Experimental Design for Processes Over Time. How Many Time Series?

Dariusz Uciński*
Institute of Control and Computation Engineering
University of Zielona Góra, ul. Podgórna 50
65–246 Zielona Góra, Poland
and
Anthony C. Atkinson†
Department of Statistics, London School of Economics
London WC2A 2AE, UK

June 20, 2005

Abstract

We develop optimum design theory for regression models when the observations have a correlation structure, but when different realizations of the process are independent. We give an algorithm for finding optimum designs when the number of observations per process is specified, but the number of realizations to be observed is left free. We find that, for appreciable correlation within a process, the optimum design consists of two different observational sequences, the relative frequencies of the two sequences depending on the correlation. Our algorithm has been implemented to be used when the model consists of a set of differential equations for which only numerical solutions are available. The numerical example illustrates this feature.

Keywords: Correlated errors; D-optimum design; Design algorithm; Global optimization; Nonlinear models; Parameter sensitivities; Several series

*e-mail: D.Ucinski@issi.uz.zgora.pl
†e-mail: A.C.Atkinson@lse.ac.uk

1 Introduction

This paper continues our work on the optimum design of experiments for regression models when the errors have a correlation structure. In Uciński and Atkinson (2004) we used D-optimality to find the best set of times at which to take measurements for estimation of the parameters of a simple nonlinear model. In that work the number of time points, q, was specified. If resources allowed taking Nq measurements, then we assumed that the series of q observations would be repeated on N realizations of the process. The important extension in this paper is to the possibility of having different series of time points for the different realizations of the series. We do however restrict attention to designs in which exactly q measurements are made on each realization.

The structure of the paper is as follows. The next section introduces the example that we use for our numerical work and discusses related models in medicine, where the realizations of the process are individual patients for each of whom a measurement schedule is required. In §3 and §4 we present the statistical models for one realization of the correlated process and for multiple independent realizations of correlated processes. Next, §5 characterises optimum designs for independent observations. In §6 we apply the criterion to independent realizations of correlated processes using an algorithm incorporating a global optimizer. The paper concludes in §7 with a numerical example in which observations on either one or two series are optimum depending on the strength of the correlation between observations.

2 Model of a general consecutive reaction

2.1 The chemical model

Our example is of a simple chemical reaction that, however, illustrates the complexity of our design problem. In the model for two consecutive reactions

$$A \stackrel{\theta_1}{\rightarrow} B \stackrel{\theta_2}{\rightarrow} C, \tag{1}$$

we start, at $t = 0$, with known concentrations $[A], [B]$ and $[C]$ of the three reactants. The change over time of these concentrations is governed by the

kinetic differential equations

$$\begin{aligned}\frac{\mathrm{d}[A]}{\mathrm{d}t} &= -\theta_1[A]^{\lambda_1},\\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= \theta_1[A]^{\lambda_1} - \theta_2[B]^{\lambda_2},\\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= \theta_2[B]^{\lambda_2},\end{aligned} \tag{2}$$

where θ_1 and θ_2 are the rates of reaction and λ_1 and λ_2 are the orders of the reactions. The purpose of the experiment is to obtain good estimates of the parameters θ, and perhaps also λ, from measurements of some or all of the concentrations at a variety of time points. If we assume that conditions, such as temperature, remain constant then designing the experiment reduces to the choice of sets of times at which measurements of the concentrations are to be taken and of the proportions of the sets of times.

Compartmental models similar to (2) are used in medicine, for example to model the passage of drugs through patients. The models may be more complicated than that in our example in having several interconnected components, but the reactions are customarily assumed to be first-order, so that the time profiles of concentrations are weighted sums of exponential terms. The experimental design problem is again that of choosing times at which measurements are to be made. If patients are to return to a clinic for observation, there will be a restricted set of values of time at which measurements can be taken. In addition, the number of measurements per patient, q, may be small. Although we do not pursue such designs here, it is clear that there is a rich field of problems to be explored. Instead we consider measurements chosen over a time continuum.

3 Statistical model for one run of the process

Consider a general multi-response regression model

$$y(t) = \eta(t, \beta^\star) + \varepsilon(t), \quad t \in T = [0, t_f], \tag{3}$$

where t stands for time, t_f is a fixed finite time horizon, $\beta^\star \in \mathbb{R}^m$ signifies an unknown constant parameter vector, $\eta : \mathbb{R}^{m+1} \to \mathbb{R}^d$ is a known (possibly non-linear) regression function, and $\varepsilon(t)$ signifies that the noise is an $\mathbb{R}^d$-valued stochastic process with zero mean and known continuous covariance kernel

$$\mathrm{E}[\varepsilon(t)\varepsilon^{\mathsf{T}}(s)] = C(t, s) \quad \text{on } T^2, \tag{4}$$

where $C(t,s)$ is a $d \times d$ positive-definite matrix.

We assume that the response can be measured at q time instants $0 \leq t_1 < t_2 < \cdots < t_q \leq t_f$, i.e. we observe

$$y(t_k) = \eta(t_k, \beta^\star) + \varepsilon(t_k), \quad k = 1, \ldots, q. \tag{5}$$

For notational convenience, in what follows we shall write

$$x = (t_1, \ldots, t_q) \tag{6}$$

for the entire time sequence.

The results of these q measurements are to serve as data in determining the following generalized least-squares estimate of the unknown parameter β:

$$\hat{\beta} = \arg\min_{\beta} \mathcal{J}(\beta; x), \tag{7}$$

where

$$\mathcal{J}(\beta; x) = \sum_{k=1}^{q} \sum_{\ell=1}^{q} [y(t_k) - \eta(t_k, \beta)]^{\mathsf{T}} W_{k\ell}(x) [y(t_\ell) - \eta(t_\ell, \beta)]. \tag{8}$$

The weighting matrices $W_{k\ell}(x) \in \mathbb{R}^{d \times d}$ above constitute the submatrices of the $qd \times qd$ matrix

$$W(x) = \begin{bmatrix} W_{11}(x) & \ldots & W_{1q}(x) \\ \vdots & \ddots & \vdots \\ W_{q1}(x) & \ldots & W_{qq}(x) \end{bmatrix} = K^{-1}(x), \tag{9}$$

where

$$K(x) = \begin{bmatrix} C(t_1, t_1) & \ldots & C(t_1, t_q) \\ \vdots & \ddots & \vdots \\ C(t_q, t_1) & \ldots & C(t_q, t_q) \end{bmatrix}. \tag{10}$$

It is clear that the accuracy of the estimates $\hat{\beta}$ depends on the choice of the times t_k, $k = 1, \ldots, q$; measurements at some points of the time interval T may supply greater amounts of information about $\beta^\star$ than others. A crucial question is thus how to select the time sequence (6) which will guarantee the best accuracy of (7). To form a basis for the comparison of different time sequences, a quantitative measure of the 'goodness' of particular sequences

is required. From a statistical point of view, such a measure is most often based on the concept of the Fisher Information Matrix (FIM). In our case, it has the explicit form:

$$\Upsilon(x) = F^{\mathsf{T}}(x)W(x)F(x), \tag{11}$$

where

$$F(x) = \begin{bmatrix} \dfrac{\partial \eta}{\partial \beta}(t_1, \beta^0) \\ \vdots \\ \dfrac{\partial \eta}{\partial \beta}(t_q, \beta^0) \end{bmatrix}, \tag{12}$$

β^0 being a prior estimate of the unknown parameter vector β. The elements of the matrix $F(x)$ are usually called the *sensitivity coefficients*. In what follows we shall assume that $F(\,\cdot\,)$ is continuous in T^q.

The above formulation is generally accepted in optimum experimental design, since the inverse of the FIM constitutes, up to a constant multiplier, the Cramér-Rao lower bound on the covariance matrix of any unbiased estimator of β. When the time horizon is large, the nonlinearity of the model with respect to its parameters is mild and the measurement errors have small magnitudes, it is legitimate to assume that the estimator $\hat{\beta}$ is *efficient* (minimum-variance) in the sense that the parameter covariance matrix achieves the lower bound. This leads to a great simplification since the minimum variance given by the Cramér-Rao lower bound can be relatively easily computed, even though the exact covariance matrix of a particular estimator is very difficult to obtain when the model is nonlinear.

Roughly speaking, a 'larger' value of $\Upsilon(x)$ reflects more precision—lower variability—in the estimate $\hat{\beta}$. Hence a natural goal in selecting x is to introduce the appropriate real-valued function Ψ defined on all possible $\Upsilon(x)$ such that a lower value of $\Psi[\Upsilon(x)]$ will make the matrix $M(x)$ 'large', and then to attempt to find an $x^\star$ which minimizes $\Psi[\Upsilon(\,\cdot\,)]$. The introduction of an optimality criterion renders it possible to formulate the problem of selecting the most suitable sequence of observation times as an optimization problem. Since it is logical to restrict our attention to sequences x for which $K(x)$ is invertible (note that $K^{-1}(x)$ does not exist if some instants among $t_1, \ldots, t_q$ coincide), we assume that an admissible compact set $\mathcal{X} \subset T^q$ has been defined such that all $x \in \mathcal{X}$ guarantee the satisfaction of this condition. For example, we could define it as

$$\mathcal{X} = \big\{ x \in T^q : \det(K(x)) \geq \epsilon \big\} \tag{13}$$

for some small $\epsilon > 0$.

Various choices exist for optimality criteria, including e.g. the following:

- The D-optimality (determinant) criterion

$$\Psi(\Upsilon) = -\ln\det(\Upsilon). \tag{14}$$

- The A-optimality (trace) criterion

$$\Psi(\Upsilon) = \mathrm{trace}(\Upsilon^{-1}). \tag{15}$$

A D-optimum design minimizes the volume of the uncertainty ellipsoid for the estimates whereas an A-optimum design minimizes the average variance of the estimates. The former criterion was extensively studied in Uciński and Atkinson (2004), where an exchange-type algorithm was proposed for numerically determining optimum sampling points.

4 Statistical model for multiple independent process runs

We are now interested in the slightly more complicated situation in which q-point series of measurements can be repeated N times, each on a process run which is independent of the other runs. This means that the random errors of measurements in different runs are independent and the time sequences $x = (t_1, \ldots, t_q)$ may vary from run to run. In mathematical terms, this framework corresponds to the statistical model

$$y_{ij}(t_k^i) = \eta(t_k^i, \beta^\star) + \varepsilon_{ij}(t_k^i), \quad k = 1, \ldots, q; \quad j = 1, \ldots, r_i; \quad i = 1, \ldots, n. \tag{16}$$

Here each time sequence $x_i = (t_1^i, \ldots, t_q^i) \in T^q$ plays the role of a value of the q-dimensional explanatory (or regressor) variable x, and we assume that $x_i \neq x_\ell$ whenever $i \neq \ell$. The additional index j is necessary when we admit of $r_i > 1$ replicated process runs for the same time sequence x_i. Thus the total number of process runs is $\sum_{i=1}^n r_i = N$.

The assumption of independent replications of the time sequences is expressed by the following characterization of the vector measurement errors:

$$\mathrm{E}[\varepsilon_{ij}(t)] = 0, \quad \mathrm{E}[\varepsilon_{ij}(t)\varepsilon_{\kappa\ell}^{\mathsf{T}}(s)] = \sigma^2 C(t,s)\delta_{i\kappa}\delta_{j\ell}, \tag{17}$$

where σ^2 is a positive constant and δ_{ij} is the Kronecker delta.

The corresponding generalized least-squares estimate of β is

$$\hat{\beta} = \arg\min_\beta \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{r_i} \mathcal{J}_{ij}(\beta; x_i), \tag{18}$$

where

$$\mathcal{J}_{ij}(\beta; x_i) = \sum_{k=1}^{q} \sum_{\ell=1}^{q} [y_{ij}(t_k^i) - \eta(t_k^i, \beta)]^{\mathsf{T}} W_{k\ell}(x_i) [y_{ij}(t_\ell^i) - \eta(t_\ell^i, \beta)]. \tag{19}$$

Then the associated FIM has the form

$$M(x_1, \ldots, x_n) = \frac{N}{\sigma^2} \sum_{i=1}^{n} p_i \Upsilon(x_i), \tag{20}$$

which is obtained after setting $p_i = r_i/N$. Note that this is a weighted sum of the information matrices $\Upsilon(x_i)$ for individual series of observations, which is crucial for the approach outlined in what follows.

There is one more helpful simplification which involves no loss of generality. Since all the design criteria used in practice satisfy the condition

$$\Psi(\beta M) = \gamma(\beta)\Psi(M), \quad \beta > 0, \tag{21}$$

γ being a positive function, we may discard the constant multiplier N/σ^2 in (20). For notational convenience, we can also introduce the matrix

$$\xi_N = \begin{Bmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{Bmatrix}, \tag{22}$$

called the *exact design* of the experiment. The proportion p_i of observations performed at x_i can be considered as the percentage of experimental effort spent at that point.

Following these remarks, we rewrite the FIM in the form

$$M(\xi_N) = \sum_{i=1}^{n} p_i \Upsilon(x_i). \tag{23}$$

Here the p_i's are rational numbers, since both r_i's and N are integers. Note that the design (22) defines a discrete probability distribution on its distinct support points, i.e., $x_1, \ldots, x_n$.

The discrete nature of N-measurement exact designs causes serious difficulties, as the resultant numerical analysis problem is not amenable to solution by standard optimization techniques, particularly when N is large. A commonly used device for this problem is to extend the definition of the design (Atkinson and Donev 1992; Silvey 1980; Fedorov and Hackl 1997; Pukelsheim 1993; Walter and Pronzato 1997; Cox and Reid 2000). In such a relaxed formulation, when N is large, the feasible p_i's can be considered

as any real numbers in the interval $[0, 1]$ which sum up to unity, and are not necessarily integer multiples of $1/N$. It is thus easier to operate on the so-called *approximate* designs of the form

$$\xi = \left\{ \begin{matrix} x_1, & x_2, & \ldots, & x_n; \\ p_1, & p_2, & \ldots, & p_n \end{matrix} \quad \sum_{i=1}^{n} p_i = 1 \right\}, \tag{24}$$

where the number of support points n is not fixed and constitutes an additional parameter to be determined, subject to the condition $n < \infty$. The key idea is then to use calculus to find the number of support points n and associated values $x_i^\star$, together with the proportions of measurements assigned to those support points $p_i^\star$ where the minimum of a suitable performance index occurs, and argue that the minimum of the same function over the integer multiples of $1/N$ will occur at the adjacents of $p_1^\star, \ldots, p_n^\star$.

The above reinterpretation of the admissible designs as discrete probability distributions on finite subsets of $\mathcal{X}$ appreciably ameliorates the intractability of the design problem. Nevertheless, there still remain many technicalities which make the resulting calculus rather cumbersome. Paradoxically, these difficulties can easily be overcome by further widening the class of admissible designs to all probability measures ξ over $\mathcal{X}$. Such an extension of the design concept allows us to replace (23) by

$$M(\xi) = \int_{\mathcal{X}} \Upsilon(x)\, \xi(\mathrm{d}x), \tag{25}$$

where the integration in (25) is to be understood in the Lebesgue-Stieltjes sense, This leads to the so-called *continuous* designs which constitute the basis of the modern theory of optimal experiments and originate in seminal works by Kiefer and Wolfowitz, e.g. Kiefer and Wolfowitz (1959). Such an approach drastically simplifies the design and the next section is devoted to this issue.

For clarity, we adopt the following notational conventions: here and subsequently, we will use the symbol $\Xi(\mathcal{X})$ to denote the set of all probability measures on $\mathcal{X}$. Then we may redefine the optimal design as a solution to the optimization problem:

$$\xi^\star = \arg \min_{\xi \in \Xi(\mathcal{X})} \Psi[M(\xi)]. \tag{26}$$

5 Characterization of optimum designs

The following technical assumptions about the design criterion $\Psi : \mathrm{NND}(m) \to \mathbb{R}$ will be needed[1]:

A1. Ψ is convex.

A2. If $M_1 \preceq M_2$ (i.e. $M_1 - M_2 \in \mathrm{NND}(m)$), then $\Psi(M_1) \geq \Psi(M_2)$ (monotonicity).

A3. There exists a finite real B such that

$$\widetilde{\Xi}(\mathcal{X}) = \big\{\xi : \Psi[M(\xi)] \leq B < \infty\big\} \neq \emptyset.$$

A4. The matrix

$$\mathring{\Psi}(\xi) = \left.\frac{\partial \Psi[M]}{\partial M}\right|_{M=M(\xi)}$$

exists for any $\xi \in \widetilde{\Xi}(\mathcal{X})$.

Note that, given $\xi \in \widetilde{\Xi}(\mathcal{X})$ and $\bar{\xi} \in \Xi(\mathcal{X})$, Assumption A3 implies the following expression for the one-sided directional derivative of Ψ:

$$\begin{aligned} \delta\Psi[M(\xi); M(\bar{\xi}) - M(\xi)] &= \left.\frac{\partial \Psi[M(\xi) + \alpha(M(\bar{\xi}) - M(\xi))]}{\partial \alpha}\right|_{\alpha=0^+} \\ &= c(\xi) - \int_{\mathcal{X}} \phi(x, \xi)\, \bar{\xi}(\mathrm{d}x), \end{aligned} \tag{27}$$

where

$$\phi(x, \xi) = -\operatorname{trace}\Big[\mathring{\Psi}(\xi)\Upsilon(x)\Big], \quad c(\xi) = -\operatorname{trace}\Big[\mathring{\Psi}(\xi)M(\xi)\Big] = \int_{\mathcal{X}} \phi(x, \xi)\, \xi(\mathrm{d}x). \tag{28}$$

Using the standard approach in optimum experimental design (Atkinson and Donev 1992; Fedorov and Hackl 1997; Pázman 1986; Uciński 2005), we can prove the following results.

Theorem 1.

(i) An optimal design exists which is purely discrete (i.e. it is of the form (24)) and comprises $n \leq m(m+1)/2$ support points.

(ii) The set of optimal designs is convex.

[1]Here $\mathrm{NND}(m)$ stands for the set of nonnegative definite $m \times m$ matrices. Similarly, $\mathrm{PD}(m)$ will denote the set of positive definite $m \times m$ matrices.

(iii) For any purely discrete optimal design $\xi^\star$, the function $\phi(\,\cdot\,,\xi^\star)$ attains its global maximum in $\mathcal{X}$ at all support points corresponding to nonzero weights.

Note that, in contrast to classical results for one-response models, the lower bound on the number of support points is now unity, and not m. This is because the matrix terms in the weighted sum (23) are non-singular provided that the matrices $F(x)$ have full column rank.

Theorem 2 (General Equivalence Theorem)**.**

The following characterizations of an optimal design $\xi^\star$ are equivalent in the sense that each implies the other two:

(i) the design $\xi^\star$ minimizes $\Psi[M(\xi)]$,

(ii) the design $\xi^\star$ minimizes $\max\limits_{x\in\mathcal{X}} \phi(x,\xi) - c(\xi)$, and

(iii) $\max\limits_{x\in\mathcal{X}} \phi(x,\xi^\star) = c(\xi^\star)$.

All the designs satisfying (i)–(iii) and their convex combinations have the same information matrix $M(\xi^\star)$ provided that the criterion Ψ is strictly convex on $\widetilde{\mathfrak{M}}(\mathcal{X}) = \big\{M(\xi) : \xi \in \widetilde{\Xi}(\mathcal{X})\big\}$.

For the most popular D-optimality criterion, the above result specializes to our next assertion.

Theorem 3 (Equivalence Theorem for D-Optimum Designs)**.**

The following conditions are equivalent:

(i) the design $\xi^\star$ maximizes $\det(M(\xi))$,

(ii) the design $\xi^\star$ minimizes $\operatorname{trace}\big[M^{-1}(\xi)\Upsilon(x)\big]$, and

(iii) $\max\limits_{x\in\mathcal{X}} \operatorname{trace}\big[M^{-1}(\xi)\Upsilon(x)\big] = m$.

6 Sequential numerical design algorithm

Apart from the characterization of optimum designs, we need efficient numerical algorithms which enable us to construct Ψ-optimum design measures in practice. It is a simple matter to adopt here the general idea of the first-order descent algorithms commonly used in optimum experimental design,

cf. Atkinson and Donev (1992), Fedorov and Hackl (1997), Pázman (1986) or Uciński (2005).

The steps in using the gradient method can be briefly summarized as follows (Fedorov and Hackl 1997; Walter and Pronzato 1997; Ermakov 1983; Rafajłowicz 1996):

Algorithm 1 (Iterative construction of Ψ-optimum designs)**.**

Step 1. *Guess a discrete nondegenerate starting design measure* $\xi^{(0)}$ *(we must have* $\Psi[M(\xi^{(0)})] < \infty$*). Choose some positive tolerance* $\eta \ll 1$*. Set* $s = 0$*.*

Step 2. *Determine* $x_0^{(s)} = \arg\max_{x\in\mathcal{X}} \phi(x, \xi^{(s)})$*. If* $\phi(x_0^{(s)}, \xi^{(s)}) < c(\xi^{(s)}) + \eta$*, then* STOP*.*

Step 3. *For an appropriate value of* $0 < \alpha_s < 1$*, set*

$$\xi^{(s+1)} = (1 - \alpha_s)\xi^{(s)} + \alpha_s \delta(x_0^{(s)}),$$

increment s *by one and go to Step 2.*

Note that $\delta(x_0^{(s)})$ in Step 3 signifies the one-point design concentrated at $x_0^{(s)}$.

It can be shown that the above algorithm converges to an optimal design provided that the sequence $\{\alpha_s\}$ is suitably chosen. For example, the following choices can be considered:

(i) Diminishing stepsize (Wynn's algorithm):

$$\lim_{s\to\infty} \alpha_s = 0, \quad \sum_{s=0}^{\infty} \alpha_s = \infty, \tag{29}$$

(ii) Limited minimization rule (Fedorov's algorithm):

$$\alpha_s = \arg\min_{\alpha} \Psi[(1 - \alpha)M(\xi^{(s)}) + \alpha\Upsilon(x_0^{(s)})]. \tag{30}$$

Computationally, Step 2 is of crucial significance. It is also the most time-consuming step in the algorithm. Complications arise, among other things, due to the necessity of calculating a global maximum of $\phi(\,\cdot\,, \xi^{(s)})$ which is usually multimodal (getting stuck in one of the local maxima leads to premature termination of the algorithm). Therefore, while implementing this part of the computational procedure an effective global optimizer is

essential. Based on numerous computer experiments it was found that the extremely simple adaptive random search (ARS) strategy from Venot, Pronzato, Walter, and Lebruchec (1986, p. 216) or Walter and Pronzato (1997) is especially suited for that purpose. The routine chooses the initial point x^0 in T^q at random. After ℓ iterations, given the current best point x^ℓ, a random displacement vector Δx is generated and the trial point

$$x^+ = \Pi_{T^q}(x^\ell + \Delta x) \tag{31}$$

is checked, where Δx follows a multinormal distribution with zero mean and covariance matrix

$$\mathrm{cov}\{\Delta x\} = \mathrm{diag}[\sigma_1, \ldots, \sigma_q], \tag{32}$$

Π_{T^q} being the projection onto T^q. Note that if $x \in T^q \setminus \mathcal{X}$ happens during computations, then we set $\phi(x, \xi^{(s)}) = -Q$, where Q is a huge number.

If $\phi(x^+, \xi^s) < \phi(x^\ell, \xi^s)$, then x^+ is rejected and consequently we set $x^{\ell+1} = x^\ell$, otherwise x^+ is taken as $x^{\ell+1}$. The adaptive strategy consists in repeatedly alternating two phases. During the first one (variance selection) $\mathrm{cov}\{\Delta x\}$ is selected from among the sequence $\mathrm{diag}({}^1\sigma), \mathrm{diag}({}^2\sigma), \ldots, \mathrm{diag}({}^5\sigma)$, where

$${}^1\sigma = \underbrace{(t_f, \ldots, t_f)}_{q \text{ times}} \tag{33}$$

and

$${}^i\sigma = {}^{(i-1)}\sigma/10, \quad i = 2, \ldots, 5. \tag{34}$$

With such a choice, ${}^1\sigma$ is large enough to allow for an easy exploration of $\mathcal{X}$, whereas ${}^5\sigma$ is small enough for a precise localization of an optimal point. In order to allow a comparison to be drawn, all the possible ${}^i\sigma$'s are used $100/i$ times, starting from the same initial value of x. The largest ${}^i\sigma$'s, designed to escape from local maxima, are therefore used more often than the smaller ones.

During the second (exploration) phase, the most successful ${}^i\sigma$ in terms of the criterion value reached during the variance selection phase is used for 100 random trials started from the best x obtained so far. The variance-selection phase then resumes, unless the decision to stop is taken.

As regards the choice of an optimal α_s in Fedorov's variant of Step 3, it should be emphasized that the situation is a bit different from the well-known case of linear regression considered in classic textbooks for which it is possible to determine a closed-form solution. Since in our case the application of the matrix-inversion lemma by no means simplifies the problem, an optimal α_s

has to be determined numerically, e.g., with the use of the golden-section search.

Furthermore, while implementing the algorithms, numerous additional problems should be addressed. For instance, it may be possible to achieve a markedly greater decrease in the value of Ψ by removing a measure from a point already in the design $\xi^{(s)}$ and distributing the measure removed among the most promising points of support (e.g., those for which $\phi(x^i, \xi^{(s)}) > c(\xi^{(s)})$), and additionally, distributing the removed measure among those points in a manner proportional to $\phi(x^i, \xi^{(s)}) - c(\xi^{(s)})$. In this way, undesirable and noninformative points of support which were included in the initial design can be eliminated.

Another complication is the tendency of the points produced in Step 2 to cluster around support points of the optimal design. This can be avoided by checking whether the newly generated point is close enough to a point of the current support so as to qualify them as coincident points. If so, the latter is replaced by the former with a simultaneous update of the weights of all the points according to the rule of Step 3. In more detail, we could check whether $x_0^{(s)}$ is close enough to a point already in the design, e.g., by determining

$$\tilde{x}^{(s)} = \arg \min_{x \in \operatorname{supp} \xi^{(s)}} \varrho(x), \tag{35}$$

where

$$\varrho(x) = \|\Upsilon(x) - \Upsilon(x_0^{(s)})\|. \tag{36}$$

If $\varrho(\tilde{x}^{(s)}) < \epsilon$, where ϵ is sufficiently small, then $\tilde{x}^{(s)}$ is to be deleted and its weight to be added to that of $x_0^{(s)}$.

As a supplementary improvement, the algorithm should include a cyclic removal of points with negligible weights in order to maintain a relatively small number of support points.

7 Numerical example

We illustrate these ideas on (2) with $\lambda_1 = \lambda_2 = 2$, so that the values of the concentrations of B and C as functions of time are found by numerical solution of the set of differential equations (2). The corresponding sensitivity coefficients (12) are then found by numerical solution of a set of auxiliary differential equations (Valko and Vajda 1984; Atkinson and Bogacka 2002; Uciński 2005). But first we recall some results from the literature obtained when both reactions are first order.

When $\lambda_1 = \lambda_2 = 1$, explicit algebraic solutions can be found for the concentrations as a function of time, given the initial concentrations of A, B

and C. If the initial concentration of A is one and that of B and C are zero, the concentration of B is given by

$$\eta(t, \beta^\star) = \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}), \tag{37}$$

where, in the more general model (2), $\beta^\star$ is an unknown vector of parameters including θ and λ. Initially $\eta(t, \beta^\star)$ is zero, rising to a maximum value and then smoothly declining. The curves of sensitivities, in this there are two, can then be found by analytical differentiation of (37).

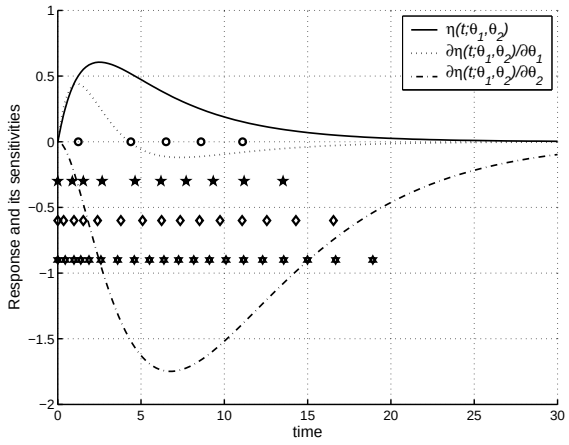
Box and Lucas (1959) found D-optimum designs for this system when only [B] was measured. They took $\theta_1^0 = 0.7$ and $\theta_2^0 = 0.2$, with $\lambda_1 = \lambda_2 = 1$ and assumed independent errors of constant variance. Their optimum design puts half the trials at $t = 1.23$ and the other half at $t = 6.86$, a design with replication that is meaningless with correlated observations.

Uciński and Atkinson (2004) found optimum designs for the same system, again with first-order kinetics, but with the measurements corrupted by zero-mean correlated noise with covariance kernel

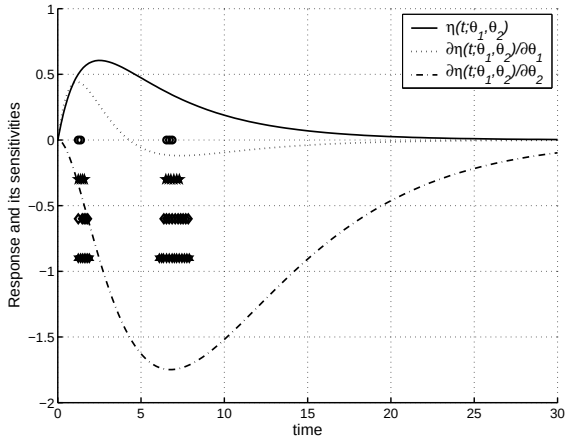
$$C(t, s) = \exp(-\tau|t - s|) \tag{38}$$

under the assumption that there was only one measurement schedule. That is, in (24) $n = 1$. The upper panel of Figure 1 shows the resulting optimum designs for strong correlation ($\tau = 1$) and $q = 5, 10, 15$ and 20. In all designs the measurement instants are spread across part of the design region. As q increases so does the spread of the points, which however are more concentrated at lower times. There is an appreciable contrast with the designs in the lower panel of the figure for weak correlation ($\tau = 50$) where the design points are clustered around the values 1.23 and 6.86. Plots for other values of τ confirm this smooth transition from well spread out designs to those close to the two point designs for independence as $\tau \to \infty$.

We now turn to the ideas developed in this paper and illustrate them on the same example, with again the only measurements being of [B]. However we now take $\lambda_1 = \lambda_2 = 2$, so that there is no longer an analytical solution and numerical solution of the sets of differential equations is necessary. As an aside, note that the sensitivities have only to be found once, since we design for given θ^0. Thereafter interpolation in the grid of numerical values of the sensitivities is used to give the values for specific values of t required by the design. In order to find optimal points for taking measurements in the interval $T = [0, 30]$, a program was written in Lahey-Fujitsu Fortran 95 v.5.7 and run on a Pentium IV PC equipped with Windows 2000. The set of ordinary differential equations (2) along with the corresponding sensitivity

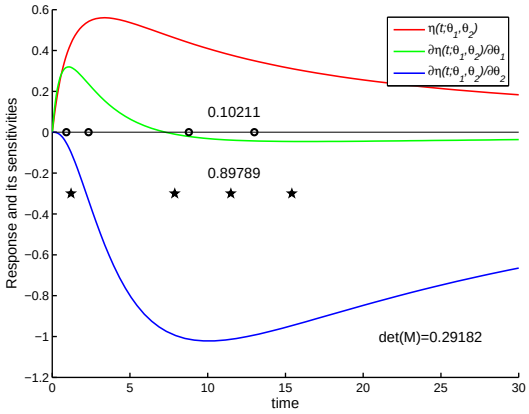


(a) considerable correlation ($\tau = 1$)

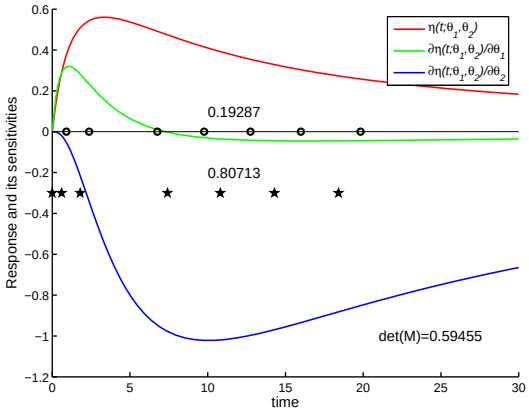


(b) small correlation ($\tau = 50$)

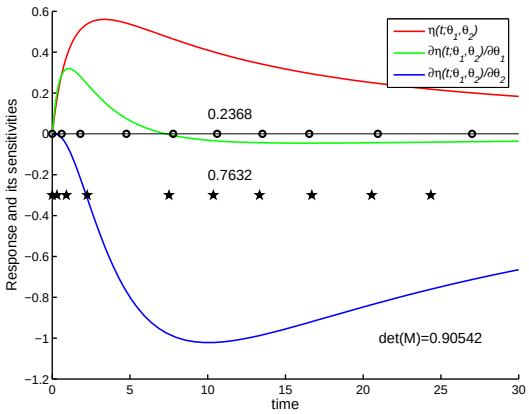
Figure 1: Time evolution of the reactant concentration $[B]$ (solid line) and its sensitivities to parameters θ_1 and θ_2. One measurement series with $\lambda_1 = \lambda_2 = 1$. Circles, pentagons, diamonds and hexagons denote D-optimum measurement moments for $q = 5, 10, 15$ and 20, respectively.



(a) $q = 4$



(b) $q = 7$



(c) $q = 10$

Figure 2: Time evolution of the reactant concentration $[B]$ (uppermost line) and its sensitivities to parameters θ_1 and θ_2 when $\lambda_1 = \lambda_2 = 2$. Unconstrained number of series. D-optimum measurement moments for $q = 4, 7$ and 10. Appreciable correlation ($\tau = 1$). The two measurement sequences are represented by circles and stars with the weights p_i given above

coefficients were solved using the `divmrk` routine (an integrator based on using Runge-Kutta pairs of various orders) from the IMSL math library Visual Numerics (1997). The one-dimensional search in Step 3 of the Algorithm was performed using the procedure `golden` from Press, Teukolsky, Vetterling, and Flannery (1996) (an implementation based on Golden Section search).

Figure 2 shows the resulting designs for appreciable correlation ($\tau = 1$) for $q = 4$, 7 and 10. For all three values of q two measurement series provide the optimum design. For $q = 4$, the two series are quite dissimilar, with weights 0.102 and 0.898. As q increases the series become more similar and the weights diverge: for $q = 10$ we obtain 0.237 and 0.763. We have not yet observed the limit of this process, which would be weights of 0.5 on two measurement series that are virtually identical. For medium correlation ($\tau = 5$) and $q = 4$ we again obtained two series of measurements with unequal weights but, for $q = 7$ and 10 only one series was required, as it was for weak correlation ($\tau = 50$). For the design for independent errors we had that $q = m(= 2)$. It is clear that, as $\tau \to \infty$ in such situations only one series is required, provided $q \geq m$. However it remains for us to find the taxonomy of situations requiring more than one series of observations.

References

Atkinson, A. C. and B. Bogacka (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems 61*, 17–33.

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Designs*. Oxford: Clarendon Press.

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika 46*, 77–90.

Cox, D. R. and N. Reid (2000). *The Theory of the Design of Experiments*. Boca Raton, FL: Chapman and Hall.

Ermakov, S. M. (Ed.) (1983). *Mathematical Theory of Experimental Design*. Moscow: Nauka. (In Russian).

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments*. Lecture Notes in Statistics. New York: Springer-Verlag.

Kiefer, J. and J. Wolfowitz (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics 30*, 271–294.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. Mathematics and Its Applications. Dordrecht, The Netherlands: D. Reidel

Publishing Company.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1996). *Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing* (2nd ed.). Cambridge: Cambridge University Press.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Probability and Mathematical Statistics. New York: John Wiley & Sons.

Rafajłowicz, E. (1996). *Algorithms of Experimental Design with Implementations in MATHEMATICA*. Warsaw: Akademicka Oficyna Wydawnicza PLJ. (In Polish).

Silvey, S. D. (1980). *Optimal Design. An Introduction to the Theory for Parameter Estimation*. London: Chapman and Hall.

Uciński, D. (2005). *Optimal Measurement Methods for Distributed-Parameter System Identification*. Boca Raton, FL: CRC Press.

Uciński, D. and A. C. Atkinson (2004). Experimental design for time-dependent models with correlated observations. *Studies in Nonlinear Dynamics & Econometrics 8*(2). Article No. 13.

Valko, P. and S. Vajda (1984). An extended ODE solver for sensitivity calculations. *Computers and Chemistry 8*, 255–271.

Venot, A., L. Pronzato, É. Walter, and J.-F. Lebruchec (1986). A distribution-free criterion for robust identification, with applications in system modelling and image processing. *Automatica 22*(1), 105–109.

Visual Numerics (1997). *IMSL. Fortran Subroutines for Mathematical Applications. Vol. 1*. Houston, TX: Visual Numerics, Inc.

Walter, É. and L. Pronzato (1997). *Identification of Parametric Models from Experimental Data*. Communications and Control Engineering. Berlin: Springer-Verlag.