Asymptotic properties of adaptive penalized optimal designs over a finite space

Luc Pronzato

Asymptotic properties of adaptive penalized optimal designs over a finite space

L. Pronzato

Abstract Adaptive optimal design with a cost constraint is considered, both for LS estimation in nonlinear regression and ML estimation in Bernoulli-type experiments, with possible applications in clinical trials. We obtain the strong consistency of the estimators for designs over a finite space, both when the cost level is fixed (and the adaptive design converges to an optimum constrained design) and when the objective is to minimize the cost. Moreover, the asymptotic normality of the estimators is obtained in the first situation, with an asymptotic covariance matrix given by the inverse of the usual information matrix, calculated as if the design were not constructed sequentially.

1 Introduction

Let $\mathscr{X}$, a compact subset of $\mathbb{R}^d$, denote the admissible domain for the experimental variables x (design points) and $\theta \in \Theta$, a compact subset of $\mathbb{R}^p$, denote the p-dimensional vector of parameters, all of interest, in a parametric model with independent observations $Y_i(x_i)$ conditionally on the x_i, $i = 1, 2\ldots$ The information matrix for parameters θ and design measure ξ (a probability measure on $\mathscr{X}$) is denoted by $\mathbf{M}(\xi,\theta) = \int_{\mathscr{X}} \mu(x,\theta)\,\xi(dx)$, with $\mu(x,\theta)$ the contribution of the design point x. We only consider the case of scalar observations, so that $\mu(x,\theta)$ is a rank-one matrix, which we denote $\mu(x,\theta) = \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)$ with $\mathbf{f}_\theta(x)$ a p-dimensional vector. We shall suppose that $\mathbf{f}_\theta(x)$ is continuously differentiable with respect to θ in the interior of Θ for all $x \in \mathscr{X}$. In a nonlinear situation, $\mathbf{M}(\xi,\theta)$ depends on θ and locally optimal design maximizes a concave function $\Psi(\cdot)$ of $\mathbf{M}(\xi,\theta)$ for some nominal value of θ. Here we shall only consider D-optimal design, i.e. $\Psi(\mathbf{M}) = \log\det(\mathbf{M})$,

Laboratoire I3S, CNRS/Université de Nice–Sophia Antipolis, France, e-mail: pronzato@i3s.unice.fr

but the extension to other global optimality criteria, such as $[\mathrm{trace}(\mathbf{M}^{-1})]^{-1}$ for instance, could be obtained by following a similar route. A rather common approach to overcome the difficulty caused by the dependence of a locally optimal design on the unknown value of the model parameters is to design the experiment sequentially.

In fully-adaptive D-optimal design, the next design point x_{n+1} after n observations is taken as

$$x_{n+1} = \arg\max_{x\in\mathscr{X}} \mathrm{trace}[\mu(x,\hat{\theta}^n)\mathbf{M}^{-1}(\xi_n,\hat{\theta}^n)]\,, \tag{1}$$

where $\hat{\theta}^n \in \Theta$ is the current estimated value for θ, based on $x_1,\ldots,x_n$ and the associated observations $Y_1,\ldots,Y_n$, and $\xi_n = (1/n)\sum_{i=1}^n \delta_{x_i}$ is the current empirical design measure. We leave aside initialisation issues and simply assume that $x_1,\ldots,x_p$ are such that $\mathbf{M}(\xi_p,\theta)$ is nonsingular for any $\theta \in \Theta$.

When $\hat{\theta}^n$ is frozen to a fixed value θ, the iteration (1) corresponds to one step of a steepest-ascent vertex-direction algorithm with step-length $1/n$ and convergence to a D-optimal design measure is proved in (Wynn 1970). The fact that $\hat{\theta}^n$ is estimated in adaptive design creates dependency among observations and makes the investigation of the asymptotic behavior of the design and estimator a much more complicated issue for which few results are available: (Ford and Silvey 1980, Wu 1985, Müller and Pötscher 1992) concern a particular example with LS estimation; (Hu 1998) is specific of Bayesian estimation by posterior mean and does not use a fully sequential design of the form (1); Lai (1994) and Chaudhuri and Mykland (1995) require the introduction of a subsequence of non-adaptive design points to ensure consistency of the estimator and Chaudhuri and Mykland (1993) require that the size of the initial experiment (non-adaptive) grows with the increase in size of the total experiment. Notice that the situation is different in clinical trials for comparing treatments: the designs considered are typically such that the number of allocations of each treatment goes to infinity a.s., which then yields the strong consistency of the ML estimators, see for instance the ML design in (Antognini and Giovagnoli 2005). It is shown in (Pronzato 2009b) that the situation becomes much simpler when $\mathscr{X}$ is a finite set and that, under reasonable assumptions, (1) yields the a.s. convergence and asymptotic normality of the estimator $\hat{\theta}^n$. Using the results in (Pronzato 2009a), we show here that similar asymptotic properties are obtained for adaptive penalized D-optimal design. We shall always assume that

$$\mathscr{X} = \{x^{(1)}, x^{(2)}, \ldots, x^{(K)}\}\,,\; K < \infty\,.$$

2 Asymptotic properties of estimators when $\mathscr{X}$ is finite

Consider a nonlinear regression model with observations

$$Y_i = Y(x_i) = \eta(x_i,\bar{\theta}) + \varepsilon_i\,, \tag{2}$$

where the ε_i are i.i.d. with zero mean and finite variance (which we take equal to one without any loss of generality) and $\eta(x,\theta)$ is a known function of θ and x. We suppose that $\bar{\theta}$, the unknown 'true' value of θ, is in the interior of Θ. We have $\mu(x,\theta) = \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)$ with $\mathbf{f}_\theta(x) = \partial\eta(x,\theta)/\partial\theta$. The LS estimator $\hat{\theta}_{LS}^n$ minimizes $S_n(\theta) = \sum_{k=1}^n [Y(x_k) - \eta(x_k,\theta)]^2$ and we define

$$D_n(\theta,\theta') = \sum_{i=1}^n [\eta(x_i,\theta) - \eta(x_i,\theta')]^2 \,. \tag{3}$$

The following properties, see (Pronzato 2009a), will be used in §3.2 and 3.3.

Theorem 1. *Suppose that* $\mathscr{X}$ *is finite. If* $D_n(\theta,\bar{\theta})$ *given by (3) satisfies for all* $\delta > 0$, $\left[\inf_{\|\theta-\bar{\theta}\|\geq\delta/\tau_n} D_n(\theta,\bar{\theta})\right] / (\log\log n) \overset{\text{a.s.}}{\to} \infty$ $(n \to \infty)$, *with* $\{\tau_n\}$ *a nondecreasing sequence of positive deterministic constants, then* $\tau_n \|\hat{\theta}_{LS}^n - \bar{\theta}\| \overset{\text{a.s.}}{\to} 0$ *as* $n \to \infty$.

Theorem 2. *Suppose that* $\mathscr{X}$ *is finite and that there exist non-random symmetric positive definite* $p \times p$ *matrices* $\mathbf{C}_n$ *such that* $\mathbf{C}_n^{-1}\mathbf{M}^{1/2}(\xi_n,\bar{\theta}) \overset{\text{p}}{\to} \mathbf{I}$, *with* $\mathbf{I}$ *the* p-*dimensional identity matrix. If* $c_n = \lambda_{\min}(\mathbf{C}_n)$ *and* $D_n(\theta,\bar{\theta})$ *satisfy* $n^{1/4}c_n \to \infty$ *and for all* $\delta > 0$, $\inf_{\|\theta-\bar{\theta}\|\geq c_n^2\delta} D_n(\theta,\bar{\theta})/(\log\log n) \overset{\text{a.s.}}{\to} \infty$ $(n \to \infty)$, *then* $\hat{\theta}_{LS}^n$ *satisfies* $\sqrt{n}\mathbf{M}^{1/2}(\xi_n,\hat{\theta}_{LS}^n)(\hat{\theta}_{LS}^n - \bar{\theta}) \overset{\text{d}}{\to} \omega \sim \mathscr{N}(\mathbf{0},\sigma^2\mathbf{I})$ *as* $n \to \infty$.

Consider now the case of dose-response experiments with

$$Y \in \{0,1\}\,, \quad \text{with } \Pr\{Y = 1|x_i,\theta\} = \eta(x_i,\theta)\,. \tag{4}$$

Suppose that the 'true' value of θ that generates the observations lies in the interior of Θ, that $\eta(x,\theta) \in (0,1)$ for any $\theta \in \Theta$ and $x \in \mathscr{X}$, and that when n observations $Y_1,\ldots,Y_n$ are performed at the design points $x_1,\ldots,x_n$, the Y_i are independent conditionally on the x_i. Also suppose that x_i is a non-random function of $Y_1,\ldots,Y_{i-1}$, $x_1,\ldots,x_{i-1}$ for all i. Theorems 1 and 2 are then also valid for the ML estimator $\hat{\theta}_{ML}^n$ in this model, see (Pronzato 2009a), and, in the rest of the paper, $\tilde{\theta}^n$ will denote indifferently $\hat{\theta}_{LS}^n$ in the model (2) or $\hat{\theta}_{ML}^n$ in (4).

3 Adaptive penalized *D*-optimal design

Consider constrained locally D-optimal design that maximizes $\log\det[\mathbf{M}(\xi,\theta)]$ under a constraint $\Phi(\xi,\theta) \leq C$ on the average cost $\Phi(\xi,\theta) = \int_{\mathscr{X}} \phi(x,\theta)\,\xi(dx)$. We suppose that $\phi(x,\theta)$, the cost induced by a single observation at x, is a positive continuous function of θ for all $x \in \mathscr{X}$. The extension to nonlinear or multiple constraints is considered, e.g., in (Cook and Fedorov 1995) and (Fedorov and Hackl 1997, Chap. 4). A necessary and sufficient condition for the optimality of ξ_C^* satisfying $\Phi(\xi_C^*,\theta) \leq C$ is the existence of a Lagrange coefficient $\lambda^* = \lambda^*(\theta) \geq 0$ satisfying $\lambda^*[C - \Phi(\xi_C^*,\theta)] = 0$ and $\forall x \in \mathscr{X}$, $\text{trace}[\mu(x,\theta)\mathbf{M}^{-1}(\xi_C^*,\theta)] \leq p + \lambda^*[\phi(x,\theta) - \Phi(\xi_C^*,\theta)]$. In practice, ξ_C^* can be determined by maximizing

$$H_\theta(\xi,\lambda_i) = \log\det[\mathbf{M}(\xi,\theta)] - \lambda_i\,\Phi(\xi,\theta) \tag{5}$$

for an increasing sequence $\{\lambda_i\}$ of coefficients, starting at $\lambda_0 = 0$ and stopping at the first λ_i such that the associated optimal design ξ^* satisfies $\Phi(\xi^*,\theta) \leq C$, see, e.g., Mikulecká (1983) (notice that for C large enough the unconstrained D-optimal design is optimal for the constrained problem). The coefficient λ in $H_\theta(\xi,\lambda)$ can thus be considered as a penalty coefficient that penalizes costly experiments and sets the tradeoff between the maximization of $\log\det[\mathbf{M}(\xi,\theta)]$ and the minimization of $\Phi(\xi,\theta)$. One may refer to Cook and Wong (1994) for the equivalence between constrained and compound optimal designs.

In adaptive constrained D-optimal design, we take x_{n+1} that gives the steepest ascent direction for $H_{\hat{\theta}^n}(\xi_n,\lambda_n)$,

$$x_{n+1} = \arg\max_{x\in\mathscr{X}} \left\{\text{trace}[\mu(x,\hat{\theta}^n)\mathbf{M}^{-1}(\xi_n,\hat{\theta}^n)] - \lambda_n\,\phi(x,\hat{\theta}^n)\right\}\,, \tag{6}$$

where different choices for λ_n are discussed below. Since (1) can be considered as a special case of (6), the results to be presented also cover the case of classical (unconstrained) adaptive D-optimal design (1) treated in (Pronzato 2009b) (they therefore also cover the case of the adaptive penalized designs considered in (Dragalin and Fedorov 2006, Dragalin, Fedorov, and Wu 2008), where the constrained problem is formulated as a standard D-optimal design problem). One may notice the similarity between (6) and the construction used in (Pronzato 2000) to optimize a parametric function, the parameters of which being estimated by least-squares in a linear regression model.

Two situations will be considered concerning the choice of the sequence $\{\lambda_n\}$ in (6), respectively in §3.2 and 3.3. In the first one, the objective is to obtain an optimal design with a specified cost: we adapt λ_n to $\hat{\theta}^n$ and take $\lambda_n = \lambda^*(\hat{\theta}^n)$, the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters $\hat{\theta}^n$. The second situation corresponds to the case where $\{\lambda_n\}$ forms an increasing sequence, which gives more and more importance to the constraint in the construction of the design. When $\phi(x,\theta)$ has a single minimum, by letting the Lagrange coefficient λ_n increase with n one may hope to be able to force the design to concentrate at the minimizer of ϕ associated with the true value of θ. In clinical trials, when $\phi(x,\theta)$ is related to the probability of success of treatment x, it means that we can focuss more and more on individual ethics by allocating treatments with increasing efficacy, see (Pronzato 2010).

3.1 A bound on the sampling rate of nonsingular designs

The key idea used below for investigating the asymptotic properties of an estimator for a design generated by (6) is to suppose first that $\{\hat{\theta}^n\}$ is *an arbitrary sequence* in Θ. We shall use the following assumptions on the design space $\mathscr{X}$, the vectors $\mathbf{f}_\theta(x)$ and the Lagrange coefficients λ_n.

$\mathbf{H}_{\mathscr{X}}$-(i): $\inf_{\theta\in\Theta}\lambda_{\min}\left[\sum_{i=1}^{K}\mathbf{f}_\theta(x^{(i)})\mathbf{f}_\theta^\top(x^{(i)})\right] > \gamma > 0$;
$\mathbf{H}_\lambda$-(i): $0 \leq \lambda_n < \bar{\lambda} < \infty,\ \forall n$;
$\mathbf{H}_\lambda$-(ii): $\{\lambda_n\}$ is a non-decreasing positive sequence and $\lim_{n\to\infty}\lambda_n = \infty$.

Theorem 3. *Let $\{\hat{\theta}^n\}$ be an arbitrary sequence in Θ used to generate design points according to (6) in a finite design space satisfying $H_{\mathscr{X}}$-(i), with an initialisation such that $\mathbf{M}(\xi_n,\theta)$ is non-singular for all θ in Θ and all $n \geq p$. Let $r_{n,i} = r_n(x^{(i)})$ denote the number of times $x^{(i)}$ appears in the sequence $x_1,\ldots,x_n$, $i = 1,\ldots,K$, and consider the associated order statistics $r_{n,1:K} \geq r_{n,2:K} \geq \cdots \geq r_{n,K:K}$. Define*

$$q^* = \max\{j : \textit{ there exists } \alpha > 0 \textit{ such that } \liminf_{n\to\infty} r_{n,j:K}/n > \alpha\},$$
$$q^{**} = \max\{j : \textit{ there exists } \alpha > 0 \textit{ such that } \liminf_{n\to\infty}\lambda_n\, r_{n,j:K}/n > \alpha\}.$$

Then H_λ-(i) implies $q^* \geq p$ and H_λ-(ii) implies $q^{**} \geq p$. When the sequence $\{\hat{\theta}^n\}$ is random, the statement holds with probability one.

The proof is similar to that of Lemma 2 in (Pronzato 2009b). $\mathscr{X}$ finite implies that q^* and $q^{**} > 1$. Supposing that $p \geq 2$, we show that assuming q^* or $q^{**} < p$ leads to a contradiction under H_λ-(i) or H_λ-(ii) respectively.

3.2 λ_n is bounded in (6)

When λ_n is bounded, for any sequence $\{\hat{\theta}^n\}$ used in (6) the conditions of Th. 3 ensure the existence of n_1 and $\alpha > 0$ such that $r_{n,j:K} > \alpha n$ for all $n > n_1$ and all $j = 1,\ldots,p$. Under the additional assumption

$\mathbf{H}_{\mathscr{X}}$-(ii): For all $\delta > 0$ there exists $\varepsilon(\delta) > 0$ such that for any subset $\{i_1,\ldots,i_p\}$ of distinct elements of $\{1,\ldots,K\}$, $\inf_{\|\theta-\bar{\theta}\|\geq\delta}\sum_{j=1}^{p}[\eta(x^{(i_j)},\theta)-\eta(x^{(i_j)},\bar{\theta})]^2 > \varepsilon(\delta)$;

we thus obtain that $D_n(\theta,\bar{\theta})$ given by (3) satisfies $\inf_{\|\theta-\bar{\theta}\|\geq\delta} D_n(\theta,\bar{\theta}) > \alpha n\varepsilon(\delta)$, $n > n_1$. Therefore, $\tilde{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ $(n\to\infty)$ from Th. 1, with $\tilde{\theta}^n = \hat{\theta}^n_{LS}$ in (2) or $\hat{\theta}^n_{ML}$ in (4). Since this holds for any sequence $\{\hat{\theta}^n\}$ in Θ, it is true in particular when $\tilde{\theta}^n$ is substituted for $\hat{\theta}^n$ in (6). One can take in particular $\lambda_n = \lambda^*(\tilde{\theta}^n)$, with $\lambda^*(\theta)$ the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters θ. The following condition then guarantees that H_λ-(i) is satisfied so that Th. 3 applies and $\tilde{\theta}^n$ is strongly consistent from Th. 1.

$\mathbf{H}_\lambda$-(i'): There exists $C' < C$ such that $\forall\theta\in\Theta$, $\exists\hat{\xi}(\theta)\in\Xi$ with $\Phi[\hat{\xi}(\theta),\theta] \leq C'$ and $\mathbf{M}[\hat{\xi}(\theta),\theta]$ has full rank.

Making the following additional assumption on $\mathscr{X}$

$\mathbf{H}_{\mathscr{X}}$-(iii): $\lambda_{\min}\left[\sum_{j=1}^{p}\mathbf{f}_{\bar{\theta}}(x^{(i_j)})\mathbf{f}_{\bar{\theta}}^\top(x^{(i_j)})\right] \geq \bar{\gamma} > 0$ for any subset $\{i_1,\ldots,i_p\}$ of distinct elements of $\{1,\ldots,K\}$, we then obtain the following concerning the convergence of $\mathbf{M}(\xi_n, \tilde{\theta}^n)$.

Theorem 4. *Suppose that the design points for $n > p$ are generated sequentially according to (6) with $\lambda_n = \lambda^*(\tilde{\theta}^n)$ and $\hat{\theta}^n = \tilde{\theta}^n$, the LS-estimator $\hat{\theta}^n_{LS}$ in (2) or the ML-estimator $\hat{\theta}^n_{ML}$ in (4). Suppose, moreover, that the first p design points are such that the information matrix is nonsingular for any $\theta \in \Theta$. Then, under $H_{\mathscr{X}}$-(i-iii) and H_λ-(i') we have $\tilde{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and $\mathbf{M}(\xi_n, \tilde{\theta}^n) \stackrel{\text{a.s.}}{\to} \mathbf{M}[\xi^*(\bar{\theta}), \bar{\theta}]$, $n \to \infty$, with $\xi^*(\bar{\theta})$ a constrained D-optimal design for $\bar{\theta}$.*

From Th. 4 we can take $\mathbf{C}_n = \mathbf{M}^{1/2}[\xi^*(\bar{\theta}), \bar{\theta}]$ in Th. 2 and obtain the usual asymptotic normality of $\tilde{\theta}^n$ for the adaptive design (6) (although the sequential construction of the design implies that $\mathbf{M}(\xi_n, \theta)$ is not the information matrix for parameters θ).

3.3 λ_n tends to infinity in (6)

For any sequence $\{\hat{\theta}^n\}$ used in (6), the conditions of Th. 3 ensure the existence of n_1 and $\alpha > 0$ such that $r_{n,j:K} > \alpha n/\lambda_n$ for all $n > n_1$ and all $j = 1, \ldots, p$. $H_{\mathscr{X}}$-(ii) then implies that $D_n(\theta, \bar{\theta})$ given by (3) satisfies $\left[\inf_{\|\theta-\bar{\theta}\| \geq \delta} D_n(\theta, \bar{\theta})\right] / (\log\log n) > \alpha n \varepsilon(\delta)/[\lambda_n (\log\log n)]$ for $n > n_1$. Therefore, if $\lambda_n(\log\log n)/n \to 0$ as $n \to \infty$, $\tilde{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ from Th. 1. Since this holds for any sequence $\{\hat{\theta}^n\}$ in Θ, it is true in particular when $\tilde{\theta}^n$ is substituted for $\hat{\theta}^n$ in (6). (One may notice that Th. 1 provides some indication about the rate of convergence of $\tilde{\theta}^n$ towards $\bar{\theta}$: for $\|\theta - \bar{\theta}\| = \delta$ small enough, $D_n(\theta, \bar{\theta})/n \approx (\theta - \bar{\theta})^\top \mathbf{M}(\xi_n, \bar{\theta})(\theta - \bar{\theta})$, which is larger than $\alpha\bar{\gamma}\delta^2/\lambda_n$ from $H_{\mathscr{X}}$-(iii); therefore, $\|\tilde{\theta}^n - \bar{\theta}\| = O(\sqrt{\lambda_n(\log\log n)}/\sqrt{n})$ a.s.) Next theorem indicates that when the following is satisfied in addition to H_λ-(ii):

H_λ-(iii): λ_n/n is non-increasing and $\lambda_n(\log\log n)/n \to 0$, $n \to \infty$;
H_ϕ: $\phi(x, \bar{\theta})$ has a unique global minimizer in $\mathscr{X}$: $\phi(x^{(i^*)}, \bar{\theta}) = \min_{x \in \mathscr{X}} \phi(x, \bar{\theta}) < \phi(x^{(i)}, \bar{\theta})$, $\forall i \in \{1, \ldots, K\}$, $i \neq i^*$;

then $\{x_n\}$ tends to accumulate at the point of minimum cost for $\bar{\theta}$.

Theorem 5. *Suppose that the design points for $n > p$ are generated sequentially according to (6), where λ_n satisfies H_λ-(ii) and H_λ-(iii). Suppose, moreover, that the first p design points are such that the information matrix is nonsingular for any $\theta \in \Theta$. Then, under $H_{\mathscr{X}}$-(i-iii) we have $\tilde{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and*

$$\Phi(\xi_n, \bar{\theta}) \stackrel{\text{a.s.}}{\to} \phi^*_{\bar{\theta}} = \min_{x \in \mathscr{X}} \phi(x, \bar{\theta}) \,,\; n \to \infty \,.$$

If, moreover, H_ϕ is satisfied, then $\xi_n(x^{(i)}) \stackrel{\text{a.s.}}{\to} 0$ for all $i \neq i^*$.

Example. Suppose that $\eta(x, \theta) = [\theta_1/(\theta_1 - \theta_2)]\,[\exp(-\theta_2 x) - \exp(-\theta_1 x)]$ in the model (2) with i.i.d. errors $\mathscr{N}(0,1)$. The objective is to maximize $\eta(x, \bar{\theta})$ for $x \in \mathscr{X}$

consisting of 1001 points regularly spaced in $[0,10]$. We take $\phi(x,\theta) = -\eta(x,\theta)$ and $\bar{\theta} = (0.7,\ 0.2)^\top$, so that $\eta(x,\bar{\theta})$ reaches its maximum value in $\mathscr{X}$ (approximately 0.606, indicated by a dashed line on Fig. 1) at $x^* = 2.51$. The design points are generated by (6) for $n \geq 2$, with $\hat{\theta}^n$ the LS estimator and $x_1 = 1.25$, $x_2 = 6.6$. Three sequences are considered for $\{\lambda_n\}$: $\lambda_n^{(a)} = \log^2 n$, $\lambda_n^{(b)} = n/(1+\log^2 n)$ and $\lambda_n^{(c)} = n^{1.1}$, $n \leq 1000$ (notice that $\lambda^{(b)} < \lambda^{(a)}$ on the horizon considered). Th. 5 is satisfied for $\lambda_n^{(a)}$ and $\lambda^{(b)}$, but $\lambda_n^{(c)}$ increases too fast and does not insure convergence of ξ_n to the delta measure at x^*, see Fig. 1 for a typical realization. Of course, the behavior is even worse for the "best intention design" (also called "forced certainty equivalence" in the control literature) $x_{k+1} = \arg\min_{x\in\mathscr{X}} \phi(x,\hat{\theta}^k)$.

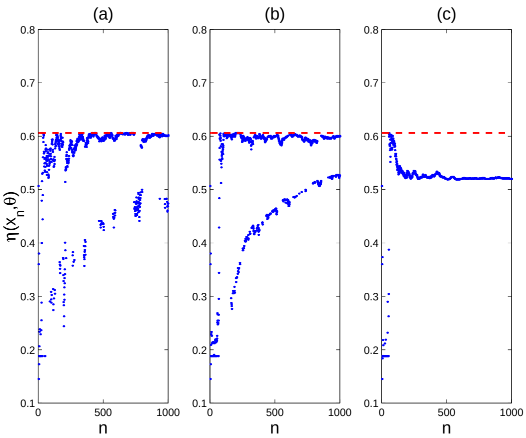

Fig. 1 Evolution of $\eta(x_n,\bar{\theta})$ as a function of n for three different sequences $\{\lambda_n\}$.

Similar results are obtained when the cost $\phi(x,\bar{\theta})$ to be minimized is not directly related to $\eta(x,\bar{\theta})$. Consider for instance a regulation problem where the objective is to set a function $\varphi(x,\bar{\theta})$ on a given target T, so that one may take $\phi(x,\theta)$ as a measure of the distance between $\varphi(x,\theta)$ and T, e.g., $\phi(x,\theta) = [\varphi(x,\theta) - T]^2$. There, "best intention design" (the "continuous reassessment method" in dose finding), or Robbins-Monro type procedures (see, e.g., Lai and Robbins (1978)) can be used when $\varphi(x,\theta) = \eta(x,\theta)$. The adaptive design (6) may be convenient in more general circumstances where the function $\varphi(x,\theta)$ to be regulated differs from $\eta(x,\theta)$.

References

Antognini, A. B. and A. Giovagnoli (2005). On the large sample optimality of sequential designs for comparing two or more treatments. *Sequential Analysis 24*, 205–217.

Chaudhuri, P. and P. Mykland (1993). Nonlinear experiments: optimal design and inference based likelihood. *Journal of the American Statistical Association 88*(422), 538–546.

Chaudhuri, P. and P. Mykland (1995). On efficiently designing of nonlinear experiments. *Statistica Sinica 5*, 421–440.

Cook, D. and V. Fedorov (1995). Constrained optimization of experimental design (invited discussion paper). *Statistics 26*, 129–178.

Cook, D. and W. Wong (1994). On the equivalence between constrained and compound optimal designs. *Journal of the American Statistical Association 89*(426), 687–692.

Dragalin, V. and V. Fedorov (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference 136*, 1800–1823.

Dragalin, V., V. Fedorov, and Y. Wu (2008). Adaptive designs for selecting drug combinations based on efficacy-toxicity response. *Journal of Statistical Planning and Inference 138*, 352–373.

Fedorov, V. and P. Hackl (1997). *Model-Oriented Design of Experiments*. Berlin: Springer.

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

Hu, I. (1998). On sequential designs in nonlinear problems. *Biometrika 85*(2), 496–503.

Lai, T. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Annals of Statistics 22*(4), 1917–1930.

Lai, T. and H. Robbins (1978). Adaptive design in regression and control. *Proc. Nat. Acad. Sci. USA 75*(2), 586–587.

Mikulecká, J. (1983). On a hybrid experimental design. *Kybernetika 19*(1), 1–14.

Müller, W. and B. Pötscher (1992). Batch sequential design for a nonlinear estimation problem. In V. Fedorov, W. Müller, and I. Vuchkov (Eds.), *Model-Oriented Data Analysis II, Proceedings 2nd IIASA Workshop, St Kyrik (Bulgaria), May 1990*, pp. 77–87. Heidelberg: Physica Verlag.

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

Pronzato, L. (2009a). Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statistics & Probability Letters 79*, 2307–2313. DOI: 10.1016/j.spl.2009.07.025.

Pronzato, L. (2009b). One-step ahead adaptive *D*-optimal design on a finite design space is asymptotically optimal. *Metrika*. (to appear, DOI: 10.1007/s00184-008-0227-y).

Pronzato, L. (2010). Penalized optimal designs for dose-finding. *Journal of Statistical Planning and Inference 140*, 283–296. DOI: 10.1016/j.jspi.2009.07.012.

Wu, C. (1985). Asymptotic inference from sequential design in a nonlinear situation. *Biometrika 72*(3), 553–558.

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