\theta}}\left\{g\left(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N\right)|\mathcal{I}^{j-1}\right\} \\ &= \frac{\int g\left(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N\right)\pi\left(\mathbf{y}_1^{j-1}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}}{\int\pi\left(\mathbf{y}_1^{j-1}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}}. \end{aligned} \tag{14}$$

Since $\int\pi(\mathbf{y}_1^{j-1}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ in (14) is a normalisation constant independent of $\boldsymbol{\xi}_j^N$, the criterion becomes $\int G(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, with $G(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N) = g(\boldsymbol{\theta},\boldsymbol{\xi}_1^{j-1},\boldsymbol{\xi}_j^N)\pi(\mathbf{y}_1^{j-1}|\boldsymbol{\theta})$.

The simplest version of the algorithm is then as follows

$$\left(\boldsymbol{\xi}_j^N\right)_{k+1} = \left(\boldsymbol{\xi}_j^N\right)_k + a_k \frac{\partial G\left(\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N\right)}{\partial \boldsymbol{\xi}_j^N}_{|\left(\boldsymbol{\xi}_j^N\right)_k}, \tag{15}$$

where $\boldsymbol{\theta}^{(k)}$ is generated according to the distribution $\pi(\boldsymbol{\theta})$ and

$$\frac{\partial G\left(\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N\right)}{\partial \boldsymbol{\xi}_j^N} = \pi\left(\mathbf{y}_1^{j-1}|\boldsymbol{\theta}^{(k)}\right) \frac{\partial g\left(\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N\right)}{\partial \boldsymbol{\xi}_j^N}.$$

The sequence $\{a_k\}$ should satisfy

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

A typical choice is $a_k = A/k$, with A a positive constant. Convergence can be accelerated by taking $a_k = A/k^{\beta}$,with $\beta \in (0,1)$, and averaging the iterates

$$\left(\hat{\boldsymbol{\xi}}_j^N\right)_k = \frac{1}{k} \sum_{i=1}^{k} \left(\boldsymbol{\xi}_j^N\right)_k. \tag{16}$$

When the posterior distribution of $\boldsymbol{\theta}$ is approximated by $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1}$, $\boldsymbol{\Sigma}_{j-1}(\tilde{\boldsymbol{\theta}}^{j-1}))$, one can use directly $G(\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N) = g(\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_1^{j-1}, \boldsymbol{\xi}_j^N)$ in (15) and generate $\boldsymbol{\theta}^{(k)}$ according to $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1}$, $\boldsymbol{\Sigma}_{j-1}(\tilde{\boldsymbol{\theta}}^{j-1}))$.

For the active strategy, the criterion to be maximized is a sum of mathematical expectations, see (6):

$$f\left(\boldsymbol{\xi}_j^N\right) = \sum_{i=j-1}^{N-1} \left(\int g\left(\boldsymbol{\theta}, \boldsymbol{\xi}_1^i, \xi^{i+1}\right) \hat{\pi}(\boldsymbol{\theta}|\mathcal{I}^i) d\boldsymbol{\theta} \right),$$

with $\hat{\pi}(\boldsymbol{\theta}|\mathcal{I}^i)$ the density of the normal distribution $\mathcal{N}(\tilde{\boldsymbol{\theta}}^{j-1}, \boldsymbol{\Sigma}_i(\tilde{\boldsymbol{\theta}}^{j-1}))$. It can be rewritten as

$$f(\boldsymbol{\xi}_j^N) = \sum_{i=j-1}^{N-1} \left(\int g\left(\boldsymbol{\theta}(\mathbf{u}), \boldsymbol{\xi}_1^i, \xi^{i+1}\right) \pi_n(\mathbf{u}) d\mathbf{u} \right), \tag{17}$$

with $\pi_n(\cdot)$ the density of the normal distribution $\mathcal{N}(\mathbf{0},\mathbf{I})$, $\boldsymbol{\theta}_i(\mathbf{u}) = \mathbf{Z}_i\mathbf{u} + \tilde{\boldsymbol{\theta}}^{j-1}$, and $\mathbf{Z}_i$ satisfying $\mathbf{Z}_i{\mathbf{Z}_i}^T = \boldsymbol{\Sigma}_i(\tilde{\boldsymbol{\theta}}^{j-1})$. Note that in the stochastic approximation algorithm for the maximization of $f(\boldsymbol{\xi}_j^N)$, one can generate at each iteration a unique vector $\mathbf{u}^{(k)}$ according to $\mathcal{N}(\mathbf{0},\mathbf{I})$ instead of $N-j$ vectors ${\mathbf{u}_i}^{(k)}$. This algorithm makes the determination of the optimal active strategy as simple as that of the *OLF* strategy.

6. EXAMPLES

Example 1. Consider the model of exponential decay $\eta(\theta,\xi) = \exp(-\theta\xi)$, with $\theta \in \mathbb{R}$ and $\mathrm{E}_\theta\{\mathbf{M}(\theta,\boldsymbol{\xi}_1^N)\} = \mathrm{E}_\theta\{(1/\sigma^2)\sum_{i=1}^N {\xi^i}^2 \exp(-2\theta\xi^i)\}$ the criterion to be maximized. The prior distribution for θ is normal $\mathcal{N}(0.5, 0.01)$. We first compare the approximations given in Section 4 for the *OLF* design criterion. We take $N = 4$ and consider the last step $j = 4 : \xi^4$ is chosen so as to maximize $\overset{3}{\hat{\mathrm{E}}}_\theta\{\mathbf{M}(\theta,\boldsymbol{\xi}_1^4)\}$. The variance of the measurement errors is equal to $\sigma^2 = 5.10^{-3}$.

Figure 1. Criterion $\overset{3}{\hat{\mathrm{E}}}_\theta\{M(\theta,\boldsymbol{\xi}_1^4)\}$ as function of ξ^4. Numericalintegration: full line, normal approximation: dashed line, approximation (9): large dots, approximation (11): small dots.

In Figure 1, the curve in full line corresponds to the exact design criterion obtained by numerical integration, a normal approximation of the posterior distribution of θ is used for the curve in dashed-line (the mathematical expectation can then be calculated analytically). The small dots correspond to the approximation (11), whereas the large dots correspond to (9). One can notice the very good agreement between (11) and the exact criterion (the curves are almost indistinguishable).

The criterion in full line in Figure 1 is now optimized with the stochastic approximation algorithm of Section 5. Figure 2 gives the evolution of $(\xi^4)_k$ given by (15) and $(\hat{\xi}^4)_k$ given by (16), with respectively $a_k = 60/k$ and $a_k = 60/k^{0.7}$. We see that although $\theta^{(k)}$ is generated according to the prior distribution, the optimization of the posterior expectation is reasonably fast.

Figure 2. Evolution of $(\xi^4)_k$ (full line) and $(\hat{\xi}^4)_k$ (dotted line) as functions of k, for the optimization of the *OLF* criterion in Example 1 (full line in Figure 1).

Consider finally the active strategy of Section 3. We take $\sigma^2 = 0.015$, $N = 2$ and consider step 1. The active strategy can then be computed without numerical integration, see [3], and $\boldsymbol{\xi}_1^2 \simeq (2.20, 2.13)$. Figure 3 (15) and $(\hat{\boldsymbol{\xi}}_1^2)_k$ given by (16), with respectively $a_k = 8/k$ and $a_k = 8/k^{0.7}$, when the criterion (17) is optimized. Again, convergence to the optimal values is reasonably fast. ■

To enhance the active character of the strategy of Section 3, we introduce constraints between successive design points, and consider the case where the experimental variable ξ corresponds to time. The sequential construction of the design then implies $\xi^{k+1} \geq \xi^k$ for any k. Since all design points play the same role in passive strategies, they can be permuted without changing the value of the criterion. The time constraint has thus no other effect than restricting the design space to $[\xi^{j-1}, \infty)$ at step j. On the other hand, the active strategy is affected by the time constraint since the successive ξ^i's play different roles in the criterion.

Figure 3. Evolution of $(\boldsymbol{\xi}_1^2)_k$ (full line) and $(\hat{\boldsymbol{\xi}}_1^2)_k$ (dotted line) as functions of k, for the optimization of (17) in Example 1.

Example 2. Consider the nonlinear regression model $\eta(\boldsymbol{\theta}, \xi) = \theta_1 \exp(-\theta_2 \xi)$, with a normal prior $\mathcal{N}(\hat{\boldsymbol{\theta}}^0, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^0 = (1, 0.5)^T$, $\boldsymbol{\Sigma}_0 = \mathrm{diag}(0.04, 0.01)$, $\sigma^2 = 5.10^{-3}$ and $N = 4$. The design criterion is $\Phi(\mathbf{M}) = -\mathrm{trace}[\mathbf{M}^{-1}]$ for *FCE* and *OLF*, $-\mathrm{trace}[(\lambda \mathbf{I} + \mathbf{M})^{-1}]$ for the active strategy, with $\lambda = 0.1$.

We compare the three strategies *FCE* (1), *OLF* (2) and *Active* (3) in terms of the final mean-squared error of estimation, and perform M independent repetitions ($M = 600$) of the sequential experiment. For each experiment, say the i-th one, a value $\bar{\boldsymbol{\theta}}^i$ is generated according to the prior distribution $\pi(\cdot)$ and N independent disturbances $\{\epsilon_k\}^i_{k=1,\ldots,N}$ are generated with the distribution $\mathcal{N}(0,\sigma^2)$. Let $(\hat{\boldsymbol{\theta}}^N)^i_m$ denote the LS estimator of the parameters at the end of the i-th experiment for strategy m, $m = 1,2,3$. We compute the differences of performances between strategies m and n in the i-th experiment ($i = 1,\ldots,M$) as

$$\triangle^i_{m-n} = \|\bar{\boldsymbol{\theta}}^i - (\hat{\boldsymbol{\theta}}^N)^i_m\|^2 - \|\bar{\boldsymbol{\theta}}^i - (\hat{\boldsymbol{\theta}}^N)^i_n\|^2.$$

We thus obtain M independent realisations of $\triangle^i_{m-n}$, and compute the empirical mean $\bar{\triangle}_{m-n}$ and variance V_{m-n} of $\triangle^i_{m-n}$. For large M, $\bar{\triangle}_{m-n}$ tends to be normally distributed, and one can use the t-test to test if $\mathrm{E}\{\triangle^i_{m-n}\} > 0$. We then say that strategy m performs significantly better (resp. worse) than n at level α if $\rho_{m-n} < -t_{M-1,\alpha}$ (resp. $\rho_{m-n} > t_{M-1,\alpha}$), where $t_{M-1,\alpha}$ has probability α of being exceeded by a random variable with Student's t-distribution with $M-1$ degrees of freedom, and

$$\rho_{m-n} = \sqrt{M}\frac{\bar{\triangle}_{m-n}}{\sqrt{V_{m-n}}}. \tag{18}$$

For M large and $\alpha = 0.05$, $t_{M-1,\alpha} \simeq 1.645$.

Table 1 gives the values of the empirical mean-squared error for the three strategies considered, with and without time constraints. Table 2 gives the values of ρ_{m-n}. One can conclude from these results that *OLF* performs significantly better than *FCE* and the active strategy significantly better than *OLF*, especially when time constraints are present. ■

	MSE without constraints	MSE with constraints
FCE	$2.47\ 10^{-2}$	$2.57\ 10^{-2}$
OLF	$2.44\ 10^{-2}$	$2.49\ 10^{-2}$
Active	$2.42\ 10^{-2}$	$2.47\ 10^{-2}$

Table 1. Empirical mean-squared error with and without time constraints, $M = 600$.

	without constraints	with constraints
ρ_{1-2}	2.57	3.39
ρ_{1-3}	2.75	3.71
ρ_{2-3}	2.42	4.38

Table 2. Values of ρ_{m-n} (18) for the three strategies: (1) *FCE*, (2) *OLF*, (3) *Active*, $M = 600$.

Example 3. Consider the nonlinear regression model $\eta(\boldsymbol{\theta}, \xi) = \frac{\theta_1}{\theta_2-\theta_1}$ $(\exp(-\theta_1\xi) - \exp(-\theta_2\xi))$, with a normal prior $\mathcal{N}(\hat{\boldsymbol{\theta}}^0, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^0 = (1, 0.5)^T$, $\boldsymbol{\Sigma}_0 = \text{diag}(0.04, 0.01)$, $N = 4$ and $\sigma^2 = 0.005$. The design criterion is $\Phi(\mathbf{M}) = -\text{trace}[\mathbf{M}^{-1}]$ for *FCE* and *OLF*, $-\text{trace}[(\lambda\mathbf{I} + \mathbf{M})^{-1}]$ for the active strategy, with $\lambda = 1$. Again we impose time constraints, that is $\xi^{k+1} \geq \xi^k$ for any k.

The influence of these constraints on the active strategy is clearly seen at step 1: one obtains $\boldsymbol{\xi}_1^4 = (3.28, 4.31, 0.71, 0.71)$ without time constraints and $\boldsymbol{\xi}_1^4 = (0.67, 0.67, 0.67, 3.40)$ with time constraints. In the second case, since $\xi^k \geq \xi^1$ for $k \geq 2$, ξ^1 must be chosen cautiously so that we get a value smaller than 0.71. For *FCE* and *OLF*, the values are the same with and without constraints: we get respectively $\boldsymbol{\xi}_1^4 = (0.67, 0.67, 0.67, 3.55)$ and $\boldsymbol{\xi}_1^4 = (0.64, 0.64, 0.64, 3.44)$. ■

The designs that are constructed depend on the prior information $\mathcal{I}^0$: *FCE* depends on $\hat{\boldsymbol{\theta}}^0$ and *OLF* and *Active* depend both on $\hat{\boldsymbol{\theta}}^0$ and $\boldsymbol{\Sigma}_0/\sigma^2$. Investigating the robustness of the performance of a given design with respect to a mispecification of the prior would require lengthy simulations, and is beyond the scope of this paper. However, the influence of the prior on the design is clearly seen from examples, as illustated hereafter.

Example 1. (continued): We consider the case $N = 2$ and $\sigma^2 = 0.015$. The optimal strategy (up to a linearization of the response) can be computed, see [3]. Figure 4 gives the evolution of ξ^1 as a function of $\hat{\boldsymbol{\theta}}^0$ for different strategies with the constraint $\xi^2 \geq \xi^1$ when $\boldsymbol{\Sigma}_0 = 0.01$: the dependence in $\hat{\boldsymbol{\theta}}^0$ is similar for all strategies and reflects the fact that ξ^1 should be chosen close to $1/\hat{\boldsymbol{\theta}}^0$.

Figure 4. ξ^1 as a function of $\hat{\boldsymbol{\theta}}^0$ for different strategies with the constraint $\xi^2 \geq \xi^1$ ($N = 2$, $\boldsymbol{\Sigma}_0 = 0.01$) *FCE* : dotted line, *OLF* : stars, *Active* : crosses, optimal strategy : full line.

Figure 5 gives the evolution of ξ^1 as a function of $\boldsymbol{\Sigma}_0$ for different strategies with the constraint $\xi^2 \geq \xi^1$ when $\hat{\boldsymbol{\theta}}^0 = 0.4$. ξ^1 increases with $\boldsymbol{\Sigma}_0$ for *OLF* and *Active*, and *Active* remains closer to the optimal strategy than *OLF*. ■

7. CONCLUSIONS

Various methods have been considered for sequential experiment design in nonlinear regression. Passive strategies ignore the future, whereas active strategies take into account the fact that future observations will take place. The active strategy proposed relies on a normal approximation of the posterior distribution of $\boldsymbol{\theta}$ which is independent of future observations. Expectations with respect to this approximated posterior are required, and Laplace approximation has been suggested to reduce the amout of computations. Numerical simulations have shown that the active strategy performs significantly better than classical passive strategies. Stochastic approximation algorithms seem particularly attractive for the determination of optimal experiments in this context. Indeed, they permit the optimization of criteria given by mathematical expectations without requiring the computation, or the approximation, of the expectations. Numerical examples have shown that such methods are quite effective for the determination of active strategies, which makes it possible to consider *active sequential experiment* design for more complicated regression models met in pratical situations.

Figure 5. ξ^1 as a function of $\boldsymbol{\Sigma}_0$ for different strategies with the constraint $\xi^2 \geq \xi^1$ ($N = 2$, $\hat{\boldsymbol{\theta}}^0 = 0.4$) *FCE* : dotted line, *OLF* : stars, *Active* : crosses, optimal strategy : full line.

References

[1] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, N. J., 1957.

[2] K. Chaloner and I. Verdinelli, *Bayesian experimental design: a review*, Statistical Science **10** (3) (1995), 273–304.

[3] R. Gautier and L. Pronzato, *Sequential design and active control*, New Developments and Applications in Experimental Design (N. Flournoy, W. F. Rosenberger and W. K. Wong, eds.), IMS Lecture Notes **34** (1998), 138–151.

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

[5] C. Kulcsár, L. Pronzato and E. Walter, *Dual control of linearly parameterized models via prediction of posterior densities*, European J. of Control **2** (1996), 135–143.

[6] H. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer, Heidelberg 1997.

[7] J. Pilz, *Bayesian Estimation and Experimental Design in Linear Regression Models*, vol. 55 Teubner-Texte zur Mathematik, Leipzig, 1983 (also Wiley, New York 1991).

[8] L. Pronzato, C. Kulcsár and E. Walter, *An actively adaptive control policy for linear models*, IEEE Trans. Autom. Cont. **41** (1996), 855–858.

[9] L. Pronzato and E. Walter, *Robust experiment design via stochastic approximation*, Mathematical Biosciences **75** (1985), 103–120.

[10] L. Pronzato, E. Walter and C. Kulcsár, *A dynamical-system approach to sequential design*, pp. 11–24 in: *Model-Oriented Data Analysis III, Proceedings MODA3*, St Petersburg, May 1992 (W. G. Müller, H. P. Wynn and A. A. Zhigljavsky, eds.), Physica Verlag, Heidelberg.

[11] W. J. Runggaldier, *Concepts of optimality in stochastic control*, pp. 101–114 in: *Reliability and Decision* (R. Barlow, *et al.*, ed.), Elsevier, Amsterdam.

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

[13] L. Tierney and J. Kadane, *Accurate approximations for posterior moments and marginal densities*, Journal of the American Statistical Association **81** (393) (1986), 82–86.

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

Received 10 December 1998
Revised 24 April 1999