SEQUENTIAL EXPERIMENTAL DESIGN AND EXTREMUM CONTROL

L. Pronzato*, **É. Thierry***

* Laboratoire I3S
CNRS/Université de Nice-Sophia Antipolis
bât. Euclide, Les Algorithmes
2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France
fax: + 33 (0)4 92 94 28 98
e-mail: {pronzato,et}@i3s.unice.fr

Keywords: Extremum control, adaptive control, experimental design, response optimisation, sequential design.

Abstract

We consider the situation where one wants to maximise a function $f(\theta, \mathbf{x})$ with respect to $\mathbf{x}$, with $f(\theta, \mathbf{x})$ linear in θ, θ unknown and estimated from observations $y_k = f(\theta, \mathbf{x}_k) + \epsilon_k$, where ϵ_k is a random error (linear regression model). Special attention is given to sequences defined by $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}} f(\hat{\theta}^k, \mathbf{x}) + \alpha_k d_k(\mathbf{x})$, with $\hat{\theta}^k$ an estimated value of θ obtained from $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_k, y_k)$ and $d_k(\mathbf{x})$ a penalty for poor estimation. Asymptotic results are given (strong consistency of $\hat{\theta}^k$) for a particular penalty function d_k and suitable weighting sequences $\{\alpha_k\}$. Approximately optimal rules are suggested for the finite horizon case where one wants to maximize $\sum_{i=1}^N w_i f(\theta, \mathbf{x}_i)$, with $\{w_i\}$ a given weighting sequence. Various examples are presented.

1 Introduction

We consider an optimisation problem, where one wants to maximise a function $f(\theta, \mathbf{x})$ with respect to $\mathbf{x} \in \mathcal{X} \subset I\!R^q$, where θ is an unknown vector of parameters, $\theta \in I\!R^p$. We assume that $f(\theta, \mathbf{x})$ is linear in θ (for instance a quadratic function of $\mathbf{x}$), that is $f(\theta, \mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\theta$. One observes

$$y_k = f(\theta, \mathbf{x}_k) + \epsilon_k\,,\ k = 1, 2, \ldots \qquad (1)$$

with ϵ_i an unobservable error such that $E\{\epsilon_k | \epsilon_1, \ldots, \epsilon_{k-1}\} = 0$ and $E\{\epsilon_k^2 | \epsilon_1, \ldots, \epsilon_{k-1}\} < \infty$, $k = 1, \ldots$ almost surely (a.s.).

We shall denote $\mathcal{F}_k$ the σ-field generated by $y_1, \ldots, y_k$, $E\{\cdot|\mathcal{F}_k\}$ the corresponding posterior expectation and $E\{\cdot|\mathcal{F}_0\}$ the prior expectation, with a prior probability measure μ for θ. The sequence of design points $\{\mathbf{x}_k\}$ and observations $\{y_k\}$ is used to estimate θ. If the issue were only to determine the value $\mathbf{x}^*$ that maximises $f(\theta, \mathbf{x})$, one could resort to optimum-design theory for choosing an appropriate sequence of inputs $\{\mathbf{x}_k\}$, see, *e.g.*, [13, 27, 7, 8, 21, 24, 12]. However, here the objective is also to have each $f(\theta, \mathbf{x}_i)$ as large as possible, so that the sequence $\{\mathbf{x}_k\}$ must simultaneously fulfill two objectives (generally contradictory): (i) help locate $\mathbf{x}^*$, (ii) be close to $\mathbf{x}^*$ in order to maximise $f(\theta, \mathbf{x}_k)$, $k = 1, 2, \ldots$ The problem is thus one of *dual control*. Note that the $\mathbf{x}_i$'s will be chosen sequentially, that is, $\mathbf{x}_i$ is $\mathcal{F}_{i-1}$-measurable, $i = 1 \ldots$ A well–known example corresponds to the so–called "self–tuning optimiser" or "self–tuning extremum control" problem, see [26], where the worth $f(\theta, \mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\theta$ is quadratic in $\mathbf{x}$. In this case, the value $\mathbf{x}_{k+1}$ maximizing $f(\hat{\theta}^k, \mathbf{x})$ is obtained analytically. However, using this value at the next step (which corresponds to "certainty equivalence control", see [4]) does not guarantee convergence of $\mathbf{x}_k$ to $\mathbf{x}^*$ that maximises $f(\bar{\theta}, \mathbf{x})$. For instance, using the ODE method of Ljung [20], Bozin and Zarrop [5] give the set of values of θ and x to which $\hat{\theta}^k$ and x_k may converge when $f(\theta, x) = \theta_1 x + \theta_2 x^2$: the set of limiting values for x_k contains $x^* = -\bar{\theta}_1/(2\bar{\theta}_2)$ but is not restricted to it. It has thus been suggested to randomly perturb the certainty equivalence control law in order to obtain convergence, see [5]. Another class of example corresponds to regulation problems, where one wishes to minimise the deviation of the response $\mathbf{r}^\top(\mathbf{x})\bar{\theta}$ from a given target. Again, the addition of random disturbances to the certainty equivalence control law can be used to obtain convergence, see, *e.g.*, [19], where the problem of how often probing inputs (disturbances) should be introduced is considered.

A rather general formulation of the problem is given in [14]:

$$\text{maximize} \quad E\{\sum_{i=1}^{\infty} w_i y_i | \mathcal{F}_0\} / \sum_{i=1}^{\infty} w_i \tag{2}$$

with respect to $\mathbf{x}_1, \mathbf{x}_2, \ldots$, with $\{w_i\}$ a weighting (discount) sequence. The choice $w_i = 1$ for $i = N+1$ and $w_i = 0$ otherwise corresponds to a pure design problem, where emphasis is put on the estimation of $\mathbf{x}^*$ after N observations; $w_i = 1$, $i = 1, \ldots, N$, $w_{N+1} = K$ and $w_i = 0$ for $i > N+1$ corresponds to the case where the best guess for $\mathbf{x}^*$ at step N is used for the next K steps. The more classical finite horizon case with no discount corresponds to $w_i = 1$, $i = 1, \ldots, N$, $w_i = 0$ for $i \geq N+1$, *etc.* Since $E\{y_i|\theta, \mathbf{x}_i\} = f(\theta, x_i)$, (2) becomes

$$\text{maximize} \quad E\{\sum_{i=1}^{\infty} w_i f(\theta, \mathbf{x}_i) | \mathcal{F}_0\} / \sum_{i=1}^{\infty} w_i \,. \tag{3}$$

The stochastic dynamic programming formulation of (3), in the finite horizon case, involves imbedded expectations and maximisations, which makes the solution extremely difficult, except in very particular situations. Simple suboptimal solutions have therefore been proposed, see, *e.g.*, [3]. Certainty equivalence control is generally not satisfactory due to its *passive* character: $\mathbf{x}_k$ does not help estimating θ. The addition of random disturbances to the certainty equivalence control law, as mentioned above, is an example of suboptimal strategy. Another suboptimal *active* strategy is proposed for instance in [23, 16], and a comparison between different strategies is presented in [1] (note that for the strategies presented in [23, 16], based on prediction of posterior covariance matrices of θ, the function to be optimized needs be nonlinear in θ, and therefore they do not apply directly here). The problem gets even more complicated when the horizon is infinite, and the restriction is then to even simpler strategies.

We shall mainly consider design sequences that are constructed as follows: at step k, $\mathbf{x}_{k+1}$ maximizes the sum of the predicted value of f, that is, $f(\hat{\theta}^k, \mathbf{x})$ with $\hat{\theta}^k$ the current estimated value of θ, and a weighted penalty term $\alpha_k d_k(\mathbf{x})$, with $d_k(\mathbf{x})$ also a function of $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, k$:

$$\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathcal{X}} f(\hat{\theta}^k, \mathbf{x}) + \alpha_k d_k(\mathbf{x}) \,. \tag{4}$$

In Section 2 we give an asymptotic result (infinite horizon, no discount) obtained when $d_k(\mathbf{x})$ is the variance function used in the construction of D-optimum designs. Other penalty functions related to L-optimum design are then suggested. Section 3 is devoted to the finite horizon case: approximately optimal strategies are suggested, and particular sequences of weights $\{\alpha_k\}$ and penalty functions $d_k(\mathbf{x})$ in (4) are obtained through series of approximations of the original problem (3), based on expansions in the noise variance. Examples are given in Section 4. Finally, Section 5 concludes and draws some perspectives.

2 Asymptotic results for linear response optimisation

2.1 D-optimum penalty

Consider the case where the horizon is infinite ($N = \infty$) and there is no discount ($w_i = 1$ for any i). We use the penalty given by the variance function used in the sequential construction of D-optimum designs, see [29, 11],

$$d_k(\mathbf{x}) = d_k^D(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x}) \mathbf{M}_k^{-1} \mathbf{r}(\mathbf{x}) \,, \tag{5}$$

with $\mathbf{M}_k$ the design matrix

$$\mathbf{M}_k = \sum_{i=1}^{k} \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \,. \tag{6}$$

We shall denote ξ_k the design measure generated, that is, the empirical measure of the $\mathbf{x}_k$'s.

Assume that $\mathbf{r}(\mathbf{x})$ is continuous on $\mathcal{X}$ compact, that the first K_0 regressors $\mathbf{r}(\mathbf{x}_1),\ldots,\mathbf{r}(\mathbf{x}_{K_0})$ are such that $\mathbf{M}_{K_0}$ is positive definite. Also assume that θ in (1) takes a deterministic (but unknown) value $\bar{\theta}$ and that $\mathbf{r}^\top(\mathbf{x})\bar{\theta}$ has a unique global maximiser $\mathbf{x}^*$ in $\mathcal{X}$; that is:

$$\begin{array}{l}\forall\beta>0\,,\ \exists\epsilon>0\ \text{ such that}\\ \mathbf{r}^\top(\mathbf{x})\bar{\theta}+\epsilon>\mathbf{r}^\top(\mathbf{x}^*)\bar{\theta}\Rightarrow\|\mathbf{x}-\mathbf{x}^*\|<\beta\,.\end{array}\tag{7}$$

We estimate θ by least squares (LS):

$$\hat{\theta}^k=\arg\min_{\theta\in\Theta}\sum_{i=1}^k[y_i-\mathbf{r}^\top(\mathbf{x}_i)\theta]^2\,,\tag{8}$$

with Θ a compact subset of $\mathbb{R}^p$ such that $\bar{\theta}\in\Theta$. The difficulty is that for suitable weighting sequences $\{\alpha_k\}$, the sequence $\{\mathbf{x}_k\}$ accumulates at the value $\mathbf{x}^*(\theta)$ that maximises $f(\theta,\mathbf{x})$ for some θ, but, at the same time, when $p=\dim(\theta)>1$, a sequence too much concentrated yields a singular design matrix, and thus does not ensure consistency of $\hat{\theta}^k$. Using the results in [18] on almost sure convergence of LS estimates, the following theorem is proved in [22].

Theorem 1 *Assume that sequence of weights $\{\alpha_k\}$ in (4) is such that $(\alpha_k/k)\log\alpha_k$ decreases monotonically and $\alpha_k/(\log k)^{1+\delta}$ increases monotonically to ∞ for some $\delta>0$. Then, the sequence $\{\mathbf{x}_k\}$ generated by (4) and (5) is such that $\hat{\theta}^k\to\bar{\theta}$, $(1/k)\sum_{i=1}^k\mathbf{r}^\top(\mathbf{x}_i)\bar{\theta}\to\mathbf{r}^\top(\mathbf{x}^*)\bar{\theta}$ and $\xi_k\xrightarrow{w}\xi_{\mathbf{x}^*}$ (in the sense of weak convergence of measures) almost surely (a.s.) as $k\to\infty$, with $\mathbf{x}^*=\mathbf{x}^*(\bar{\theta})=\arg\max_{\mathbf{x}\in\mathcal{X}}\mathbf{r}^\top(\mathbf{x})\bar{\theta}$ and $\xi_{\mathbf{x}}$ the discrete measure that puts weight 1 at the point $\mathbf{x}$.*

One can note that, from the Lebesgue dominated convergence theorem, Theorem 1 implies

$$E\{(1/k)\sum_{i=1}^k\mathbf{r}^\top(\mathbf{x}_i)\bar{\theta}\}\to E\{\mathbf{r}^\top[\mathbf{x}^*(\bar{\theta})]\bar{\theta}\}$$

when the prior μ for $\bar{\theta}$ is supported on Θ compact. Also note that taking a penalty function of the form $d_k(\mathbf{x})=\lambda\det\mathbf{M}_{k+1}/\det\mathbf{M}_k$, as suggested in [3], which corresponds to taking $\alpha_k=\alpha$ constant in (4), does not guarantee that $\lambda_{\min}(\mathbf{M}_k)\to\infty$, one of the conditions in [18]. An example in [22] illustrates the convergence problems due to this insufficient penalization of poor estimation. The approach suggested in [2], which corresponds to modifying the certainty–equivalence control law only when $\operatorname{trace}\mathbf{M}_k$ is smaller than some predefined threshold, suffers the same difficulty, that is, does not guarantee $\lambda_{\min}(\mathbf{M}_k)\to\infty$.

Using the Bayesian imbedding approach of [25, 17], one can obtain a.s. convergence results with weaker conditions on $\{\alpha_k\}$ under the assumption that the errors ϵ_i are i.i.d. and Gaussian $\mathcal{N}(0,\sigma^2)$. Indeed, for $\Theta=\mathbb{R}^p$ and $k>K_0$ the LS estimator $\hat{\theta}^k$ then coincides with the Bayesian estimator $E\{\theta|\mathcal{F}_k\}$ for the prior μ given by $\mathcal{N}(\hat{\theta}^{K_0},\sigma^2\mathbf{M}_{K_0}^{-1})$ ($\mathbf{M}_{K_0}$ is positive definite by assumption). Let Q denote the probability measure induced by $\{\epsilon_k\}$, and write $(\mu\times Q)$-a.s. for a property almost sure in the sense of the product measure $\mu\times Q$. From the martingale convergence theorem, $\hat{\theta}^\infty=\lim_{k\to\infty}\hat{\theta}^k$ exists and is finite and $\mathbf{r}^\top(\mathbf{x})\hat{\theta}^\infty$ is bounded on $\mathcal{X}$, $(\mu\times Q)$-a.s., and the posterior covariance matrix $\mathbf{C}_k$ tends to some limit $\mathbf{C}_\infty$, $(\mu\times Q)$-a.s. When the smallest eigenvalue of $\mathbf{M}_k$ satisfies $\lambda_{\min}(\mathbf{M}_k)\to\infty$, $(\mu\times Q)$-a.s., $\mathbf{C}_\infty$ is the null matrix and $\hat{\theta}^k$ converges to $\bar{\theta}$, the value of θ that generates the observations, $(\mu\times Q)$-a.s. A straightforward extension of Theorem 1 is then as follows.

Corollary 1 *Assume that the errors ϵ_i are i.i.d. $\mathcal{N}(0,\sigma^2)$ and that sequence of weights $\{\alpha_k\}$ in (4) is such that $\alpha_k\to\infty$ and $\alpha_k/k\to 0$, Then, the sequence $\{\mathbf{x}_k\}$ generated by (4) and (5) is such that $\hat{\theta}^k\to\bar{\theta}$, $(1/k)\sum_{i=1}^k\mathbf{r}^\top(\mathbf{x}_i)\bar{\theta}\to\mathbf{r}^\top(\mathbf{x}^*)\bar{\theta}$ and $\xi_k\xrightarrow{w}\xi_{\mathbf{x}^*}$ (in the sense of weak convergence of measures), $(\mu\times Q)$-a.s., as $k\to\infty$.*

The condition on α_k is weaker in Corollary 1 than in Theorem 1, but note that there may be a singular set (with respect to the Lebesgue measure) for $\bar{\theta}$ for which $\hat{\theta}^k$ is not consistent.

Using the results in [15], the assumption of normality in Corollary 1 can be relaxed, provided (i) the errors ϵ_i are i.i.d. with an almost everywhere strictly positive density h with respect to the Lebesgue measure, such that h'' is continuous and $(\log h)'' < 0$, (ii) the prior measure μ is absolutely continuous with respect to the Lebesgue measure and (iii) the LS estimates $\hat{\theta}^k$ are replaced by $E\{\theta|\mathcal{F}_k\}$.

2.2 L-optimum penalty

Using an idea similar to previous section, we can use a penalty function related to L-optimum design; that is,

$$d_k(\mathbf{x}) = d_k^L(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\mathbf{M}_k^{-1}\mathbf{H}\,\mathbf{M}_k^{-1}\mathbf{r}(\mathbf{x})\,, \quad (9)$$

with $\mathbf{H}$ a non negative definite matrix. It is shown in [28] that when $\mathbf{H}$ is positive definite and the sequence $\{\mathbf{x}_k\}$ is generated by $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} d_k^L(\mathbf{x})$, the design measure ξ_k converges to a L-optimum design measure ξ_L^* that minimises $\mathrm{trace}\,\mathbf{H}\,\mathbf{I}^{-1}(\xi)$, where $\mathbf{I}(\xi) = \int_{\mathcal{X}} \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})\xi(d\mathbf{x})$. Further work is required to check if a property similar to Theorem 1 can be obtained in this case.

A case that has retained much attention in when the design objective corresponds to the estimation of the point $\mathbf{x}^*$ where $f(\theta,\mathbf{x})$ achieves its maximum, see [24], especially when f is a quadratic function on $\mathbb{R}$, see, *e.g.*, [12]. Assume that

$$f(\theta,x) = \theta_0 + \theta_1 x + \theta_2 x^2/2\,, \quad (10)$$

with $x \in \mathcal{X}$, a compact subset of the real line. One has $x^*(\theta) = -\theta_1/\theta_2$, and, in the case where the errors ϵ_i are i.i.d., the asymptotic variance of $x^*(\hat{\theta}^k)$, with $\hat{\theta}^k$ the LS estimator (8), is proportional to $\mathbf{c}^\top\mathbf{M}_k^{-1}\mathbf{c}$, with

$$\mathbf{c} = \mathbf{c}(\theta) = \frac{\partial x^*(\theta)}{\partial\theta} = (-1/\theta_2)(0 \ \ 1 \ \ x^*)^\top\,.$$

Choosing the $\mathbf{x}_k$'s in order to maximise the accuracy of the estimation of $\mathbf{x}^*(\theta)$ corresponds to c-optimum design, that is, to L-optimum design with $\mathbf{H}$ given by the rank-one matrix $\mathbf{c}\mathbf{c}^\top$. Note that the dependence of $\mathbf{c}$ in θ makes the problem nonlinear. A Bayesian approach is used in [7, 10], based on the design criterion $E\{\mathbf{c}^\top\mathbf{M}_k^{-1}\mathbf{c}\}$, where $E\{\cdot\}$ denotes expectation with respect to θ for a given prior. Sequential approaches are considered in [13, 21]. One can also refer to [6] for the use of c-optimal design in the context of Bayesian estimation and to [9] for a survey on Bayesian experimental design. Following (9), a penalty function related to c-optimal design is thus

$$d_k(\mathbf{x}) = d_k^c(\mathbf{x}) = [\mathbf{r}^\top(\mathbf{x})\mathbf{M}_k^{-1}\mathbf{c}(\hat{\theta}^k)]^2\,. \quad (11)$$

Another approach, used in [24], is to derive the design criterion from the construction of a Bayesian risk related to the maximisation of $f(\theta,\mathbf{x})$. Assume that θ has a normal prior $\mathcal{N}(\hat{\theta}^0,\sigma^2\Omega)$, that the errors ϵ_i are i.i.d. $\mathcal{N}(0,\sigma^2)$ and that the discount factors w_i satisfy $w_i = 1$ for $i = N+1$ and $w_i = 0$ otherwise. When $\mathbf{x}_1,\ldots,\mathbf{x}_N$ are all chosen at the same time, this leads to the following (non sequential) L_B-optimal design problem (B stands for Bayesian):

$$\text{minimise} \ \ \mathrm{trace}\,\mathbf{H}(\hat{\theta}^0)\,(\mathbf{M}_k + \Omega^{-1})^{-1}$$

with

$$\mathbf{H}(\hat{\theta}) = \frac{\partial^2 f[\theta,\mathbf{x}^*(\theta)]}{\partial\theta\partial\theta^\top}_{|\hat{\theta}}\,.$$

The matrix $\mathbf{H}$ can easily be proved to be non negative definite when f is linear in θ, see [24], and

$$\mathbf{H}(\theta) = \frac{\partial\mathbf{r}(\mathbf{x})}{\partial\mathbf{x}^\top}_{|\mathbf{x}^*(\theta)}\,\frac{\partial\mathbf{x}^*(\theta)}{\partial\theta^\top}\,, \quad (12)$$

which can be expressed analytically when f is quadratic in $\mathbf{x}$. For instance, in the case where f is given by (10) (with $\theta_2 < 0$ in order to have a function concave in x), one gets

$$\begin{aligned}\mathbf{H}(\hat{\theta}^0) &= (-1/\hat{\theta}_2^0)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \mathbf{x}^*(\hat{\theta}^0) \\ 0 & \mathbf{x}^*(\hat{\theta}^0) & [\mathbf{x}^*(\hat{\theta}^0)]^2 \end{pmatrix} \\ &= -\hat{\theta}_2^0\,\mathbf{c}(\hat{\theta}^0)\mathbf{c}^\top(\hat{\theta}^0)\,.\end{aligned}$$

This suggests substitution of $\mathbf{H}(\hat{\theta}^k)$ for $\mathbf{H}$ in (9) in the case of sequential design. Again, in the infinite horizon case with no discount, the choice of a weighting sequence $\{\alpha_k\}$ in (4) ensuring convergence of $\mathbf{x}^*(\hat{\theta}^k)$ to $\mathbf{x}^*(\bar{\theta})$ and of ξ_k to $\xi_{\mathbf{x}^*(\bar{\theta})}$ remains an open issue.

3 Linear response optimisation with finite horizon

Assume that the errors ϵ_k in (1) are i.i.d. $\mathcal{N}(0,\sigma^2)$. An expansion in σ^2 will be used to derive an approximate solution to the problem (3). Our result is based on the following Lemma.

Lemma 1 *Let $g(\cdot)$ and $h(\cdot)$ be two times continuously differentiable functions on $\mathcal{X}$, a compact set of $I\!R^q$. Assume that $g(\cdot)$ has a unique global maximum at $\mathbf{x}^*$, an interior point of $\mathcal{X}$, with $\partial^2 g(\mathbf{x})/\partial\mathbf{x}\partial\mathbf{x}^\top_{|\mathbf{x}^*}$ negative definite, and let $\hat{\mathbf{x}}$ denote the point where $f(\cdot)=g(\cdot)+uh(\cdot)$ reaches its maximum in $\mathcal{X}$. Then, $\|\hat{\mathbf{x}}-\mathbf{x}^*\|=O(u)$ and $|f(\mathbf{x}^*)-f(\hat{\mathbf{x}})|=O(u^2)$, $u\to 0$.*

Proof. For u small enough, $\hat{\mathbf{x}}$ is an interior point of $\mathcal{X}$ so that

$$\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top}_{|\hat{\mathbf{x}}}=0 &= \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}^\top}_{|\mathbf{x}^*}+(\hat{\mathbf{x}}-\mathbf{x}^*)^\top\frac{\partial^2 g(\mathbf{x})}{\partial\mathbf{x}\partial \mathbf{x}^\top}_{|\mathbf{x}^*} \\
&\quad +u\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}^\top}_{|\hat{\mathbf{x}}}+o(\|\hat{\mathbf{x}}-\mathbf{x}^*\|) \\
&= (\hat{\mathbf{x}}-\mathbf{x}^*)^\top\frac{\partial^2 g(\mathbf{x})}{\partial\mathbf{x}\partial \mathbf{x}^\top}_{|\mathbf{x}^*} \\
&\quad +u\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}^\top}_{|\hat{\mathbf{x}}}+o(\|\hat{\mathbf{x}}-\mathbf{x}^*\|)\,,
\end{aligned}$$

and $\|\hat{\mathbf{x}}-\mathbf{x}^*\|=O(u)$. Therefore,

$$\begin{aligned}
f(\mathbf{x}^*) &= f(\hat{\mathbf{x}})+(\mathbf{x}^*-\hat{\mathbf{x}})^\top\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}_{|\hat{\mathbf{x}}} \\
&\quad +\frac{1}{2}(\mathbf{x}^*-\hat{\mathbf{x}})^\top\frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{x}\partial \mathbf{x}^\top}_{|\hat{\mathbf{x}}}(\mathbf{x}^*-\hat{\mathbf{x}}) \\
&\quad +o(\|\hat{\mathbf{x}}-\mathbf{x}^*\|^2)=f(\hat{\mathbf{x}})+O(u^2)\,.
\end{aligned}$$

■

Next theorem gives an approximation to the optimal solution of problem (3).

Theorem 2 *Assume that $w_i>0$, $i=1,\ldots,N$ and $w_i=0$ otherwise; that $\mathbf{r}(\cdot)$ is two times continuously differentiable; that the errors ϵ_i are i.i.d. $\mathcal{N}(0,\sigma^2)$ and the prior distribution for θ is normal $\mathcal{N}(\hat{\theta}^0,\sigma^2\Omega)$. Denote $\mathbf{x}^*(\theta)=\arg\max_{\mathbf{x}\in\mathcal{X}}\mathbf{r}^\top(\mathbf{x})\theta$ and $\tilde{\theta}^k=E\{\theta|\mathcal{F}_k\}$, and assume that $f(\tilde{\theta}^j,\mathbf{x})$ has a unique global maximum at $\mathbf{x}^*(\tilde{\theta}^j)$ which lies in the interior of $\mathcal{X}$, with $\partial^2 f(\tilde{\theta}^j,\mathbf{x})/\partial\mathbf{x}\partial\mathbf{x}^\top_{|\mathbf{x}^*(\tilde{\theta}^j)}$ negative definite, $j=0,\ldots,N-2$. Define $j_{k+1}(\mathbf{x})$ as the expected optimal gain to go at step k when $\mathbf{x}$ is applied:*

$$\begin{aligned}
j_{k+1}(\mathbf{x}) &= E\{w_{k+1}f(\theta,\mathbf{x}) \\
&\qquad +\max_{\mathbf{z}\in\mathcal{X}} w_{k+2}j_{k+2}(\mathbf{z})|\mathcal{F}_k\}\,, \\
&\qquad\quad k=0,\ldots,N-2\,, \\
j_N(\mathbf{x}) &= E\{w_N f(\theta,\mathbf{x})|\mathcal{F}_{N-1}\}\,.
\end{aligned}$$

It satisfies

$$j_{k+1}(\mathbf{x})=J_{k+1}(\mathbf{x})+O(\sigma^4)\,,\ k=0,\ldots,N-2\,, \tag{13}$$

where

$$\begin{aligned}
J_{k+1}(\mathbf{x}) &= (w_N+\cdots+w_{k+2})\{\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^k)]\tilde{\theta}^k \\
&+ \frac{\sigma^2}{2}\mathrm{trace}\,[\mathbf{H}(\tilde{\theta}^k)(\Omega^{-1}+\mathbf{M}_k)^{-1}]\} \\
&- \frac{\sigma^2}{2}\mathrm{trace}\,\{\mathbf{H}(\tilde{\theta}^k)\sum_{j=0}^{N-k-2}w_{k+j+2}[\Omega_{j,k}^{-1} \\
&+ \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})]^{-1}\}+w_{k+1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^k
\end{aligned} \tag{14}$$

with $\Omega_{j,k}=\left\{\Omega^{-1}+\mathbf{M}_k+j\mathbf{r}[\mathbf{x}^*(\tilde{\theta}^k)]\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^k)]\right\}^{-1}$ and $\mathbf{M}_k$, $\mathbf{H}(\theta)$ respectively given by (6), (12). Moreover, the strategy defined by $\mathbf{x}_N=\mathbf{x}^*(\tilde{\theta}^{N-1})$ and $\mathbf{x}_{k+1}=\arg\max_{\mathbf{x}\in\mathcal{X}}J_{k+1}(\mathbf{x})$, $k=0,\ldots,N-2$, is equivalently defined by

$$\begin{aligned}
\mathbf{x}_{k+1} &= \arg\max_{\mathbf{x}\in\mathcal{X}}\{w_{k+1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^k \\
&+\frac{\sigma^2}{2}\sum_{j=0}^{N-k-2}w_{k+j+2}\frac{\mathbf{r}^\top(\mathbf{x})\Omega_{j,k}\mathbf{H}(\tilde{\theta}^k)\Omega_{j,k}\mathbf{r}(\mathbf{x})}{1+\mathbf{r}^\top(\mathbf{x})\Omega_{j,k}\mathbf{r}(\mathbf{x})}\}\,, \\
&\qquad k=0,\ldots,N-2\,,
\end{aligned} \tag{15}$$

and satisfies

$$\begin{aligned}
E\{\sum_{i=k+1}^{N}w_i f(\theta,\mathbf{x}_i)|\mathcal{F}_k\} &= J_{k+1}(\mathbf{x}_{k+1})+O(\sigma^4)\,, \\
k &= 0,\ldots,N-2\,,
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\|\hat{\mathbf{x}}_{k+1}-\mathbf{x}_{k+1}\| &= O(\sigma^4)\,,\ k=0,\ldots,N-2\,, \\
\hat{\mathbf{x}}_N &= \mathbf{x}_N\,,
\end{aligned} \tag{17}$$

where $\hat{\mathbf{x}}_{k+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} j_{k+1}(\mathbf{x})$ *corresponds to the optimum strategy.*

Proof. Straightforward matrix manipulation shows that $\mathbf{x}_{k+1}$ given by (15) maximises (14). We prove (13) and (16) by backward induction on k. For $k = N-2$, we have

$$\begin{aligned} j_{N-1}(\mathbf{x}) &= w_N E\{\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^{N-1})]\theta|\mathcal{F}_{N-2}\} \\ &\quad + w_{N-1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{N-2} \\ &= w_N E\{\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^{N-1})]\tilde{\theta}^{N-1}|\mathcal{F}_{N-2}\} \\ &\quad + w_{N-1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{N-2}\,. \end{aligned}$$

When $\mathbf{x}^*(\tilde{\theta}^{N-2})$ is an interior point of $\mathcal{X}$, a second-order Taylor expansion around $\tilde{\theta}^{N-2}$ similar to that used in [24] gives

$$\begin{aligned} j_{N-1}(\mathbf{x}) &= w_N\{\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^{N-2})]\tilde{\theta}^{N-2} \\ &+ \frac{\sigma^2}{2}\mathrm{trace}\,[\mathbf{H}(\tilde{\theta}^{N-2})(\Omega^{-1}+\mathbf{M}_{N-2})^{-1}]\} \\ &- w_N\frac{\sigma^2}{2}\mathrm{trace}\,\{\mathbf{H}(\tilde{\theta}^{N-2})\left[\mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})\right. \\ &\quad \left.+\,\Omega^{-1}+\mathbf{M}_{N-2}\right]^{-1}\} \\ &+ w_{N-1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{N-2} + O(\sigma^4)\,, \end{aligned}$$

which proves (13) for $k = N-2$. Since $E\{\sum_{i=N-1}^N w_i f(\theta,\mathbf{x}_i)|\mathcal{F}_{N-2}\} = j_{N-1}(\mathbf{x}_{N-1})$, it also proves (16) for $k = N-2$.

Assume that (13) is true at step k. At step $k-1$ we have $j_k(\mathbf{x}) = E\{w_k f(\theta,\mathbf{x}) + \max_{\mathbf{z}\in\mathcal{X}} j_{k+1}(\mathbf{z})|\mathcal{F}_{k-1}\}$. Using Lemma 1 with $u = \sigma^2$, we get

$$\begin{aligned} j_k(\mathbf{x}) &= w_k\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{k-1} + E\{J_{k+1}[\mathbf{x}^*(\tilde{\theta}^k)]|\mathcal{F}_{k-1}\} \\ &\quad + O(\sigma^4)\,. \end{aligned}$$

Grouping the terms $\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^k)]\tilde{\theta}^k$ and using a second-order Taylor expansion around $\tilde{\theta}^{k-1}$, we obtain

$$\begin{aligned} j_k(\mathbf{x}) &= (w_N + \cdots + w_{k+2} + w_{k+1}) \\ &\times \{\mathbf{r}^\top[\mathbf{x}^*(\tilde{\theta}^{k-1})]\tilde{\theta}^{k-1} \\ &\quad + \frac{\sigma^2}{2}\mathrm{trace}\,[\mathbf{H}(\tilde{\theta}^{k-1})(\Omega^{-1}+\mathbf{M}_{k-1})^{-1}]\} \\ &- (w_N + \cdots + w_{k+2} + w_{k+1}) \\ &\times \frac{\sigma^2}{2}\mathrm{trace}\,\{\mathbf{H}(\tilde{\theta}^{k-1})[\Omega^{-1}+\mathbf{M}_{k-1} \\ &\quad + \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})]^{-1}\} \\ &+ (w_N + \cdots + w_{k+2}) \\ &\times \frac{\sigma^2}{2}\mathrm{trace}\,\{\mathbf{H}(\tilde{\theta}^{k-1})[\Omega^{-1}+\mathbf{M}_{k-1} \\ &\quad + \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})]^{-1}\} \\ &- \frac{\sigma^2}{2}\mathrm{trace}\,\{\mathbf{H}(\tilde{\theta}^{k-1})\sum_{j=0}^{N-k-2} w_{k+j+2} \\ &\quad \times[\Omega_{j+1,k-1}^{-1} + \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})]^{-1}\} \\ &+ w_k\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{k-1} + O(\sigma^4)\,. \end{aligned}$$

A simplification of the terms on the last nine lines gives $j_k(\mathbf{x}) = J_k(\mathbf{x}) + O(\sigma^4)$, which proves (13). Since $\mathbf{x}_{k+1}$ maximises (14), similar arguments using (16) and Lemma 1 give $E\{\sum_{i=k}^N w_i f(\theta,\mathbf{x}_i)|\mathcal{F}_{k-1}\} = J_k(\mathbf{x}_k) + O(\sigma^4)$, and (16) is proved by induction.

Finally, since $\mathbf{x}_{k+1}$ maximises $J_{k+1}(\mathbf{x})$, (13) and Lemma 1 with $u = \sigma^4$ give (17).

■

Remark 1 *The assumptions on* $f(\tilde{\theta}^j,\mathbf{x})$ *and* $\mathbf{x}^*(\tilde{\theta}^j)$ *for* $j = 0,\ldots,N-2$, *in Theorem 2 are most often difficult, if not impossible, to check beforehand. Note, however, that it is always possible to apply the strategy and check the assumptions afterwards: if they are satisfied, the theorem applies and the strategy used is approximately optimal in the sense of the theorem.*

The property stated in Theorem 2 suggests a simpler suboptimal strategy: in the backward induction, we substitute $\mathbf{r}(\mathbf{x})$ for $\mathbf{r}[\mathbf{x}^*(\tilde{\theta}^k)]$ in $\Omega_{j,k}$. It satisfies the following property.

Corollary 2 *Under the same conditions, and with the same notations, as in Theorem 2, the strategy defined by* $\mathbf{x}'_N = \mathbf{x}^*(\tilde{\theta}^{N-1})$ *and*

$$\begin{aligned} \mathbf{x}'_{k+1} &= \arg\max_{\mathbf{x}\in\mathcal{X}}\{w_{k+1}\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^k \\ + \frac{\sigma^2}{2} & \sum_{j=1}^{N-k-1} j\, w_{k+j+1}\frac{\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{H}(\tilde{\theta}^k)\Omega_{0,k}\mathbf{r}(\mathbf{x})}{1 + j\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{r}(\mathbf{x})}\}\,, \\ & k = 0,\ldots,N-2\,, \end{aligned} \tag{18}$$

is approximately optimal in the following sense:

$$\max_{\mathbf{x}\in\mathcal{X}} j_{k+1}(\mathbf{x}) - E\{\sum_{i=k+1}^{N} w_i f(\theta, \mathbf{x}'_i)|\mathcal{F}_k\} = O(\sigma^4)\,, \quad (19)$$

$$and \quad |\hat{\mathbf{x}}_{k+1} - \mathbf{x}'_{k+1}| = O(\sigma^2)\,, \quad (20)$$

for $k = 0, \ldots, N-1$.

Proof. We first prove that, for $k = N-2, \ldots, 1$,

$$E\{\sum_{i=k+1}^{N} w_i f(\theta, \mathbf{x}'_i)|\mathcal{F}_k\} = J_{k+1}(\mathbf{x}'_{k+1}) + O(\sigma^4)\,.$$

For $k = N-2$ we have again

$$E\{\sum_{i=N-1}^{N} w_i f(\theta, \mathbf{x}'_i)|\mathcal{F}_{N-2}\} = j_{N-1}(\mathbf{x}'_{N-1})$$

which equals $J_{N-1}(\mathbf{x}'_{N-1}) + O(\sigma^4)$ from Theorem 2. Assume that the property is true at step k. At step $k-1$ we have

$$E\{w_k f(\theta, \mathbf{x}) + \sum_{i=k+1}^{N} w_i f(\theta, \mathbf{x}'_i)|\mathcal{F}_{k-1}\} =$$
$$w_k \mathbf{r}^\top(\mathbf{x})\tilde{\theta}^{k-1} + E\{J_{k+1}(\mathbf{x}'_{k+1})|\mathcal{F}_{k-1}\} + O(\sigma^4)\,,$$

and easy matrix manipulation using (18) and (14) shows that $\mathbf{x}'_{k+1} = \arg\max_{\mathbf{x}\in\mathcal{X}}[J_{k+1}(\mathbf{x}) + \sigma^2 h(\mathbf{x})]$, for some $h(\mathbf{x})$. Lemma 1 with $u = \sigma^2$ gives $J_{k+1}(\mathbf{x}'_{k+1}) = J_{k+1}[\mathbf{x}^*(\tilde{\theta}^k)] + O(\sigma^4)$, and similarly to the proof of Theorem 2, $E\{w_k f(\theta, \mathbf{x}) + \sum_{i=k+1}^{N} w_i f(\theta, \mathbf{x}'_i)|\mathcal{F}_{k-1}\} = J_k(\mathbf{x}) + O(\sigma^4)$.

Finally, since the optimal strategy $\hat{\mathbf{x}}_{k+1}$ maximises $j_{k+1}(\mathbf{x}) = J_{k+1}(\mathbf{x}) + O(\sigma^4)$, whereas $\mathbf{x}'_{k+1}$ maximises a function that takes the form $J_{k+1}(\mathbf{x}) + \sigma^2 h(\mathbf{x})$, Lemma 1 gives (19-20).

■

Remark 2 *It is clear from the proof of Corollary 2 that properties similar to (19-20) can be obtained for other strategies than (18). What makes this strategy attractive is that* $\mathbf{r}(\mathbf{x}'_{k+1})$ *can be expected to be close to* $\mathbf{r}[\mathbf{x}^*(\tilde{\theta}^k)]$ *and the decrease of performance can be expected to be small. The rule (18) is, however, suboptimal compared to (15) that directly maximises* $J_{k+1}(\mathbf{x})$ *(compare (20) to (17)).*

Assume that $w_i = 1$, $i = 1, \ldots, N$ and $w_i = 0$ otherwise. One can write (18) as

$$\begin{aligned}\mathbf{x}'_{k+1} &= \arg\max_{\mathbf{x}\in\mathcal{X}}\{\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^k \\ &+\frac{\sigma^2}{2}\, \frac{\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{H}(\tilde{\theta}^k)\Omega_{0,k}\mathbf{r}(\mathbf{x})}{\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{r}(\mathbf{x})} \sum_{j=1}^{N-k-1} \frac{j}{a_k + j}\,, \\ & k = 0, \ldots, N-2\,,\end{aligned}$$

with $a_k = 1/[\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{r}(\mathbf{x})]$. When $a_k = 0$, $S_N = \sum_{j=1}^{N-k-1} j/(a_k + j) = N-k-1$ and when $N \to \infty$ with k fixed, $S_N = N-k-1-a_k \log(N-k-1) + a_k \Psi(a_k + 1) + O(1/N)$, with $\Psi(\cdot)$ the digamma function, $\Psi(x) = d\log\Gamma(x)/dx$, and $\Psi(x) = \log x - 1/(2x) + O(1/x^2)$, $x \to \infty$. A reasonable approximation of the strategy (18) is thus as follows:

$$\begin{aligned}\mathbf{x}''_{k+1} &= \arg\max_{\mathbf{x}\in\mathcal{X}}\{\mathbf{r}^\top(\mathbf{x})\tilde{\theta}^k \\ +(N-k-1)\frac{\sigma^2}{2} & \frac{\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{H}(\tilde{\theta}^k)\Omega_{0,k}\mathbf{r}(\mathbf{x})}{\mathbf{r}^\top(\mathbf{x})\Omega_{0,k}\mathbf{r}(\mathbf{x})}\}\,, \\ & k = 0, \ldots, N-2\,, \quad (21)\end{aligned}$$

which has the form (4).

4 Examples

We assume that $f(\theta, x)$ is given by (10), with $\mathcal{X} = [-1, 1]$, and that the observations y_k are generated by (1) with $\theta = \bar{\theta} = (0\ \ 3.2\ \ -8)^\top$ and errors ϵ_k i.i.d. $\mathcal{N}(0, \sigma^2)$. The prior for θ is $\mathcal{N}(\hat{\theta}^0, \sigma^2\Omega)$, with $\Omega = 10^6\ \mathbf{I}_3$, where $\mathbf{I}_3$ is the 3-dimensional identity matrix. We take $w_i = 1$, $i = 1, \ldots, N$ and $w_i = 0$ otherwise. We shall compare four different strategies: S_1 corresponds to (4) with the penalty (5) and $\alpha_k = \sigma^2(\log k)^2$, and S_2, S_3 and S_4 correspond respectively to (21), (18) and (15). The parameters θ are estimated by recursive least-squares, that is, $\hat{\theta}^k = E\{\theta|\mathcal{F}_k\}$.

First we illustrate Theorem 2 and Corollary 2. Since we do not know the optimal sequences $\{\hat{x}_k\}$ and $\{j(\hat{x}_k)\}$, we take $\hat{\theta}^0 = \bar{\theta}$ and compare the values of $(1/N)\sum_{i=1}^{N} f(\bar{\theta}, x_i)$ to $f[\bar{\theta}, x^*(\bar{\theta})] = 0.64$ for the three strategies S_2, S_3 and S_4. Figure 1 gives the empirical mean of $(1/N)\sum_{i=1}^{N} f(\bar{\theta}, x_i)$, obtained from 500 independent repetitions for these three strategies, as a function of σ^2, with $N = 4$

(the same values of observations errors ϵ_k are used for the three strategies in each of the 500 experiment). The full line corresponds to S_4. Strategies S_2 and S_3 are indistinguishable (dashed line). The figure indicates that the decrease of performance of the strategies varies as σ^4.

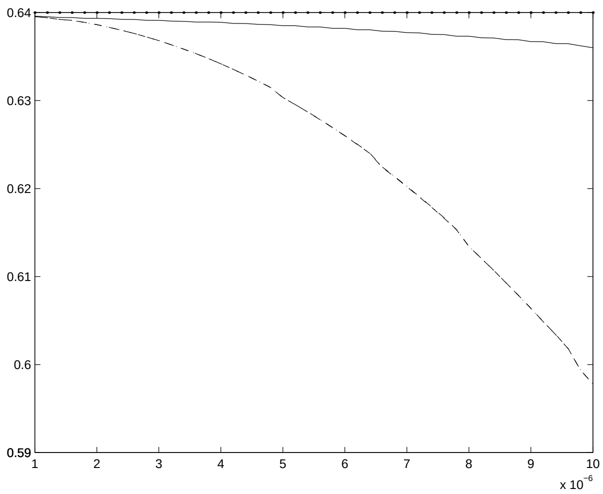

Figure 1: Empirical means of $(1/N)\sum_{i=1}^{N} f(\bar{\theta}, x_i)$ for strategies S_2, S_3 and S_4 as functions of σ^2 ($N = 4$, $\hat{\theta}^0 = \bar{\theta}$, 500 repetitions). Full line for S_4, dashed line for $S_2 \simeq S_3$.

Assume now that $\sigma = 1$, $N = 100$ and take $\hat{\theta}^0 = (2\ -4\ -1)^\top$. Note that this gives a prior guess for x^* at -4, whereas the true location is at 0.4. Also notice the large amplitude of the measurement noise. Table 1 presents the results obtained for 100 independent repetitions of the experiment (in each experiment, the same values of observations errors ϵ_k are used for the four strategies). As expected, performances improve from S_1 to S_4.

	mean	std
S_1	0.1793	0.0445
S_2	0.3759	0.0278
S_3	0.4263	0.0260
S_4	0.4773	0.0454

Table 1: Empirical means and standard deviations (std) of $(1/N)\sum_{i=1}^{N} f(\bar{\theta}, x_i)$ for strategies $S_1, \ldots, S_4$ ($N = 100$, $\hat{\theta}^0 = (2\ -4\ -1)^\top$, 100 repetitions).

Figure 2 (resp. 3) presents a typical realization of the sequences $\{x_k\}$ (resp. $\{f(\bar{\theta}, x_k)\}$) generated by the four strategies. One can notice in Figure 2 that x_k converges to $x^* = -\bar{\theta}_1/\bar{\theta}_2 = 0.4$ for S_1 (see Theorem 1), which does not seem to be the case for the three other strategies. However, the performance measured in terms of (3) is much better for these strategies, in particular S_4 that makes a particularly good compromise between estimation and optimisation, see Figure 3 where the optimum value $f[\bar{\theta}, x^*(\bar{\theta})]$ is indicated by the dashed line.

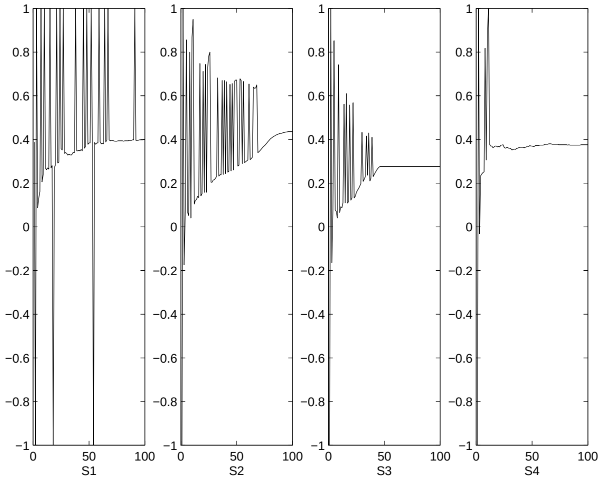

Figure 2: Sequences $\{x_k\}$ generated by strategies S_1 to S_4 ($N = 100$, $\sigma = 1$, $\hat{\theta}^0 = (2\ -4\ -1)^\top$).

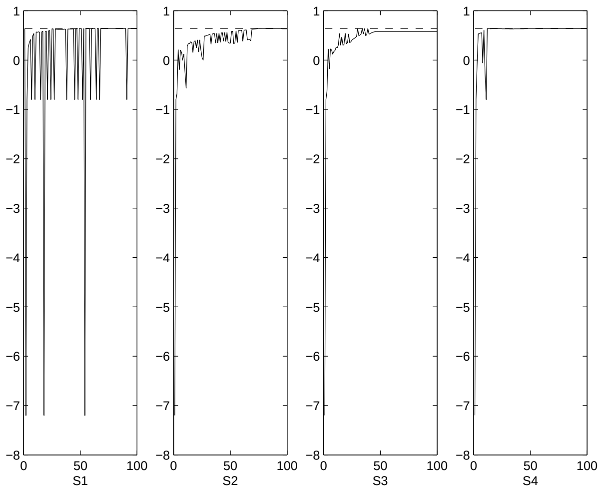

Figure 3: Sequences $\{f(\bar{\theta}, x_k)\}$ generated by strategies S_1 to S_4 ($N = 100$, $\sigma = 1$, $\hat{\theta}^0 = (2\ -4\ -1)^\top$). The optimum value $f[\bar{\theta}, x^*(\bar{\theta})]$ is indicated by the dashed line.

5 Conclusions

We considered the problem of choosing a sequence of values $\mathbf{x}_k$ that maximize a function $f(\theta, \mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\theta$ observed with errors (linear regression model), with θ unknown. Different strategies have been suggested.

Approximately optimal strategies have been constructed when the horizon is finite, using an expansion in the noise variance σ^2. Simulation results confirm that the loss of optimality is negligible when σ^2 is small. In more general situations, we considered sequences constructed according to the rule $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}} f(\hat{\theta}^k, \mathbf{x}) + \alpha_k d_k(\mathbf{x})$, with $\hat{\theta}^k$ an estimated value of θ obtained from $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_k, y_k)$ and $d_k(\mathbf{x})$ a penalty for poor estimation. The asymptotic behavior of such strategies is difficult to study, due to the intricate connection between estimation of θ and optimisation. Only the linear regression case, with a penalty related to D-optimum design, seems to have been considered so far, see [22]. Extensions to other penalty functions, *e.g.*, related to L-optimality, and to nonlinear regression problems will require further developments.

References

[1] B. Allison, P. Tessier, and G. Dumont. Comparison of suboptimal dual adaptive controllers. In *Proc. 3rd European Control Conf.*, pages 36–43, Rome, September 1995.

[2] J. Alster and P. Blanger. A technique for dual adaptive control. *Automatica*, 10:627–634, 1974.

[3] K. Aström and B. Wittenmark. *Adaptive Control.* Addison Wesley, 1989.

[4] Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5):494–500, 1974.

[5] A. Bozin and M. Zarrop. Self tuning optimizer — convergence and robustness properties. In *Proc. 1st European Control Conf.*, pages 672–677, Grenoble, July 1991.

[6] K. Chaloner. Optimal bayesian experimental design for linear models. *Annals of Statistics*, 12(1):283–300, 1984.

[7] K. Chaloner. Bayesian design for estimating the turning point of a quadratic regression. *Commun. Statist.-Theory Meth.*, 18(4):1385–1400, 1989.

[8] K. Chaloner and K. Larntz. Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208, 1989.

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

[10] H. Dette. A note on Bayesian c- and D-optimal designs in nonlinear regression models. *Annals of Statistics*, 24(3):1225–1234, 1996.

[11] V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

[12] V. Fedorov and W. Müller. Another view on optimal design for estimating the point of extremum in quadratic regression. *Metrika*, 46:147–157, 1997.

[13] I. Ford and S. Silvey. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika*, 67(2):381–388, 1980.

[14] J. Ginebra and M. Clayton. Response surface bandits. *Journal of Royal Statistical Society*, B57(4):771–784, 1995.

[15] I. Hu. Strong consistency of Bayes estimates in stochastic regression models. *Journal of Multivariate Analysis*, 57:215–227, 1996.

[16] C. Kulcsár, L. Pronzato, and E. Walter. Dual control of linearly parameterised models via prediction of posterior densities. *European J. of Control*, 2(1):135–143, 1996.

[17] P. Kumar. Convergence of adaptive control schemes using least-squares parameter estimates. *IEEE Transactions on Automatic Control*, 35(4):416–424, 1990.

[18] T. Lai and C. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, 10(1):154–166, 1982.

[19] T. Lai and C. Wei. Asymptotically efficient self-tuning regulators. *SIAM J. Control and Optimization*, 25(2):466–481, 1987.

[20] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22(4):551–575, 1977.

[21] W. Müller and B. Pötscher. Batch sequential design for a nonlinear estimation problem. In V. Fedorov, W. Müller, and I. Vuchkov, editors, *Model-Oriented Data Analysis II, Proceedings 2nd IIASA Workshop, St Kyrik (Bulgaria), May 1990*, pages 76–87. Physica Verlag, Heidelberg, 1992.

[22] L. Pronzato. Adaptive optimisation and D-optimum experimental design. *Annals of Statistics*, 28(6):1743–1761, 2000.

[23] L. Pronzato, C. Kulcsár, and E. Walter. An actively adaptive control policy for linear models. *IEEE Transactions on Automatic Control*, 41(6):855–858, 1996.

[24] L. Pronzato and E. Walter. Experimental design for estimating the optimum point in a response surface. *Acta Applicandae Mathematicae*, 33:45–68, 1993.

[25] J. Sternby. On consistency for the method of least squares using martingale theory. *IEEE Transactions on Automatic Control*, 22(3):346–352, 1977.

[26] P. Wellstead and M. Zarrop. *Self–Tuning Sytems*. Wiley, Chichester, 1991.

[27] C. Wu. Optimal design for percentile estimation of a quantal response curve. In Y. Dodge, V. V. Fedorov, and H. P. Wynn, editors, *Optimal Design and Analysis of Experiments*, pages 213–223. North Holland, Amsterdam, 1988.

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

[29] H. Wynn. The sequential generation of D-optimum experimental designs. *Annals of Math. Stat.*, 41:1655–1664, 1970.