Experimental design for time-dependent models with correlated observations

Dariusz Uciński* Anthony C. Atkinson†

1 Introduction

This paper makes a start on the design of optimum experiments for the parameters in a regression model when the errors have a correlation structure. We motivate our work in the next section by an example drawn from chemical kinetics, stressing the difference between chemical models and statistical models; the latter including the errors of observation, which are crucial to the design of an informative scheme of observations. In §3 we summarise the theory of optimum experimental design for parameter estimation. The overwhelming majority of the work has assumed independent errors. Although it is straightforward to write down the extension of the design criteria to correlated errors, it is difficult to build an efficient algorithm for finding designs. We develop such an algorithm in §4, giving numerical results in §5.

The paper concludes in §6 with discussion of outstanding issues. Our algorithm allows us to find "exact" designs, that is designs for a specified number of observations, in which the response is defined by a system of differential equations requiring numerical solution. It also covers the simpler case of linear models, such as polynomials, but again with correlated errors. We stress the contrast with earlier theoretical work, which was concerned with the asymptotic properties of designs, but did not lead to an algorithm

*Institute of Control and Computation Engineering, University of Zielona Góra, ul. Podgórna 50, 65–246 Zielona Góra, Poland, e-mail: D.Ucinski@irio.pz.zgora.pl

†Department of Statistics, London School of Economics, London WC2A 2AE, UK, e-mail: A.C.Atkinson@lse.ac.uk

for constructing exact designs. In §6 we also consider a variety of design problems with correlated errors. Variants of our algorithm should lead to the solution of many of them.

2 Model of a general consecutive reaction

2.1 The chemical model

Our example is of a simple chemical reaction which is, however, sufficiently rich to illustrate all required points. In the model for two consecutive reactions

$$A \overset{\theta_1}{\to} B \overset{\theta_2}{\to} C, \tag{1}$$

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 an experiment reduces to the choice of times at which measurements of the concentrations are to be taken. We discuss extensions to optimally varying conditions in §6.

In general 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), with an analytical expression for $[A]$. If, however, we take both reactions as first order, that is $\lambda_1 = \lambda_2 = 1$, an explicit algebraic solution 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

$$\eta(t, \beta^\star) = \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}), \tag{3}$$

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.

Although, in our numerical example in §5, we take both reactions as first order, we have programmed the procedure to work directly on the differential equations (2) and so to be unrestricted to specific values of λ. We do however restrict ourselves to experiments in which just one concentration, that of B, is measured.

2.2 The statistical model

To design experiments we need a statistical model of the process, that is one including error. Consider the model in the form

$$y(t) = \eta(t, \beta^\star) + \varepsilon(t), \quad t \in T = [0, t_f], \tag{4}$$

where t_f is a fixed finite time horizon, $\beta^\star \in \mathbb{R}^m$ signifies an unknown constant parameter vector, $\eta(t, \beta^\star)$ is a known (possibly non-linear) real-valued regression function. It is the error process $\varepsilon(t)$ that is the crucial difference between our work and the large statistical literature on optimum experimental design.

The nonlinear model (3) was introduced into the statistical literature by Box and Lucas (1959) who assumed that the errors of observation were independent and of constant variance σ^2, i.e.

$$\mathrm{E}[\varepsilon(t)\varepsilon(s)] = \begin{cases} \sigma^2 & \text{if } s = t, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The experiment consists in running the chemical reaction once from $t = 0$ and measuring the response $y(t)$ at time moments $t_1, \ldots, t_n$ for a given n, , i.e.

$$y(t_i) = \eta(t_i, \beta^\star) + \varepsilon(t_i), \quad i = 1, \ldots, n. \tag{6}$$

The parameter $\beta^\star$ is estimated by nonlinear least squares.

The customary design criterion for this and other nonlinear models is D-optimality, which we discuss in more detail in §3. This criterion maximises the determinant of the information matrix of the parameter estimates. Examples for several nonlinear models are in Chapter 18 of Atkinson and Donev (1992). All use the assumption of independent errors. This assumption is appropriate if each experiment results in a single observation at a time t_i. The next observation then comes from restarting the reaction and again taking a single observation. In this way a sample of n observations is built up from n runs

of the experiment. For the two parameter nonlinear model (3) this leads, if n is even, to a design in which $n/2$ observations are taken at one time and the remaining $n/2$ at another time, the values of the two time points depending upon the values of the parameters θ_1 and θ_2. If n is odd, one condition is repeated once more than the other — it does not matter which.

In contrast to this standard formulation, we suppose that the experiment is run only once, and that all n observations come from the same run. We then have to allow for the correlation between the observations, one effect of which is that replicate observations at a time point provide no more information than a single observation. The design for independent observations is therefore inappropriate. Accordingly, in the next section we write down the standard expression for the covariance matrix of a stochastic process with mean function zero and known covariance kernel, and use this in the calculation of D-optimum designs. We find that, the higher the correlation between adjacent observations, the more are the observations spread over the experimental region. Conversely, a low correlation leads to designs with two clusters of points, of course without replication; sometimes the clusters of observations are not of quite equal sizes.

3 Experimental design problem for correlated observations

We continue to write the regression model in the form (4), but now the error term $\varepsilon(t)$ is a stochastic process with zero mean and known continuous covariance kernel

$$\mathrm{E}[\varepsilon(t)\varepsilon(s)] = k(t,s) \quad \text{on } T^2. \tag{7}$$

The generalized least-squares estimate of the unknown parameter β from the n experimental observations (6) is given by

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}[y(t_i) - \eta(t_i,\beta)][y(t_j) - \eta(t_j,\beta)]. \tag{8}$$

The weights w_{ij} above constitute the elements of the matrix

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} = \begin{bmatrix} k(t_1,t_1) & \cdots & k(t_1,t_n) \\ \vdots & \ddots & \vdots \\ k(t_n,t_1) & \cdots & k(t_n,t_n) \end{bmatrix}^{-1}. \tag{9}$$

Clearly, the accuracy of the estimates will depend on the choice of moments t_i, $i = 1, \ldots, n$, as some points of the time interval T may give more information about $\beta^\star$ than others. A crucial question is then how to select time moments which will guarantee the best accuracy of (8). From a statistical point of view, the parameter estimation accuracy is quantified by some scalar measure of performance Ψ defined on the *Fisher Information Matrix* (FIM)

$$M = F^{\mathrm{T}} W F, \tag{10}$$

where

$$F = \begin{bmatrix} \nabla_\beta^{\mathrm{T}} \eta(t_1, \beta^0) \\ \vdots \\ \nabla_\beta^{\mathrm{T}} \eta(t_n, \beta^0) \end{bmatrix}, \quad \nabla_\beta \eta(t_i, \beta^0) = \begin{bmatrix} \partial \eta(t_i, \beta^0)/\partial \beta_1 \\ \vdots \\ \partial \eta(t_i, \beta^0)/\partial \beta_m \end{bmatrix}, \quad i = 1, \ldots, n, \tag{11}$$

β^0 being a prior estimate of the unknown parameter vector β. The elements of the matrix F are usually called the *sensitivity coefficients*. Such a 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.

Various choices exist for $\Psi(M)$, but the determinant of the FIM, i.e.

$$\Psi(M) = \det(M), \tag{12}$$

is the most widely criterion in the statistical literature. Its maximization leads to the so-called *D-optimum* solutions (*designs*). They minimize the volume of the uncertainty ellipsoid for the estimates.

For notational convenience, in what follows, we shall use the following compact notation to denote a solution to our problem:

$$\xi = \{t_1, \ldots, t_n\}. \tag{13}$$

Each ξ will then be called a *discrete* (or *exact*) *design* and its elements *support points*. Since all measurements are to be taken on one run of the process (4), so that replications are not allowed, the admissible designs must satisfy the condition $t_i \neq t_j$ for $i, j = 1, \ldots, n$ and $i \neq j$. Thus, from among all discrete designs we wish to select the one which maximizes the determinant of the information matrix $M(\xi)$. We call such a design a *D-optimum design*.

This design problem differs from the much studied classical design problem because the matrix W may here be non-diagonal, due to the correlations between different observations. This constitutes a major obstacle when trying directly to apply the numerical algorithms already existing in the optimum experimental design literature, which were constructed for experiments with uncorrelated observations. The technical point is that the algorithms make explicit use of the fact that, for uncorrelated observations, the FIM can be expressed as the sum of the FIMs of the individual points forming the design. This is no longer true when the observations are correlated.

4 Exchange-type algorithm for computation of D-optimum designs

4.1 Scheme of the algorithm

There are extremely few results regarding practical determination of D-optimum designs in the context of correlated observations. Most authors assume a correlation-free framework and then apply well-developed sequential algorithms to generate designs claimed to be acceptable. Such an assumption can sometimes hardly be justified as far as dynamic processes over time are considered where correlations between observations are rather natural.

In order to calculate exact designs, we adapt the numerical algorithm presented by Brimkulov, Krug, and Savanov (1986) for determining the D-optimum sampling points for estimation of the parameters in linear models for expectations of random fields. It is an exchange-type algorithm that starts from an arbitrary initial n-point design $\xi^{(0)}$; at each iteration one support point is deleted from the current design and a new point

from the domain T is included in its place in such a way as to maximize the resulting increase in the determinant of the FIM. Since the results from Brimkulov et al. (1986) are not widely known, as they were published in Russian, we restate them, in some detail if necessary.

The general scheme of the algorithm is outlined below. For iteration ℓ, we use the following notation: $F^{(\ell)}$ is the current matrix of sensitivity coefficients in (11), $W^{(\ell)}$ stands for the current weighting matrix (9), $M^{(\ell)}$ means the resulting FIM and $D^{(\ell)} = \left\{M^{(\ell)}\right\}^{-1}$. Moreover, $I = \{1, \ldots, n\}$ and $\xi^{(\ell)}_{t_i \leftrightarrows t}$ denotes the design in which the support point t_i is replaced by t.

Algorithm

Step 1. Select an initial design $\xi^{(0)} = \left\{t_1^{(0)}, \ldots, t_n^{(0)}\right\}$ such that $t_i^{(0)} \neq t_j^{(0)}$ for $i, j \in I$ and $i \neq j$. Calculate the matrices $F^{(0)}$, $W^{(0)}$ and $M^{(0)}$, and then the value of $\det M^{(0)}$. If $\det M^{(0)} = 0$, select a new initial design and repeat this step.

Step 2. Calculate $D^{(0)}$. Set $\ell = 0$.

Step 3. Determine

$$(i^*, t^*) = \arg \max_{(i,t) \in I \times T} \Delta(t_i, t),$$

where $\Delta(t_i, t) = \left\{\det M(\xi^{(\ell)}_{t_i \leftrightarrows t}) - \det M(\xi^{(\ell)})\right\} / \det M(\xi^{(\ell)})$.

Step 4. If $\Delta(t_i^*, t^*) \leq \delta$, where δ is some given positive tolerance, then STOP. Otherwise, set $\xi^{(\ell+1)} = \xi^{(\ell)}_{t_{i^*} \leftrightarrows t^*}$ and determine $F^{(\ell+1)}$, $W^{(\ell+1)}$, $M^{(\ell+1)}$ and $D^{(\ell+1)}$ corresponding to $\xi^{(\ell+1)}$. Set $\ell \leftarrow \ell + 1$ and go to Step 3.

Remark 1. From the point of view of non-linear programming, if we treat the design problem as an optimization problem with decision variables $t_1, \ldots, t_n$ and performance index $\det M$, the algorithm outlined above constitutes a Gauss-Seidel-like algorithm. In the classical Gauss-Seidel relaxation scheme (also called the block coordinate ascent method, cf. Bertsekas 1999), each iteration consists of n one-dimensional search steps with respect to variables $t_1, \ldots, t_n$ taken in cyclic order. If an increase in $\det M$ is attained for some i after performing the search, then the corresponding variable t_i is immediately updated. The design algorithm outlined above differs from the Gauss-Seidel scheme in the update,

which takes place only for the coordinate t_i for which the resulting increase in $\det M$ is the largest. As a result, the delineated design algorithm possesses convergence properties similar to those of the Gauss-Seidel scheme, in particular only convergence to a local maximum can be guaranteed. This implies that several restarts must often be performed from different initial designs so as to obtain an approximation to a global maximum. Owing to optimized updates of the matrices in Steps 3 and 4 through elimination of matrix inverses (see the next section), the algorithm is fast enough in spite of this inconvenience.

4.2 Implementation details

The speed of the algorithm can be greatly improved by elimination of time-consuming operations, such as inversions of matrices and the calculation of determinants in Step 3. We here give some details, based on derivations from Brimkulov et al. (1986). Without loss of generality, we assume that we are always to replace point t_n by a point $t \in T$. Indeed, deletion of a point t_i for $i < n$ can be converted to this situation by interchanging positions of t_i and t_n in $\xi^{(\ell)}$, which then should be followed by swapping the i-th and n-th rows in $F^{(\ell)}$, as well as interchanging the i-th and n-th rows and the i-th and n-th columns in $W^{(\ell)}$. Such a replacement greatly simplifies the resulting formulae and makes the implementation easier.

Assume that we are to pass from $\xi^{(\ell)}$ to $\xi^{(\ell)}_{t_n \leftrightarrows t}$. This task requires two stages: the removal of t_n from $\xi^{(\ell)}$ and the augmentation of the resulting design by t. Clearly, both stages imply changes in all the matrices corresponding to the current design. Applying the matrix-inversion theorem, we can derive the forms of the necessary updates which guarantee the extreme efficiency of the calculations.

Stage 1: Deletion of t_n from $\xi^{(\ell)}$. Write $F^{(\ell)}$ as

$$F^{(\ell)} = \begin{bmatrix} F_r \\ \cdots\cdots \\ f_r^{\mathrm{T}} \end{bmatrix}, \tag{14}$$

where $F_r \in \mathbb{R}^{(n-1)\times n}$ and $f_r \in \mathbb{R}^n$. Deletion of t_n from $\xi^{(\ell)}$ implies removing f_r^{T} from $F^{(\ell)}$, so that we then have F_r instead of $F^{(\ell)}$. But some changes are also necessary in matrices

$W^{(\ell)}$ and $M^{(\ell)}$. Namely, decomposing the symmetric matrix $W^{(\ell)}$ into

$$W^{(\ell)} = \left[\begin{array}{c:c} V_r & b_r \\ \hdashline b_r^{\mathrm{T}} & \gamma_r \end{array}\right], \tag{15}$$

where $V_r \in \mathbb{R}^{(n-1)\times(n-1)}$, $b_r \in \mathbb{R}^{n-1}$, $\gamma_r \in \mathbb{R}$, we set

$$g_r = \frac{1}{\sqrt{\gamma_r}} F_r^{\mathrm{T}} b_r + \sqrt{\gamma_r} f_r, \tag{16}$$

and then calculate respectively the following counterparts of $W^{(\ell)}$ and $M^{(\ell)}$:

$$W_r = V_r - \frac{1}{\sqrt{\gamma_r}} b_r b_r^{\mathrm{T}}, \quad M_r = M^{(\ell)} - g_r g_r^{\mathrm{T}}. \tag{17}$$

Stage 2: Inclusion of t into the design resulting from Stage 1. We construct

$$F_a = \left[\begin{array}{c} F_r \\ \hdashline f_a^{\mathrm{T}}(t) \end{array}\right], \tag{18}$$

where $f_a(t) = \nabla_\beta \eta(t, \beta^0)$. Such an extension of the matrix of sensitivity coefficients influences the form of matrices W_r and M_r obtained at Stage 1. In order to determine their respective updated versions $W_a(t)$ and $M_a(t)$, we define

$$v_a(t) = [k(t,t_1), \ldots, k(t,t_n)]^{\mathrm{T}}, \qquad \gamma_a(t) = 1/[k(t,t) - v_a^{\mathrm{T}}(t) W_r v_a(t)], \tag{19}$$

$$g_a(t) = \sqrt{\gamma_a(t)}[f_a(t) - F_r^{\mathrm{T}} W_r v_a(t)], \qquad b_a(t) = -\gamma_a(t) W_r v_a(t). \tag{20}$$

Then we have

$$W_a(t) = \left[\begin{array}{c:c} W_r + \dfrac{1}{\gamma_a(t)} b_a(t) b_a^{\mathrm{T}}(t) & b_a(t) \\ \hdashline b_a^{\mathrm{T}}(t) & \gamma_a(t) \end{array}\right], \quad M_a(t) = M_r + g_a(t) g_a^{\mathrm{T}}(t). \tag{21}$$

It can be proven that, once the point t_n has been replaced by t in the above two-stage procedure, the ratio $\Delta(t_n, t)$ in Step 3 of the Algorithm can be calculated as

$$\Delta(t_n, t) = d_a(t) - d_r - d_r d_a(t) + d_{ra}(t), \tag{22}$$

where

$$d_r = g_r^{\mathrm{T}} D^{(\ell)} g_r, \quad d_a = g_a^{\mathrm{T}}(t) D^{(\ell)} g_a(t), \quad d_{ra} = g_r^{\mathrm{T}} D^{(\ell)} g_a(t). \tag{23}$$

Completion of Step 3 means that points $t_{i^*} \in \xi^{(\ell)}$ (identified with t_n) and $t^* \in T$ which guarantee the largest increase in the D-optimality criterion for the current iteration have been found. At this juncture, we may update the matrices as follows:

$$F^{(\ell+1)} = F_a(t^*), \quad W^{(\ell+1)} = W_a(t^*), \quad M^{(\ell+1)} = M_a(t^*), \tag{24}$$

$$\begin{aligned} D^{(\ell+1)} = D^{(\ell)} - \frac{1}{1+\Delta(t_n,t^*)} D^{(\ell)} \Big\{ & (1-d_r) g_a(t^*) g_a^{\mathrm{T}}(t^*) \\ & + d_{ra}(t^*) \big[g_r g_a^{\mathrm{T}}(t^*) + g_a(t^*) g_r^{\mathrm{T}} \big] - [1+d_a(t^*)] g_r g_r^{\mathrm{T}} \Big\} D^{(\ell)}. \end{aligned} \tag{25}$$

What is more, we get

$$\det M^{(\ell+1)} = \det M^{(\ell)} [1 + \Delta(t_n, t^*)]. \tag{26}$$

The relatively simple form of the above formulae guarantees the efficiency of the algorithm which becomes thus very similar to Fedorov's exchange algorithm for determining exact designs (Fedorov 1972a).

Remark 2. It can easily be easily verified that $g_a(t_i) = 0$ for $i = 1, \ldots, n-1$. This means from (21) that inclusion of a point coinciding with a point which already exists in the design does not alter the information matrix. In other words, an additional observation at the same point does not provide more information about the estimated parameters. This implies that the optimal design will automatically be replication-free.

5 Numerical example

To illustrate the ideas presented in the paper, consider the reaction described by

$$\begin{cases} \dfrac{\mathrm{d}[A]}{\mathrm{d}t} = -\theta_1 [A], & [A]_{t=0} = 1, \\ \dfrac{\mathrm{d}[B]}{\mathrm{d}t} = \theta_1 [A] - \theta_2 [B], & [B]_{t=0} = 0, \\ \dfrac{\mathrm{d}[C]}{\mathrm{d}t} = \theta_2 [B], & [C]_{t=0} = 0, \end{cases} \tag{27}$$

which is (2) when both reactions are first order. We assume that the observed response is the reactant concentration $[B]$, the expected value of which is given by (3). The parameters θ_1 and θ_2 are to be estimated based on measurements made at n different time moments. The measurements are assumed to be corrupted by zero-mean correlated noise with covariance kernel

$$k(t,s) = \exp(-\rho |t-s|). \tag{28}$$

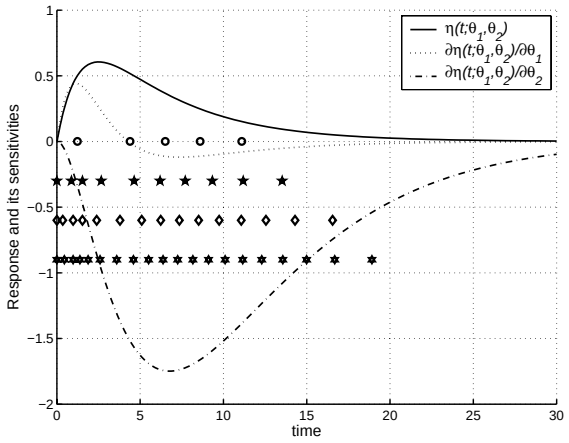
The nominal parameter values $\theta_1^0 = 0.7$ and $\theta_2^0 = 0.2$ were used. For independent errors, Box and Lucas (1959) show that the optimum design then puts equal numbers of observations at $t = 1.23$ and 6.86.

In order to find optimal points for taking measurements in the interval $T = [0, 30]$, a program was written in Matlab v.6.5 Release 13 and run on a Pentium IV PC equipped with Windows 2000. The set of ordinary differential equations (27) along with the corresponding sensitivity coefficients were solved using the built-in subroutine `ode45` (an integrator based on an explicit Runge-Kutta formula). The one-dimensional search for a point to be included in the design was performed using the procedure `fminbnd` from the Optimization toolbox (an implementation based on Golden Section search and parabolic interpolation). We thus chose not to use the analytical form of the sensitivities found by differentiation of (3).

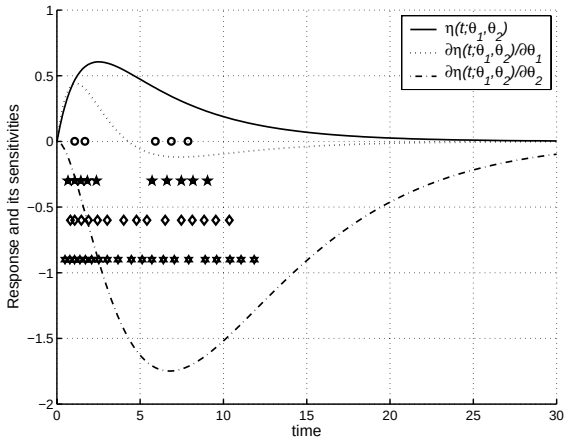
The results obtained for various numbers of measurements n are presented in Fig. 1. For each n, the program was restarted several times from randomly generated initial conditions in order to avoid getting stuck in a local maximum. The time of each run did not exceed twenty seconds.

The influence of the correlation between observations was tested by varying the coefficient ρ in (28). In particular, the ρ values of 1, 5 and 50 were chosen as representative ones for considerable, medium and small correlations, respectively (as a justification for this terminology, note that the model (28) implies that, in practice, statistical interactions between different points are certainly negligible at distances of 5, 1 and 0.1, respectively).

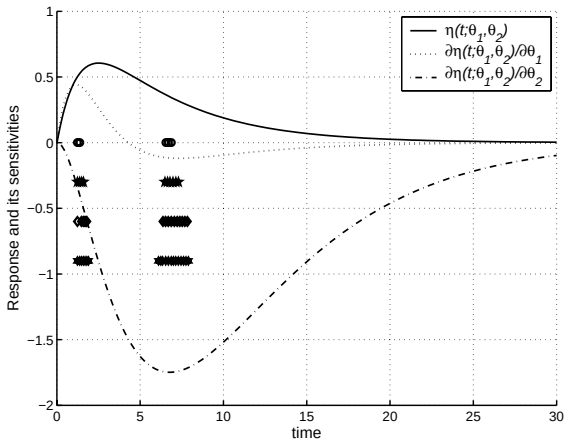
Two main conclusions can be drawn from our results. First, the greater the correlation, the more the optimal observations are spread over the domain where measurements can be taken. If the correlation is less significant, observations tend to form two clusters with approximately equal numbers of elements. From Panel (c) of the figure it is clear that the design points are clustering around the times 1.23 and 6.86. The second conclusion is that extrema of the sensitivity coefficients (denoted by dotted and dash-dotted lines) turn out to be close to the best points for taking measurements in this case. Box and Lucas (1959) discuss the geometry of optimum designs for independent observations, as do Atkinson and Donev (1992); the points of the optimum design for $n = 2$ form, together with the origin, the triangle of maximum area in the space of the sensitivities. This



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



(b) medium correlation ($\rho = 5$)



(c) small correlation ($\rho = 50$)

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

confirms the intuition that times with high values of the parameter sensitivities will yield informative readings. Note that the higher the correlation, the lower is the value of the D-optimality criterion. This phenomenon occurs because higher correlations impose stronger relationships between noise realizations at different points; this "structural" noise interferes with estimation of the parameters θ, so that the information in the experiment is reduced.

6 Discussion

6.1 Literature

There is a large statistical literature on the design of experiments. A survey of the work of the last 100 years is given by Atkinson and Bailey (2001). Our paper builds on several strands in this work.

One is optimum design, first given its modern formulation by Kiefer (1959). Book length treatments include Fedorov (1972b), Pazman (1986), Atkinson and Donev (1992), Pukelsheim (1993) and, briefly, Fedorov and Hackl (1997a). Despite this list, only Brimkulov et al. (1986) consider correlated errors in any detail. Their examples however are for simple linear models. Optimum designs for the four-parameter nonlinear model (2), but with independent errors, are found by Atkinson and Bogacka (2002).

There is also some work on optimum design in the control theory literature, but the relevant works consider optimal input design for discrete-time linear dynamic systems. Surveys are given by Mehra (1974), Goodwin and Payne (1977) and Titterington (1980), as well as more recently by Walter and Pronzato (1997). The design procedures can be divided into the categories of time-domain methods and frequency-domain methods, with the former being more general, but also more consuming of computational time. These methods cannot, however, be directly applied to the problem considered in the present paper.

Correlated observations constitute a fundamental problem in spatial statistics (Fedorov and Hackl 1997b; Müller 1998). Although the framework is somewhat different, as the main problem is to make a best spatial prediction, some interactions with this work are to be expected.

Correlated errors have been considered for treatment allocation designs in which t treatments have to be allocated along a line or over a plane. The main requirement is that of neighbour balance (Williams 1952). Kiefer and Wynn (1984) explore the relationship with coding theory.

Experimental designs for regression problems with correlated errors were thoroughly studied from the theoretical point of view in Parzen (1961), Sacks and Ylvisaker (1966), Sacks and Ylvisaker (1968), Sacks and Ylvisaker (1970) and Wahba (1971). The problem considered there differed from ours in that an optimal number of support points was sought in addition to their coordinates. The main difficulty in such a formulation stems from the fact that every new observation gives a new piece of information about the parameters, so usually a solution with a finite number of support points does not exist. An exception is when the basis functions multiplying the estimated parameters in the linear regression model can be represented in the form $f(t) = \sum_{i=1}^{N} a_i k(t, t_i)$ for some finite N and fixed numbers a_i and t_i (more generally, they are elements of the so-called reproducing kernel Hilbert space associated with the kernel $k(t, s)$). But this assumption is too strong to be satisfied in most practical situations. Some relaxation of this condition comes from introducing the notion of the so-called asymptotic optimality for a sequence of designs; but this can hardly be applied in the framework of the present paper when we are seeking designs for a specified n.

6.2 Extensions and Related Problems

The main contribution of our paper is to provide a numerical method for the construction of optimum designs in the presence of correlation. The crucial step is the algebra of §4.2 which leads to a rapid exchange-type algorithm. There are however a number of extensions that require attention.

One is the extension to the measurement of multiple responses. The extension of Box and Lucas (1959) to multiresponse design for independent error is given, for example, by Draper and Hunter (1966). A second is the extension to multiple runs. Suppose we can take n_k observations in the kth run. Even though the errors between runs will be independent, it is not clear whether the optimum design over all consists of the assembly of the optimum deign for each n_k. Clearly, when all $n_k = 1$ for our two parameter example,

there is an effect of the designs for individual runs on each other.

We have extended optimum design from the standard error assumption of independence to allow for correlation. However, there is evidence that, in some kinetic models, a power transformation of the response is needed to provide homogeneity of variance. We need to extend the procedure for design with transformations of Atkinson (2003) from independent errors to allow for correlation.

Our motivation has come from examples in chemical kinetics in which we can expect that the errors are correlated over time, as they would be in pharmacokinetic experiments. However (Rocke and Lorenzato 1995) there is some evidence that there may be two error strata, the correlated observations having independent errors added to them arising from instrument and measurement errors. In such cases there would be some information from replications. The structure of the resulting designs is not clear to us.

We have assumed that the spacing of the design points is solely determined by the design algorithm responding to the sensitivities and the correlation structure. However, there are situations in which there is physically a minimum time between observations, for example because a sampling instrument is blocked for a fixed time period until the sample or analysis have been transmitted. Some examples are given by Bohachevsky, Johnson, and Stein (1986) for independent errors. Inclusion of such constraints in our algorithm for correlated errors seems to pose difficult numerical problems unless a search is performed on a sufficiently dense spatial grid of points.

There are also situations in which the time points are not an issue, measurements for example being taken at regular intervals. Then the experimental variables are the starting conditions and the profiles of other variables during the run. Fedorov and Hackl (1997b, p. 72) give a simple example for a first-order polynomial model in the presence of correlation. Often industrial processes are not of the batch kind described in §2, but are continuous; there may be a stirred reactor to which components are added and from which product is withdrawn. Then the experimental design will again consist of time profiles of, for example, flows, temperatures and concentrations, together perhaps with time points of observation. An example is Bauer, Bock, Körkel, and Schlöder (2000) where, however, independent errors are assumed.

Finally, we wonder whether any of the numerical methods associated with time series

analysis can simplify the search for optimum designs. For example, can the Kalman filter and state-space representations provide us with a more transparent derivation of our algorithm which can be extended to the solution of the harder problems listed above?

Acknowledgment

This work was supported by the British Council and the State Committee for Scientific Research within the framework of a Polish-British Research Partnership Programme.

References

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Applied Statistics 52*. (in press).

Atkinson, A. C. and R. A. Bailey (2001). One hundred years of the design of experiments on and off the pages of. *Biometrika. Biometrika 88*, 53–97.

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: Oxford University Press.

Bauer, I., H. G. Bock, S. Körkel, and J. P. Schlöder (2000). Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics 120*, 1–25.

Bertsekas, D. P. (1999). *Nonlinear Programming* (2-nd ed.). Optimization and Computation Series. Belmont: Athena Scientific.

Bohachevsky, I. O., M. E. Johnson, and M. L. Stein (1986). Generalized simulated annealing for function optimization. *Technometrics 28*, 209–217.

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

Brimkulov, U. N., G. K. Krug, and V. L. Savanov (1986). *Design of Experiments in Investigating Random Fields and Processes*. Moscow: Nauka. (in Russian).

Draper, N. R. and W. G. Hunter (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika 53*, 596 – 599.

Fedorov, V. V. (1972a). *Theory of Optimal Experiments*. New York: Academic Press.

Fedorov, V. V. (1972b). *Theory of Optimal Experiments*. New York: AcademicPress.

Fedorov, V. V. and P. Hackl (1997a). *Model-Oriented Design of Experiments*. Lecture Notes in Statistics 125. New York: Springer.

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

Goodwin, G. C. and R. L. Payne (1977). *Dynamic System Identification. Experiment Design and Data Analysis*. Mathematics in Science and Engineering. New York: Academic Press.

Kiefer, J. (1959). Optimum experimental designs (with discussion). *J. Roy. Statist. Soc. B 21*, 272–319.

Kiefer, J. and H. P. Wynn (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error processes. *Annals of Statistics 12*, 414–450.

Mehra, R. K. (1974, December). Optimal input signals for parameter estimation in dynamic systems-survey and new results. *IEEE Transactions on Automatic Control 19*(6), 753–768.

Müller, W. G. (1998). *Collecting Spatial Data. Optimum Design of Experiments for Random Fields*. Contributions to Statistics. Heidelberg: Physica-Verlag.

Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics 32*(4), 951–989.

Pazman, A. (1986). *Foundations of Optimum Experimental Design*. Dordrecht: Reidel.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

Rocke, D. M. and S. Lorenzato (1995). A two-component model for measurement error in analytical chemistry. *Technometrics 37*, 176–184.

Sacks, J. and D. Ylvisaker (1966). Designs for regression problems with correlated errors. *The Annals of Mathematical Statistics 37*(1), 66–89.

Sacks, J. and D. Ylvisaker (1968). Designs for regression problems with correlated errors: Many parameters. *The Annals of Mathematical Statistics 39*(1), 49–69.

Sacks, J. and D. Ylvisaker (1970). Designs for regression problems with correlated errors III. *The Annals of Mathematical Statistics 41*(6), 2057–2074.

Titterington, D. M. (1980). Aspects of optimal design in dynamic systems. *Technometrics 22*(3), 287–299.

Wahba, G. (1971). On the regression design problem of Sacks and Ylvisaker. *The Annals of Mathematical Statistics 42*(3), 1035–1053.

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

Williams, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika 39*, 151–167.