Chapter 11

SEQUENTIAL CONSTRUCTION OF AN EXPERIMENTAL DESIGN FROM AN I.I.D. SEQUENCE OF EXPERIMENTS WITHOUT REPLACEMENT

Luc Pronzato
Laboratoire I3S, CNRS–UNSA, Sophia-Antipolis, France
pronzato@unice.fr

Abstract We consider a regression problem, with observations $y_k = \eta(\boldsymbol{\theta}, \xi_k) + \epsilon_k$, where $\{\epsilon_k\}$ is an i.i.d. sequence of measurement errors and where the experimental conditions ξ_k form an i.i.d. sequence of random variables, independent of $\{\epsilon_k\}$, which are observed sequentially. The length of the sequence $\{\xi_k\}$ is N but only $n < N$ experiments can be performed. As soon as a new experiment ξ_k is available, one must decide whether to perform it or not. The problem is to choose the n values $\xi_{k_1}, \ldots, \xi_{k_n}$ at which observations $y_{k_1}, \ldots, y_{k_n}$ will be made in order to estimate the parameters $\boldsymbol{\theta}$. An optimal rule for the on–line selection of $\xi_{k_1}, \ldots, \xi_{k_n}$ is easily determined when $p = \dim \boldsymbol{\theta} = 1$. A suboptimal open–loop feedback–optimal rule is suggested in Pronzato (1999b) for the case $p > 1$. We propose here a different suboptimal solution, based on a one–step–ahead optimal approach. A simple procedure, derived from an adaptive rule which is asymptotically optimal, Pronzato (1999a), when $p = 1$ ($N \to \infty$, n fixed), is presented. The performances of these different strategies are compared on a simple example.

Keywords: Sequential design, random experiments, expected determinant

Introduction

Consider a regression model, with observations

$$y_k = \eta(\bar{\boldsymbol{\theta}}, \xi_k) + \epsilon_k \,, \tag{11.1}$$

where $\{\epsilon_k\}$ is an i.i.d. sequence of measurement errors, assumed for simplicity to be distributed $\mathcal{N}(0, \sigma^2)$, and $\bar{\boldsymbol{\theta}} \in \Theta$ is the unknown true value of the model parameters to be estimated, with Θ an open subset of $\mathbb{R}^p$. The function $\eta(\boldsymbol{\theta}, \xi)$

A. Atkinson et al. (eds.), Optimum Design 2000, 113–122.

is assumed continuously differentiable in $\boldsymbol{\theta}$, uniformly in ξ. The experimental conditions $\xi_k \in \mathcal{X}$ form an i.i.d. sequence, independent of $\{\epsilon_k\}$, of length N. Only $n < N$ experiments can be made. As soon as a new experiment ξ_k becomes available, one must decide whether to perform it or not, and one wishes to select n values $\xi_{k_1}, \ldots, \xi_{k_n}$ at best.

Let ν denote the probability measure of the ξ_k's, assumed to be known, and let $\{u_k\}$ denote the decision sequence: $u_k = 1$ if y_k is observed, with experimental conditions ξ_k, and $u_k = 0$ otherwise. We denote the random sequence $\xi_1, \ldots, \xi_N$ by $\boldsymbol{\xi}_1^N$ and the decision sequence $u_1, \ldots, u_N$ by $\boldsymbol{u}_1^N$, with, for any admissible policy,

$$u_j \in \mathcal{U}_j \subseteq \{0,1\}\,,\ j = 1, \ldots, N\,,\ \sum_{j=1}^{N} u_j = n\,. \tag{11.2}$$

We consider design criteria $\Phi(\cdot)$ for the estimation of $\boldsymbol{\theta}$ that are increasing functions of the Fisher information matrix, evaluated at a prior value $\hat{\boldsymbol{\theta}}^0$ for $\boldsymbol{\theta}$ if $\eta(\boldsymbol{\theta}, \xi)$ is nonlinear in $\boldsymbol{\theta}$ (local design). We define the rank–one matrix $\mathcal{M}(\xi)$ as

$$\mathcal{M}(\xi) = \frac{1}{\sigma^2} \frac{\partial \eta(\boldsymbol{\theta}, \xi)}{\partial \boldsymbol{\theta}} {}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^0} \frac{\partial \eta(\boldsymbol{\theta}, \xi)}{\partial \boldsymbol{\theta}^\top} {}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^0}$$

and we write $J(\boldsymbol{u}_1^N, \boldsymbol{\xi}_1^N) = \Phi[\sum_{k=1}^N u_k \mathcal{M}(\xi_k)]$. The measure ν for ξ induces a measure μ for $\mathcal{M}(\xi)$; we shall denote $\mathcal{M}_k = \mathcal{M}(\xi_k)$, $\mathbf{M}_j = \sum_{k=1}^j u_k \mathcal{M}_k$ and we assume that $\mathbf{M} = E\{\mathcal{M}_k\}$ and $E\{\Phi(\mathbf{M}_N)\}$ exist. Note that ξ_k, and thus $\mathcal{M}_k$, is known when u_k is chosen. The problem is then to maximise

$$EJ(\boldsymbol{u}_1^N) = E\{\Phi(\mathbf{M}_N)\} = E\{\Phi(\sum_{k=1}^{N} u_k \mathcal{M}_k)\} \tag{11.3}$$

w.r.t. $\boldsymbol{u}_1^N$ satisfying (11.2), where the expectation $E\{\cdot\}$ is w.r.t. the product measure $\mu^{\otimes N}$ of $\mathcal{M}_1, \ldots, \mathcal{M}_N$. For any sequence $\boldsymbol{u}_1^N$ and any step j, $1 \leq j \leq N$, a_j will denote the number of observations already made; that is,

$$a_j = \sum_{k=1}^{j-1} u_k\,, \tag{11.4}$$

with $a_1 = 0$. The problem corresponds to a discrete–time stochastic control problem, where j represents time, $\mathcal{S}_j = (a_j, \mathbf{M}_{j-1}, \mathcal{M}_j)$ and $u_j \in \mathcal{U}_j \subseteq \{0,1\}$ respectively represent the state and control at time j. For each $j \in \{1, \ldots, N\}$ and each policy $\boldsymbol{u}_j^N$, the optimal decisions are obtained by solving the following

problem:

$$\max_{u_j \in \mathcal{U}_j} [E_{\mathcal{M}_{j+1}} \{ \max_{u_{j+1} \in \mathcal{U}_{j+1}} [E_{\mathcal{M}_{j+2}} \{ \max_{u_{j+2} \in \mathcal{U}_{j+1}} [\dots$$

$$E_{\mathcal{M}_{N-1}} \{ \max_{u_{N-1} \in \mathcal{U}_{N-1}} [E_{\mathcal{M}_N} \{ \max_{u_N \in \mathcal{U}_N} [\Phi(\sum_{k=1}^{N} u_k \mathcal{M}_k)]\}]\} \dots]\}]\}] \quad (11.5)$$

where $E_{\mathcal{M}_k}\{.\}$ denotes the expectation w.r.t. $\mathcal{M}_k$, distributed with the measure μ, and, for any $k = j, \dots, N$,

$$\mathcal{U}_k = \mathcal{U}_k(a_k) = \begin{cases} \{0\} & \text{if } a_k = n\,, \\ \{1\} & \text{if } a_k + N - k + 1 \leq n\,, \\ \{0,1\} & \text{otherwise.} \end{cases} \quad (11.6)$$

One can notice that the formulation of the problem would be much more complicated for a criterion function evaluated at the expected value of $\mathbf{M}_N$ or $\mathbf{M}_N^{-1}$, that is, when the problem is to maximize $\Phi(E\{\mathbf{M}_N\})$, or to minimize $\Phi(E\{\mathbf{M}_N^{-1}\})$, hence the choice (11.3).

The case $p = \dim \boldsymbol{\theta} = 1$ is considered in Section 1 The optimal (closed–loop) solution is given by a backward recurrence equation. A simple open–loop solution, asymptotically optimal for $N \to \infty$ with n fixed, is presented. Section 2 concerns the multidimensional case $p > 1$, for which the optimal solution cannot be obtained in close form. Several suboptimal solutions (open–loop feedback–optimal, one–step–ahead optimal) are proposed. Finally, the different strategies are compared on an illustrative example in Section 3

1. ESTIMATION OF A SCALAR PARAMETER

When $p = 1$, $\mathcal{M}$ is scalar, and we simply take $J(\boldsymbol{u}_1^N, \boldsymbol{\xi}_1^N) = \sum_{k=1}^N u_k \mathcal{M}_k$, that is, we maximize the expected information (note that the expected variance is not additive, which makes its minimisation more difficult). The problem to be solved at step j then becomes

$$\max_{u_j \in \mathcal{U}_j} [u_j \mathcal{M}_j + E_{\mathcal{M}_{j+1}} \{ \max_{u_{j+1} \in \mathcal{U}_{j+1}} [u_{j+1} \mathcal{M}_{j+1} + \cdots$$

$$E_{\mathcal{M}_{N-1}} \{ \max_{u_{N-1} \in \mathcal{U}_{N-1}} [u_{N-1} \mathcal{M}_{N-1} + E_{\mathcal{M}_N} \{ \max_{u_N \in \mathcal{U}_N} u_N \mathcal{M}_N \}]\} \cdots]\} \quad (11.7)$$

with the constraints (11.6) on the sets $\mathcal{U}_k$. It presents some similarities with the well-known secretary problem, see Pronzato (1999a) for a discussion. The state $\mathcal{S}_j$ at step j reduces to $\mathcal{S}_j = (a_j, \mathcal{M}_j)$, with a_j given by (11.4). Let $C(j, a_j, \mathcal{M}_j, \boldsymbol{u}_j^N)$ denote the optimal conditional expected gain–to–go at step j given $\mathcal{S}_j$, $C(j, a_j, \mathcal{M}_j, \boldsymbol{u}_j^N) = E\{\sum_{k=j}^N u_k \mathcal{M}_k | a_j, \mathcal{M}_j\}$, let $\tilde{C}(j, a_j, \mathcal{M}_j)$ denote its value when $\boldsymbol{u}_j^N$ is chosen optimally, and $c(j, a_j, \boldsymbol{u}_j^N)$, $\tilde{c}(j, a_j)$ respectively denote the expected values of $C(j, a_j, \mathcal{M}_j, \boldsymbol{u}_j^N)$ and $\tilde{C}(j, a_j, \mathcal{M}_j)$ w.r.t.

$\mathcal{M}_j$. One has, for $n-N+j-1 < a_j < n$, $\tilde{C}(j, a_j, \mathcal{M}_j) = \max_{u_j \in \{0,1\}}[u_j \mathcal{M}_j + \tilde{c}(j+1, a_j + u_j)]$. The optimal decision is thus

$$\tilde{u}_j(a_j, \mathcal{M}_j) = \begin{cases} 0 & \text{if } \mathcal{M}_j \leq \tilde{s}(j, a_j) = \tilde{c}(j+1, a_j) - \tilde{c}(j+1, a_j+1)\,, \\ 1 & \text{otherwise,} \end{cases} \tag{11.8}$$

which gives $\tilde{C}(j, a_j, \mathcal{M}_j) = \max[\mathcal{M}_j + \tilde{c}(j+1, a_j+1), \tilde{c}(j+1, a_j)]$ and the following backward recurrence equation for $\tilde{c}(j, a_j)$:

$$\tilde{c}(j, a_j) = E\{\max[\mathcal{M} + \tilde{c}(j+1, a_j+1), \tilde{c}(j+1, a_j)]\}\,. \tag{11.9}$$

The constraints (11.2) give $\tilde{c}(j, j+n-N-1) = (N-j+1)E\{\mathcal{M}\}$ and $\tilde{C}(j, n, \mathcal{M}_j) = 0$ for all $\mathcal{M}_j$, which initialises the recurrence (11.9) used to compute the optimal thresholds $\tilde{s}(j, a_j)$ in (11.8). The asymptotic properties ($N \to \infty$, n fixed) of this optimal solution are considered in Pronzato (1999a). In particular, it is shown that when the support of μ is unbounded, for any $a_j < n$, $\tilde{s}(j, a_j)$ and $\tilde{c}(j, a_j)$ tend to infinity as $N \to \infty$, and, when the tail of the distribution function of $\mathcal{M}$ decreases fast enough, a simple open–loop rule is asymptotically optimal.

Assume that μ is absolutely continuous w.r.t. the Lebesgue measure, with $\varphi(\cdot)$ its density, let $F(\cdot)$ denote the distribution function of $\mathcal{M}$, and consider the following assumption on the tail of $F(\cdot)$.

H1: $\bar{F}(\cdot) = 1 - F(\cdot)$ is twice differentiable, the density $\varphi(\cdot)$ is such that $\varphi(s) > 0$ and its derivative $\varphi'(\cdot)$ satisfies $\varphi'(s) < 0$ for s larger than some s_1. Moreover, $\bar{F}(\cdot)$ has the representation

$$\bar{F}(s) = \bar{F}(s_0) \exp\left[-\int_{s_0}^{s} \frac{1}{a(t)} dt\right]\,, \quad s \geq s_0\,,$$

where the auxiliary function $a(t) > 0$ is absolutely continuous w.r.t. Lebesgue measure, with derivative $a'(t)$ having limit $\lim_{t\to\infty} a'(t) = \bar{a} \in [0, 1)$.

Note that $a(s) = \bar{F}(s)/\varphi(s)$ for $s > s_0$. When $\bar{a} = 0$, $F(\cdot)$ is a von Mises function, see Embrechts *et al.* (1997), p. 138, a class which contains for instance the exponential, normal, lognormal, Weibull and Gamma distributions, all with a tail decreasing faster than any power law $s^{-\alpha}$. In that case, $\lim_{t\to\infty} \bar{F}(t)\varphi'(t)/[\varphi(t)]^2 = -1$ and $\lim_{t\to\infty} a(t)/t = 0$, see Embrechts *et al.* (1997), p. 140. When $a(t) = t/b(t)$ with $b(t) \to \alpha \in (1, \infty)$ as $t \to \infty$, $\bar{a} = 1/\alpha$ and $\bar{F}(\cdot)$ is regularly varying with index $-\alpha$; that is (see Embrechts *et al.* (1997), p. 566),

$$\bar{F}(\cdot) \in \mathcal{R}_{-\alpha} : \lim_{s\to\infty} \frac{\bar{F}(ts)}{\bar{F}(s)} = t^{-\alpha}\,, \; t > 0\,.$$

We define

$$A(s) = \frac{h(s)\varphi(s)}{[\bar{F}(s)]^2}\,, \quad \bar{A} = \lim_{s\to\infty} A(s)\,,$$

where $h(s) = E\{\max(\mathcal{M} - s, 0)\}$. When $F(\cdot)$ is a von Mises function, direct application of L'Hôpital's rule shows that $\bar{A} = 1$. When $\bar{F}(\cdot) \in \mathcal{R}_{-\alpha}$, $\alpha \in (1, \infty)$, $\varphi(\cdot) \in \mathcal{R}_{-(\alpha+1)}$ and from Feller (1966), vol. 2, p. 281, $\bar{A} = \alpha/(\alpha - 1)$.

The following property is proved in Pronzato (1999a).

Theorem 1. *Assume that $F(\cdot)$ satisfies* **H1**. *Any admissible open–loop decision rule $\boldsymbol{u}_1^N$, defined by $u_j(a_j, \mathcal{M}_j) = 0$ if $\mathcal{M}_j \le s(j, a_j)$, $u_j(a_j, \mathcal{M}_j) = 1$ otherwise, with thresholds $s(j, a_j)$ such that $(N - j)\bar{A}\bar{F}[s(j, a_j)] \to n - a_j$ when $N \to \infty$, is asymptotically optimal in the following sense:*
when $n - a_j = 1$,

(i) $c(j, n-1, \boldsymbol{u}_j^N)/\tilde{c}(j, n-1) \to 1$ as $N \to \infty$ if $\liminf_{s\to\infty} a(s) > c > 0$;

(ii) $\tilde{c}(j, n-1) - c(j, n-1, \boldsymbol{u}_j^N) \to 0$ as $N \to \infty$ if $\limsup_{s\to\infty} a(s) < C < \infty$;
when $n - a_j > 1$,

(iii) $c(j, a_j, \boldsymbol{u}_j^N)/\tilde{c}(j, a_j) \to 1$ as $N \to \infty$ if $\bar{a} = 0$ and $\liminf_{s\to\infty} a(s) > c > 0$;

(iv) $\tilde{c}(j, a_j) - c(j, a_j, \boldsymbol{u}_j^N) \to 0$ as $N \to \infty$ if $\bar{a} = 0$ and $a(s) \to 0$ as $s \to \infty$.

In particular, the open–loop rule defined by

$$u_j(a_j, \mathcal{M}_j) = \begin{cases} 0 & \text{if } \bar{F}(\mathcal{M}_j) \ge \frac{n - a_j}{\bar{A}(N-j+1) + (1-\bar{A})(n-a_j) - \epsilon}\,, \\ 1 & \text{otherwise,} \end{cases} \tag{11.10}$$

with $0 < \epsilon \ll 1$, satisfies (11.6) and the conditions in Theorem 1, and is thus asymptotically optimal.

When the distribution of the $\mathcal{M}_k$'s is unknown, their empirical distribution can be plugged in to the construction of the optimal rule (11.8) or the open-loop rule (11.10). This (suboptimal) approach corresponds to Certainty Equivalence (CE) control. First, we delay the decision for a few steps in order to initialise the construction of the empirical distribution of the $\mathcal{M}_k$'s, see Pronzato (1999a) for a discussion; then, at each step j, we substitute $\hat{F}_j(\cdot)$ for $F(\cdot)$, with $\hat{F}_j(\cdot)$ the empirical distribution function of $\mathcal{M}$ based on previous observed values, including $\mathcal{M}_j$. In the case of (11.10), this CE open–loop rule can be expressed directly in terms of the order statistics $\{\mathcal{M}^*_{i,j-1}\}$, ($\{\mathcal{M}^*_{i,j-1}\}$ denotes the sequence obtained by ordering the $\mathcal{M}_k$'s, $k \le j - 1$, by decreasing values, with for any j, $\mathcal{M}^*_{i,j-1} = -\infty$ for $i > j - 1$ and $\mathcal{M}^*_{0,j-1} = \infty$). The rule (11.10) then becomes

$$u_j^{CE}(a_j, \mathcal{M}_j) = \begin{cases} 0 & \text{if } \mathcal{M}_j < \mathcal{M}^*_{l_j, j-1}\,, \\ 1 & \text{otherwise,} \end{cases} \tag{11.11}$$

with

$$l_j = \left\lceil \frac{j(n-a_j)}{\bar{A}(N-j+1)+(1-\bar{A})(n-a_j)-\epsilon} \right\rceil , \qquad (11.12)$$

where $\lceil x \rceil$ rounds x to the nearest larger integer.

2. ESTIMATION OF A PARAMETER VECTOR

When $p = \dim(\boldsymbol{\theta}) > 1$, the problem (11.5) cannot be solved analytically, and we restrict our attention to suboptimal solutions.

2.1. OPEN–LOOP FEEDBACK–OPTIMAL DECISIONS

A suboptimal approach, called Open–Loop Feedback–Optimal (OLFO) in control theory, assumes that at each step j, all the decisions $\boldsymbol{u}_j^N$ may only depend on the current state $\mathcal{S}_j$. The OLFO solution is then obtained by solving the optimisation problem:

$$\max_{\boldsymbol{u}_j^N} [E_{\mathcal{M}_{j+1},\ldots,\mathcal{M}_N}\{\Phi(\sum_{k=j}^{N} u_k \mathcal{M}_k)|\mathcal{S}_j\}] ,$$

with $\boldsymbol{u}_j^N$ satisfying the constraints (11.2) and being a function of $\mathcal{S}_j = (a_j, \mathbf{M}_{j-1}, \mathcal{M}_j)$ only. It appears as a deterministic quantity in the conditional expectation above, with $\sum_{k=j}^{N} u_k = \max(0, n-a_j)$. For $n-N+j-1 < a_j < n$, the decision rule is thus:

$$\hat{u}_j(\mathcal{S}_j) = \begin{cases} 0 & \text{if } E_{\mathcal{M}_{j+1},\ldots,\mathcal{M}_{n+j-a_j}}\{\Phi(\mathbf{M}_{j-1}+\mathcal{M}_j+\sum_{k=j+1}^{n+j-a_j}\mathcal{M}_k)\} \leq \\ & E_{\mathcal{M}_{j+1},\ldots,\mathcal{M}_{n+j-a_j+1}}\{\Phi(\mathbf{M}_{j-1}+\sum_{k=j+1}^{n+j-a_j+1}\mathcal{M}_k)\} , \\ 1 & \text{otherwise} . \end{cases} \qquad (11.13)$$

Note that $\hat{u}_j$ is independent of N. It can be easily determined when the criterion takes the form $\Phi(\mathbf{M}) = \det(\boldsymbol{\Omega}^{-1}+\mathbf{M})$, with $\boldsymbol{\Omega}$ a positive–definite matrix. Indeed, expectations of determinants are obtained analytically as follows, see Pronzato (1999b).

Theorem 2. *Assume that the vectors $\mathbf{z}_i \in \mathbb{R}^p$ are i.i.d., with a probability measure μ_z such that $\bar{\mathbf{M}} = E\{\mathbf{z}\mathbf{z}^\top\}$ exists, then for any regular matrix $\mathbf{Q}$,*

$$E_{\mathbf{z}_1,\ldots,\mathbf{z}_k}\{\det[\mathbf{Q}+\sum_{l=1}^{k}\mathbf{z}_l\mathbf{z}_l^T]\} = \det[\mathbf{Q}] \times \left[1+\sum_{l=1}^{\min(k,p)} \mathrm{C}_k^l P_l(\mathbf{t}_1^l)\right] ,$$

where $\mathrm{C}_k^l = k!/[l!(k-l)!]$, $\mathbf{t}_1^l = (t_1,\ldots,t_l)$, with $t_l = tr[(\mathbf{Q}^{-1}\bar{\mathbf{M}})^l]$, and

$$P_l(\mathbf{t}_1^l) = l! \sum_{1^{n_1}\ldots l^{n_l}} \frac{(-1)^{l+\sum_{i=1}^{l} n_i}}{(1^{n_1}\ldots l^{n_l})(n_1!\ldots n_l!)} t_1^{n_1}\ldots t_l^{n_l} ,$$

where the summation $\sum_{1^{n_1}\ldots l^{n_l}}$ *is over all non-negative integers* $n_1, \ldots, n_l$ *such that* $\sum_{i=1}^{l} i n_i = l$, *that is, over all partitions of* l *which have* n_i *parts of size* i.

The proof is by induction on k. One has for $l = 1, \ldots, 4$:

$$P_1(t_1) = t_1\,, \qquad P_2(\mathbf{t}_1^2) = t_1^2 - t_2\,,$$
$$P_3(\mathbf{t}_1^3) = t_1^3 - 3t_1t_2 + 2t_3\,, \quad P_4(\mathbf{t}_1^4) = t_1^4 - 6t_1^2t_2 + 8t_1t_3 + 3t_2^2 - 6t_4\,.$$

The OLFO thresholds in (11.13) are obtained by applying Theorem 2 two times, with $\mathbf{Q} = \mathbf{\Omega}^{-1} + \mathbf{M}_{j-1} + \mathcal{M}_j$ and $\mathbf{Q} = \mathbf{\Omega}^{-1} + \mathbf{M}_{j-1}$ respectively. Note that this approach is not restricted to parameter estimation in regression models (11.1); the only requirement for Theorem 2 to apply is that the information matrix can be written as a sum of i.i.d. rank–one matrices.

2.2. ONE–STEP–AHEAD DECISIONS

If the experimental conditions could be chosen freely in a given set $\mathcal{X}$, at step j a steepest ascent algorithm, see Fedorov (1972); Wynn (1970), would select $\xi_j^* = \arg\max_{\xi_j \in \mathcal{X}} d(\mathbf{M}_{j-1}, \mathcal{M}_j)$, with

$$d(\mathbf{M}_{j-1}, \mathcal{M}_j) = \frac{\partial \Phi[(1-\alpha)\mathbf{M}_{j-1} + \alpha \mathcal{M}_j]}{\partial \alpha}_{|\alpha=0}\,.$$

When $\Phi(\mathbf{M}) = \det(\mathbf{\Omega}^{-1} + \mathbf{M})$, choosing ξ_j^* produces the maximal increase in Φ, since

$$\Phi(\mathbf{M}_{j-1} + \mathcal{M}_j) = \Phi(\mathbf{M}_{j-1})\{1 + \text{trace}[\mathcal{M}_j(\mathbf{\Omega}^{-1} + \mathbf{M}_{j-1})^{-1}]\}$$

and

$$\begin{aligned} d(\mathbf{M}_{j-1}, \mathcal{M}_j) \;=\; & \Phi(\mathbf{M}_{j-1})\{\text{trace}[\mathcal{M}_j(\mathbf{\Omega}^{-1} + \mathbf{M}_{j-1})^{-1}] \\ & -\text{trace}[\mathbf{M}_{j-1}(\mathbf{\Omega}^{-1} + \mathbf{M}_{j-1})^{-1}]\}\,. \end{aligned}$$

Assume that at step j there is only one additional observation to make, that is, $a_j = n - 1$. The state $\mathcal{S}_j$ is known, and the optimal solution of (11.5) is obtained as in Section 1, by solving a problem similar to (11.7) with the i.i.d. scalar variables $\mathcal{M}_k$, $k \geq j$, replaced by $\mathcal{M}'_k = d(\mathbf{M}_{j-1}, \mathcal{M}_k)$. In particular, Theorem 1 applies, and, under the conditions of the theorem, the open–loop rule (11.10) is asymptotically optimal.

When more observations are to be made, we still define $\mathcal{M}'_k = d(\mathbf{M}_{j-1}, \mathcal{M}_k)$ and maximise at step j:

$$EJ'(\boldsymbol{u}_j^N) = E_{\mathcal{M}'_{j+1}, \ldots, \mathcal{M}'_N}\{\sum_{k=j}^{N} u_k \mathcal{M}'_k | \mathcal{S}_j\}\,. \tag{11.14}$$

The optimal solution is given by (11.8), which provides a suboptimal solution for (11.5). The strategy can be further simplified by using (11.10). The OLFO rule can also be used to get a suboptimal solution for the maximisation of (11.14): ξ_j is accepted if $\mathcal{M}'_j > E\{\mathcal{M}'_k|\mathcal{S}_j\}$ and rejected otherwise.

Note that the distribution of $\mathcal{M}'_k$ may be difficult to derive from that of ξ_k, whereas using CE and empirical distributions is straightforward, see in particular (11.11,11.12).

3. EXAMPLE

Consider the following linear regression model, quadratic in the design variable ξ_k : $\eta(\boldsymbol{\theta}, \xi_k) = \theta_0 + \theta_1 \xi_k + \theta_2 \xi_k^2$, with $\Theta = I\!R^3$, $\mathcal{X} = I\!R$, $\bar{\boldsymbol{\theta}} = (-4, 4, -1)^\top$, so that the response $\eta(\bar{\boldsymbol{\theta}}, \xi)$ reaches its maximum at $\xi^* = 2$. The observation errors ϵ_k are i.i.d. $\mathcal{N}(0, \sigma^2)$, with $\sigma = 0.1$, and the ξ_k's are i.i.d. $\mathcal{N}(1, 1)$. $N = 100$ values of ξ_k are proposed, and $n = 10$ observations can be made.

Four strategies are compared, all based on $\Phi(\mathbf{M}) = \det(\boldsymbol{\Omega}^{-1} + \mathbf{M})$, with $\boldsymbol{\Omega} = 10^6 \mathbf{I}$:

- S_1 corresponds to the OLFO rule (11.13);
- S_2 corresponds to the OLFO rule for the solution of the one–step–ahead problem (11.14);
- S_3 corresponds to the optimal rule (11.8) for the maximisation of (11.14);
- S_4 corresponds to the open–loop rule (11.11,11.12) for the maximisation of (11.14), with $\epsilon = 0.01$ and $\bar{A} = 1$.

All the strategies use CE: in S_1, $\bar{\mathbf{M}}$ is replaced by the empirical mean of $\mathcal{M}_k$, see Theorem 2; S_2 (resp. S_3) uses the empirical mean (resp. empirical distribution) of $\mathcal{M}'_k$. Ten random samples $\xi_{-9}, \ldots, \xi_0$ are used to initialise the construction of the empirical quantities. Two criteria are used for the comparison: the determinant $\mathcal{D}$ of the information matrix per sample: $\mathcal{D}(\boldsymbol{u}_1^N) = \det[\mathbf{M}_N/n]$, and the squared error $\mathcal{E}$ for the estimation of ξ^*,

$$\mathcal{E}(\boldsymbol{u}_1^N) = [\bar{\theta}_1/(2\bar{\theta}_2) - \hat{\theta}_1^N/(2\hat{\theta}_2^N)]^2 ,$$

where $\hat{\boldsymbol{\theta}}^N$ is the maximum *a posteriori* estimator, with a normal prior $\mathcal{N}(\hat{\boldsymbol{\theta}}^0, \boldsymbol{\Omega})$, $\hat{\boldsymbol{\theta}}^0 = (0, 3, -0.5)^\top$.

The results obtained for $q = 1000$ independent repetitions of the experiment are given in Table 11.1. From a result in Pronzato (1998a), taking n experiments at random gives on the average

$$E\{\det[\sum_{k=1}^{N} \mathcal{M}_k/n]\} = \frac{n!}{n^p(n-p)!} \det[\bar{\mathbf{M}}] \simeq 1.44\ 10^6 ,$$

and the four strategies above thus yield an important improvement in terms of $\mathcal{D}$. Note that the formula above shows that the approximate design problem defined by the maximisation of $E\{\det[\mathbf{M}_N/n]\}$ with $n \to \infty$ (for instance, $n = \gamma N$, $0 < \gamma < 1$, $N \to \infty$) corresponds to the determination of an optimal constrained measure, a problem for which one can refer to Wynn (1982), Fedorov(1989) and in this volume, to Sahm and Schwabe (2000).

	$\mathcal{D}$	$\mathcal{E}$
S_1	$8.92\ 10^6$ $(9.5\ 10^6)$	$5.09\ 10^{-5}$ $(1.0\ 10^{-4})$
S_2	$8.71\ 10^6$ $(9.3\ 10^6)$	$5.03\ 10^{-5}$ $(1.3\ 10^{-4})$
S_3	$1.82\ 10^7$ $(1.3\ 10^7)$	$2.84\ 10^{-5}$ $(4.3\ 10^{-5})$
S_4	$1.73\ 10^7$ $(1.3\ 10^7)$	$2.87\ 10^{-5}$ $(4.7\ 10^{-5})$

Table 11.1 Average performances of strategies S_1 to S_4 ($N = 100$, $n = 10$, 1000 repetitions, standard deviations are in brackets)

Since the standard deviations in Table 1 are rather large, we perform a statistical analysis of the results, based on the method of paired comparisons. First, we compute the differences between the performances of each pair of strategies, the same ξ_k's and ϵ_k's being used for all strategies in each experiment. For comparing S_j with S_k, we compute $\Delta^{j,k}_{\mathcal{D},i} = [\mathcal{D}(S_j)]_i - [\mathcal{D}(S_k)]_i$, with $[\mathcal{D}(S)]_i$ the value of $\mathcal{D}$ obtained for strategy S in the i-th experiment. This gives $q = 1000$ independent realisations of $\Delta^{j,k}_{\mathcal{D},i}$, with empirical mean $E(\Delta^{j,k}_{\mathcal{D}})$ and standard deviation $\sigma(\Delta^{j,k}_{\mathcal{D}})$. The same is done with the criterion $\mathcal{E}$, which yields $E(\Delta^{j,k}_{\mathcal{E}})$ and $\sigma(\Delta^{j,k}_{\mathcal{E}})$. Then, we test if S_j performs significantly better or worse than S_k in terms of $\mathcal{D}$ and $\mathcal{E}$ by computing the ratios $\rho^{j,k}_{\mathcal{D}} = \sqrt{q}E(\Delta^{j,k}_{\mathcal{D}})/\sigma(\Delta^{j,k}_{\mathcal{D}})$ and $\rho^{j,k}_{\mathcal{E}} = \sqrt{q}E(\Delta^{j,k}_{\mathcal{E}})/\sigma(\Delta^{j,k}_{\mathcal{E}})$. If the two decision rules have similar average performances, which corresponds to the null hypothesis, then $\rho^{j,k}_{\mathcal{D}}$ and $\rho^{j,k}_{\mathcal{E}}$ approximately follow Student's t-distribution with $q - 1$ degrees of freedom. For large values of q, which is the case here, the distribution is approximately $\mathcal{N}(0, 1)$, and the critical value, for a level of significance 0.5% (one sided–test), is 2.576. Values $\rho^{j,k}_{\mathcal{D}}$ larger than 2.576 (resp. $\rho^{j,k}_{\mathcal{E}}$ smaller than -2.576) thus indicate that S_j performs significantly better than S_k in terms of $\mathcal{D}$ (resp. $\mathcal{E}$). The values of $\rho^{j,k}_{\mathcal{D}}$ and $\rho^{j,k}_{\mathcal{E}}$ are given in Table 11.2, with indices j, k corresponding respectively to lines and columns.

Table 11.2 indicates that the performances of S_1 and S_2 are not significantly different, and that S_3 and S_4 perform significantly better than S_1 and S_2, both in terms of $\mathcal{D}$ and $\mathcal{E}$. We conclude this example by indicating the average computing time for the four strategies. They are respectively 2.3, 0.14, 27.9

$\rho_{\mathcal{D}}^{j,k}$	S_1	S_2	S_3	S_4	$\rho_{\mathcal{E}}^{j,k}$	S_1	S_2	S_3	S_4
S_1	·	1.45	-23.12	-23.70	S_1	·	0.12	6.89	6.53
S_2	-1.45	·	-24.14	-24.64	S_2	-0.12	·	5.32	5.15
S_3	23.12	24.14	·	4.05	S_3	-6.89	-5.32	·	-0.16
S_4	23.70	24.64	-4.05	·	S_4	-6.53	-5.15	0.16	·

Table 11.2 Values of $\rho_{\mathcal{D}}^{j,k}$ and $\rho_{\mathcal{E}}^{j,k}$ (method of paired comparisons) for strategies S_1 to S_4 ($N = 100$, $n = 10$, 1000 repetitions)

and 0.35 seconds. The evaluation of expected values of determinants in S_1, see Theorem 2, explains the larger computing time than for S_2. Taking the small computing time of S_4 into account, together with the fact that the performances of S_3 and S_4 are very close, we conclude that S_4 is the most attractive strategy.

References

Embrechts, P., Klüppelberg, C. and Mikosch T. (1997). *Modelling Extremal Events*. Berlin: Springer-Verlag.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. N.Y.: Academic Press.

Fedorov, V.V. (1989). Optimal design with bounded density: optimization algorithms of the exchange type. *JSPI* **22**, 1–13.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*. N.Y.: Wiley.

Pronzato, L. (1998a). On a property of the expected value of a determinant. *Statistics & Probability Letters* **39**, 161–165.

Pronzato, L. (1998b). Optimal selection of information with restricted storage capacity. In *Proc. ICASSP'98* **4**, pp. 2285–2288. Seattle.

Pronzato, L. (1999a). Optimal and asymptotically optimal decision rules for sequential screening and resource allocation. Technical Report 99–19, Laboratoire I3S, CNRS–UPRES-A 6070, Sophia Antipolis, France, `http://www.i3s.unice.fr/~pronzato/` .

Pronzato, L. (1999b). Sequential selection of observations in randomly generated experiments. *Tatra Mountains Math. Pub.* **17**, 167–175.

Wynn, H.P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Stat.* **41**, 1655–1664.

Wynn, H.P. (1982). Optimum submeasures with applications to finite population sampling. In *Statistical Decision Theory and Related Topics III. Proc. 3rd Purdue Symp* **2** (Eds S.S. Gupta and J.O. Berger), 485-495. N.Y.: Academic Press.