Journal of Statistical Planning and Inference 139 (2009) 3947 – 3961

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Multiplicative algorithms for computing optimum designs

B. Torsney[a,1], R. Martín-Martín[b,*]

[a]Department of Statistics, University of Glasgow, Glasgow G12 8QW, UK
[b]School of Computer Science, Department of Mathematics, The Institute of Applied Mathematics in Science and Engineering, University of Castilla-La Mancha, Paseo de la Universidad, 4 13071 Ciudad Real, Spain

ARTICLE INFO

Article history:
Received 26 February 2008
Received in revised form
7 April 2009
Accepted 8 May 2009
Available online 19 May 2009

Keywords:
Exact design
Approximate design
Multiplicative algorithm
Covariance function

ABSTRACT

We study a new approach to determine optimal designs, exact or approximate, both for the uncorrelated case and when the responses may be correlated. A simple version of this method is based on transforming design points on a finite interval to proportions of the interval. Methods for determining optimal design weights can therefore be used to determine optimal values of these proportions. We explore the potential of this method in a range of examples encompassing linear and non-linear models, some assuming a correlation structure and some with more than one design variable.

Contents

* Corresponding author. Tel.: +34 926 295300x3711; fax: +34 926 295354.
E-mail addresses: ben@stats.gla.ac.uk (B. Torsney), Raul.MMartin@uclm.es (R. Martín-Martín).
[1] Tel.: +44 141 330 5149; fax: +44 141 330 4814.

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.05.007

1. Optimal design problems

Consider the linear model

$$y(\underline{x}) = \underline{\eta}^T(\underline{x})\underline{\theta} + \varepsilon(\underline{x}),$$

where the components of $\underline{\eta}^T(\underline{x}) = (\eta_1(\underline{x}), \ldots, \eta_k(\underline{x}))$ are k linearly independent continuous functions on some compact space χ, $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ is a vector of unknown parameters to be estimated and the error terms $\varepsilon(\underline{x})$ are assumed uncorrelated and homoscedastic.

1.1. Exact design problem

An *exact design* is defined by a set of experimental conditions $\underline{\widetilde{x}}_1, \ldots, \underline{\widetilde{x}}_N$, which are not necessarily distinct at which observations are to be taken.

If the $\underline{\widetilde{x}}_i$'s are taken on J distinct values $\underline{x}_j, j = 1, \ldots, J$ and if the point $\underline{x}_j$ appears n_j times in the design, $j = 1, \ldots, J$, we denote the design by

$$\lambda = \begin{Bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_J \\ n_1 & n_2 & \ldots & n_J \end{Bmatrix}.$$

Clearly $g_j = n_j/N$ will be the proportion of experimental runs at $\underline{x}_j$. Conversely given a probability measure $\xi(\underline{x})$ on χ an exact design of N runs can be approximately defined by $\lambda(\underline{x}) = N\xi(\underline{x})$. Rounding will be necessary to convert $\lambda(\underline{x})$ to a non-negative integer vector $\forall \underline{x} \in \chi$. The exact design problem can be viewed as one of determining these proportions optimally subject to them being rational. In contrast the approximate or continuous optimal design problem relaxes this condition. The calculus of the general equivalence theorem (see Whittle, 1973) can then be brought to bear in determining conditions of optimality; in particular conditions defining optimal weights. In this paper we use an idea presented by Torsney (2007) to deal with the problem of constructing both types of design. One aspect is the transformation of the exact design problem to one of determining optimally a different set of namely proportions of the design space. We first outline the approximate design problem.

1.2. Approximate design problem

Following Kiefer (1971) any probability measure, ξ on χ with finite support is called an approximate design; e.g.

$$\xi = \begin{Bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_J \\ p_1 & p_2 & \ldots & p_J \end{Bmatrix},$$

where $\xi(\underline{x}_j) = p_j$ define design weights and satisfy the conditions $p_j \geq 0, j = 1, \ldots, J$ and $\sum p_j = 1$. The $x_i's$ now are said to be the design or support points. The per observation information matrix is defined by

$$M(\xi) = \sum_{j=1}^{J} \eta(\underline{x}_j)\eta^T(\underline{x}_j)p_j = VPV^T,$$

where $P = \mathrm{diag}(p_1, p_2, \ldots, p_J)$ and the ith column of the matrix V is $\eta(\underline{x}_i)$ denoted by $\underline{v}_i = \underline{v}(\underline{x}_i)$. That is $M(\xi) = \sum_j I(\underline{\theta}, x_j)p_j$, where $I(\underline{\theta}, \underline{x})$ is the expected information matrix of a single observations under the linear model, $I(\underline{\theta}, \underline{x}) = \underline{\eta}(\underline{x})\underline{\eta}^T(\underline{x})$.

We note that non-linear models of the form, $y(\underline{x}) = \eta^T(\underline{x}, \underline{\theta}) + \varepsilon(\underline{x})$ can be subsumed here with $\varepsilon(\underline{x})$ as above. These models depend non-linearly on the values of the parameters $\underline{\theta}$. It is common to replace them by their linear approximations in terms of a Taylor expansion about a prior point estimate $\underline{\theta}^0$. In this way the corresponding information matrix is $M(\xi) = \sum_{j=1}^{N} g(\underline{x}_j)g^T(\underline{x})_j p_j$ where $g(\underline{x}) = (\partial\eta(\underline{x}, \underline{\theta})/\partial\underline{\theta})_{\underline{\theta}=\underline{\theta}^0}$. If we have heteroscedastic errors with variances known up to a constant, then $M(\xi)$ has the form $M(\xi) = \sum_{j=1}^{N} g(\underline{x}_j)g^T(\underline{x})_j \omega_j p_j$. Also for a class of generalized linear models the information matrix has the same form with weights ω_j depending on the linear predictor; see Ford et al. (1992).

For an N-point exact design ξ_N we can without loss of generality assume, $J = N, p_j = 1/N$, so

$$M(\xi_N) = \frac{1}{N} V V^T \propto V V^T.$$

An optimal design maximizes an appropriate function of the information matrix, say $\phi(\xi) = \Psi[M(\xi)]$; these functions are homogeneous and concave. Standard criteria are D-optimality, c-optimality and A-optimality which, respectively, maximize

$\Psi[M(\xi)]=\log\det[M(\xi)]$, $\Psi[M(\xi)]=-\underline{c}^T M^{-1}(\xi)\underline{c}$, $\Psi[M(\xi)]=-\mathrm{tr}(M^{-1})$. We note that in the case of the approximate design problem, we have, for fixed $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_J$, an example of the following type of problem:

Maximize a criterion $\phi(\underline{p})$ over $\mathfrak{P}=\{\underline{p}=(p_1,\ldots,p_J): p_j\geq 0, \sum p_j=1\}$.

Conditions of optimality are needed to identify when a design $\xi^*(\underline{x})=\underline{p}^*$ is optimal for a given problem. We define the necessary first-order conditions in terms of point to point directional derivatives. The directional derivative $F_\phi(\underline{p},\underline{q})$ of a criterion function $\phi(\cdot)$ at $\underline{p}$ in the direction of $\underline{q}$ is defined as

$$F_\phi(\underline{p},\underline{q})=\lim_{\varepsilon\downarrow 0}\frac{\phi\{(1-\varepsilon)\underline{p}+\varepsilon\underline{q}\}-\phi(\underline{p})}{\varepsilon}=\left.\frac{\partial\phi}{\partial\varepsilon}\right|_{\varepsilon=0^+}.$$

This derivative exists even if $\phi(\cdot)$ is not differentiable but if it is then the directional derivative can be written as follow:

$$F_\phi(\underline{p},\underline{q})=(\underline{q}-\underline{p})'\frac{\partial\phi}{\partial\underline{p}}=\sum_{j=1}^{J}(q_j-p_j)d_j \quad \text{where } d_j=\frac{\partial\phi}{\partial p_j},\ j=1,\ldots,J.$$

Let $\underline{e}_j$ be the j th unit vector in $\mathbb{R}^J$. The vertex directional derivative of $\phi(\cdot)$ at $\underline{p}$ is defined to be $F_j=F_\phi(\underline{p},\underline{e}_j)=d_j-\sum p_j d_j$. Note that $F_\phi(\underline{p},\underline{q})=\sum_j q_j F_j$. From this it follows that if $\phi(\cdot)$ is differentiable at $\underline{p}^*$, a necessary condition for $\phi(\underline{p}^*)$ to be a local maximum is

$$F_j^*=F_\phi(\underline{p}^*,\underline{e}_j)=\begin{cases}=0 & \text{for } p_j^*>0,\\ \leq 0 & \text{for } p_j^*=0.\end{cases}$$

Furthermore if $\phi(\cdot)$ is concave on its feasible region then this condition is both necessary and sufficient for optimality. This of course is the general equivalence theorem in the design context (Whittle, 1973; Kiefer, 1974).

The efficiency of a design ξ with respect to the criterion Ψ will be

$$\text{eff}=\frac{\Psi[M(\xi)]}{\Psi[M(\xi^*)]}.$$

Thus if a design has 70% efficiency then the optimal design ξ^* will produce the same precision as ξ with 30% fewer observations.

In this paper we show how a new approach combined with a multiplicative algorithm can be used to construct optimal designs in a wide variety of situations. We provide an exploration of the consequences of this idea with numerous examples. The approach involves transforming the values of design points to proportions of the design interval. We then choose these proportions optimally to determine exact designs or choose both these proportions and design weights optimally to determine approximate designs. Section 2 gives a summary of the properties of the multiplicative algorithm and a brief discussion of its history. In Section 3 the new approach combined with the algorithm is presented in order to compute exact designs and designs for models with correlated observations. Section 3 also addresses the computation of designs with two-factor models where the construction of optimal designs is more complicated. (See Schwabe, 1996). In Section 4 further modifications and extensions of the new approach of Section 3 are discussed in order to calculate approximate optimal designs. The last section provides conclusions and comments on most of the results.

2. Algorithms

2.1. Brief review

In general numerical techniques are needed to determine optimal designs, be they exact or approximate. Different techniques have been developed for each.

Approximate design theory is used to help with the more intractable N-trial exact design problems. These designs can be found by integer approximation to the approximate designs in several ways (see Fedorov, 1972, p. 157). However if N is not too large then a poor approximation is obtained. Many numerical algorithms for the construction of exact designs are based on exchange schemes and have been proposed to construct D-optimal designs (Fedorov, 1972, p. 164; Mitchell and Miller, 1970; Wynn, 1970), the DETMAX algorithm of Mitchell (1974), the modified Fedorov algorithm of Cook and Nachtsheim (1980), the KL-exchange algorithm of Atkinson and Donev (1989) which is another modification of the Fedorov's algorithm. Haines (1987) applied the annealing algorithm to construct D-optimal designs, Meyer and Nachtsheim (1995) described a cyclic coordinate-algorithm for constructing D-optimal and linear-optimal experimental designs for continuous design spaces.

For the case of D-optimality for constructing approximate designs algorithms have been provided and investigated by Wynn (1970, 1972) and Fedorov (1972, p. 102). They are called Wynn–Fedorov algorithms; so also did Silvey and Titterington (1973), Wu (1978) and Wu and Wynn (1978). To accelerate algorithms for D-optimum designs Pronzato (2003) and Harman and Pronzato (2007) proposed some inequalities in order to remove non-optimal support points.

So separate methods are available for the two types of design problem. Below we advocate a method which can be used for both problems. It exploits a multiplicative algorithm developed for the approximate case.

2.2. Multiplicative algorithms

By a multiplicative algorithm we mean one which updates all weights simultaneously according to the multiplicative form

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(d_j^{(r)}, \delta)}{\sum_{i=1}^{J} p_i^{(r)} f(d_i^{(r)}, \delta)} \quad \text{or} \quad p_j^{(r+1)} = \frac{p_j^{(r)} f(F_j^{(r)}, \delta)}{\sum_{i=1}^{J} p_i^{(r)} f(F_i^{(r)}, \delta)},$$

where $f(x,\delta)$ is positive, $\partial f(x,\delta)/\partial x > 0$ and, if $\delta = 0$, $f(x,\delta)$ is constant. Properties of this iteration, which can be seen at Torsney (2007), are

1. $\underline{p}^{(r)}$ is always feasible, i.e., $\underline{p}^{(r)} \in \mathfrak{P}$.
2. $F_\phi(\underline{p}^{(r)}, \underline{p}^{(r+1)}) \geq 0$, with equality when the $d_j's$ corresponding to nonzero $p_j's$ have a common value $d(=\sum p_i d_i)$, in which case $\underline{p}^{(r)} = \underline{p}^{(r+1)}$.
3. If $\delta = 0$ there is no change in $p^{(r)}$, given $f(x,\delta) =$ constant.
4. So the algorithm should be monotonic for small positive δ.
5. An iterate $\underline{p}^{(r)}$ is a fixed point of the iteration if derivatives $d_j^{(r)}$ corresponding to nonzero $p_j^{(r)}$are equal; i.e., if corresponding vertex directional derivatives $F_j^{(r)}$ are zero.

A proof of the second statement follows given that:

$$F_\phi(\underline{p}^{(r)}, \underline{p}^{(r+1)}) = \frac{\mathrm{Cov}(D, f(D,\delta))}{E(f(D,\delta))},$$

where D is a discrete random variable with probability distribution $P(D = d_j^{(r)}) = p_j^{(r)}$. It can then be argued that $\mathrm{Cov}(D, f(D,\delta)) > 0$ if $\partial f(D,\delta)/\partial D > 0$, while $E(f(D,\delta)) > 0$ if $f(D,\delta) > 0$.

This class of algorithm evolved from a result of Fellman (1974). This was devoted to linear criteria, not algorithms, but in effect he proved that $f(d,\delta) = d^\delta$ with $\delta = \frac{1}{2}$ yields monotonicity for c-optimality. We note that all design criteria have positive derivatives. Torsney (1983) extends this result to A-optimality, while Titterington (1976) proved monotonicity of $f(d,\delta) = d^\delta$ with $\delta = 1$ for D-optimality. This latter choice is also monotonic for finding the maximum likelihood estimators of the mixing weights, given data from a mixture of distributions. Indeed it is an EM algorithm; see Torsney (1977). Both choices also appear to be monotonic in determining, respectively, c-optimal and D-optimal conditional designs, i.e., in determining several optimizing distributions; see Martín-Martín et al. (2007).

Finally Silvey et al. (1978) is an empirical study of the choice of δ in $f(d,\delta) = d^\delta$. Note that this is a feasible choice for standard design criteria and the mixture likelihood since they enjoy positive partial derivatives.

Other choices of $f(\cdot,\delta)$ are needed if $\phi(\underline{p})$ can have negative derivatives, as in some maximum likelihood estimation problems, or if the partial derivatives d_j are replaced by the vertex directional derivatives F_j. Torsney (1988) considers the case of $f(d,\delta) = e^{\delta d}$, while objective bases for choosing $f(x,\delta)$ are to be found in Torsney and Alahmadi (1992), Torsney and Mandal (2004) and Mandal and Torsney (2006). Also Torsney and Mandal (2001) apply these algorithms to finding constrained optimal designs, while Mandal and Torsney (2006) explore their use in a clustering approach under which designs points are collected into sets. The algorithm is then used to determine an optimal marginal design across sets and optimal conditional designs within sets. This is similar in spirit to the approach now proposed.

In a new approach to determine optimal designs, in particular optimal support points of exact or approximate designs, we generate another example of the above problem, but one in which ϕ is not guaranteed to have positive derivatives; so we need to choose an $f(x,\delta)$, which is defined for negative x.

3. A novel approach for determining support points

In optimal experimental design observation are usually taken at a small subset of points in a continuous design interval. If these points were known or we could remove some of them the problem could be simplified; see Pronzato (2003) and Harman and Pronzato (2007). In Torsney (2007) a new idea is proposed for determining support points. Consider a linear model with one design variable x, to be selected from a design interval $\chi = [a,b]$. A simple case is the determination of the best N-point ϕ-optimal design. Let $x_1, x_2, \ldots, x_N$ be its support points. Let

$$W_t = \frac{x_t - x_{t-1}}{b-a}, \quad t = 1, \ldots, N+1,$$

where $x_0 = a$ and $x_{N+1} = b$.

We have linearly transformed from N variables to $(N+1)$ variables, $\underline{W}$, $\underline{W} = A^T\underline{x} + \underline{c}$ where $\underline{x} = (x_1, \ldots, x_N)^T$,

$$A = \frac{1}{b-a}[(I_N|0_N) - (0_N|I_N)],$$

with $0_N = (0, 0, \ldots, 0)^T \in \mathbb{R}^N$, I_N is identity matrix of order N and $\underline{c} = (1/(b-a))(-a, 0, 0, \ldots, 0, b) \in \mathbb{R}^{N+1}$. These must satisfy

$$W_h \geq 0, \quad \sum_{h=1}^{N+1} W_h = 1.$$

We note that it follows that these variables are invariant to prior linear transformations of $\underline{x}$; also $W_h = 0$ if $x_{h+1} = x_h$, i.e., two potentially distinct design points coincide.

Now consider maximizing the ϕ criterion with respect to $W_1, W_2, \ldots, W_{N+1}$, subject to these constraints.

Thus the optimization problem for a general criterion can be stated as a problem in the new variables W_h. We need to choose $W_1, \ldots, W_{N+1}$ optimally and our problem is of the form: maximize a criterion $\phi(\underline{W})$ over $\mathfrak{W} = \{\underline{W} = (W_1, \ldots, W_{N+1}) : W_h \geq 0, \sum_{h=1}^{N+1} W_h = 1\}$, where $\phi(\underline{W}) = \phi(A^T\underline{x} + \underline{c}) = \phi_x(\underline{x}) = \Psi[M(\xi_N)]$.

The first-order conditions for a local maximum are

$$F_h^* = F_\phi(\underline{W}^*, \underline{e}_h) = \begin{cases} = 0 & \text{for } W_h^* > 0, \\ \leq 0 & \text{for } W_h^* = 0, \end{cases} \tag{1}$$

where $\underline{e}_h$ is the h th unit vector in $\mathbb{R}^N$, $F_h = F_\phi(\underline{W}, \underline{e}_h) = d_h - \sum_{h=1}^N W_h d_h$ with

$$d_h = \frac{\partial \phi}{\partial W_h} = \sum_{i=1}^N \frac{\partial \phi_x}{\partial x_i}\frac{\partial x_i}{\partial W_h} = \sum_{i=1}^N \frac{\partial \Psi}{\partial x_i}\frac{\partial x_i}{\partial W_h}, \quad h = 1, \ldots, N \tag{2}$$

and

$$\frac{\partial \Psi}{\partial x_i} = \sum_r \sum_s \frac{\partial \Psi}{\partial M_{rs}}\frac{\partial M_{rs}}{\partial x_i}.$$

It is to be noted that we did not need to define W_{N+1}. Since there exists a one to one transformation between $x_1, \ldots, x_N$ and $W_1, \ldots, W_N$ the transformed criterion, say ϕ, does not explicitly depend on W_{N+1}. So it is a fact that $\partial\phi/\partial W_{N+1} = d_{N+1} = 0$.

We note two consequences of this transformation:

1. Condition (1) offers a set of first-order conditions for local optimality of exact designs, an advantage, we argue, over other approaches where equivalence theorems cannot be applied.
2. The above class of multiplicative algorithm can be used to determine optimizing $\underline{W}$, namely

$$W_h^{(r+1)} = \frac{W_h^{(r)} f(F_h^{(r)}, \delta)}{\sum_{t=1}^{N+1} W_t^{(r)} f(F_t^{(r)}, \delta)},$$

where F_h is the h th vertex directional derivative of $\phi(\cdot)$ at W and one choice of f is $f(F, \delta) = \Phi(\delta F)$ where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. We choose the first arguments of $f(\cdot, \delta)$ to be the vertex directional derivatives, so that we are centering partial derivatives with respect to $\underline{W}$ on zero.

We note that when the partial derivatives are always positive, as in the case of design criteria, we can consider $f(d, \delta)$ defined for positive d such as $f(x, \delta) = f(d, \delta) = d^\delta$. However here derivatives with respect to W_i are not guaranteed to be positive. Torsney and Mandal (2004, 2006) consider this issue and make recommendations for objective choices of $f(x, \delta)$. We opt for $f(F, \delta) = \Phi(\delta F)$, and always take $\underline{W}^{(0)}$ to be uniform.

3.1. Computing exact D-optimal designs

We now use these ideas for the construction of exact optimum designs. The exact D-optimum designs measure ξ_N^* maximizes the determinant of the matrix $M = V V^T$ where V^T is the design matrix. In order to compute the directional derivatives we need to calculate d_h which, as seen in (2), depends in turn on $\partial\Psi/\partial x_r$. In general

$$\frac{\partial \Psi}{\partial x_r} = \sum_i \sum_j \frac{\partial \Psi}{\partial M_{ij}}\frac{\partial M_{ij}}{\partial x_r} = \operatorname{tr}\left\{\frac{\partial \Psi}{\partial M}\frac{\partial M}{\partial x_r}\right\}. \tag{3}$$

Consider the D-optimal criterion function $\Psi[M] = \log\det(M)$,

$$\frac{\partial \Psi}{\partial x_r} = \sum_i \sum_j M_{ij}^{-1}\frac{\partial M_{ij}}{\partial x_r} = \operatorname{tr}\left\{M^{-1}\frac{\partial M}{\partial x_r}\right\}.$$

After some algebra (see Appendix) the vector of these derivatives is

$$\frac{\partial \Psi}{\partial \underline{x}} = 2\,\mathrm{Diag}(V^T M^{-1} D),$$

where $D = (\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_N)$ with $\underline{d}_r = \partial \underline{v}_r/\partial x_r$.

Table 1
Number of iterations needed to achieve $\max\{F_h\} \leq 10^{-n}$ and support points and efficiencies with respect to *D*-optimal global approximate design.

N	n	Iterations	Support points								Eff.
4	1	12	0.1156	0.4089	0.7439	0.9991					
	2	34	0.0894	0.3871	0.7374	0.9999					
	3	68	0.0823	0.3815	0.7347	1					1.0000
6	1	31	0.0563	0.2137	0.3915	0.5554	0.7733	0.9999			
	2	175	0.0367	0.2036	0.3999	0.6615	0.7839	1			
	3	519	0.0378	0.2057	0.4089	0.6924	0.7718	1			0.9560
8	1	31	0.0526	0.1286	0.2953	0.4021	0.5404	0.7166	0.8220	0.9999	
	2	311	0.0531	0.0812	0.2779	0.3648	0.4812	0.7282	0.7613	1	
	3	851	0.0621	0.0735	0.2718	0.3733	0.4653	0.7362	0.7496	1	0.9353

Note: In all examples we start from uniform values of $W_h^{(0)}$ (and latter of $p_j^{(0)}$).

Example 1. Wynn's quadratic/trigonometric example.

We consider the trigonometric model proposed by Wynn (1972), $E[y(x)]=f^T(x)\theta$, where the vector function $f(x)=(x, x^2, \sin 2\pi x, \cos 2\pi x)^T$ and the design space is the interval $\chi=[0,1]$. We run the algorithm to find the best N-point *D*-optimal exact design for $N=4,6,8$ with $f(F,\delta)=\Phi(\delta F)$, $\delta=0.1$. In Table 1 we report for $n=1,2,3$ the number of iterations needed to achieve $\max\{F_h\} \leq 10^{-n}$ and the support points.

We note that the global *D*-optimal design has four support points with equal weights. This is the design to which we have converged in the case of $N=4$. We cannot converge to this design in the case $N=6$ equally weighted points since this is not a multiple of 4, whereas we could have done so when $N=8$. This has not happened, an issue that we address in the Discussion. However the *D*-efficiency of the 8 point design is 93.5%.

3.2. Two-factor models

In many practically relevant experimental situations models with more than one factor are more appropriate than models with a single factor. Thus in a wide range of technical and industrial applications these models are suitable (Schwabe, 1996).

The calculation of optimum designs for such models is more complicated than for models with one factor.

For illustrative purposes we consider the second-order polynomial regression model in order to extend the new approach of the multiplicative algorithm presented above.

Example 2. Second-order polynomial regression model.

Consider the non-additive model for second-order response surface in two factors

$$E[y(\underline{x})] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2, \quad (x_1, x_2) \in \chi_1 \times \chi_2 = [-1,1]^2.$$

Here we have a two dimensional design space. We need to extend our *W*-transformation to this case. To avoid imposing constraints on the design we consider transforming x_1-values, say $x_i^{(1)}$, $i=1,\ldots,N$, to a marginal distribution or set of weights W_{1h}, $h=1,\ldots,N+1$, and transforming the x_2-values paired with each value of x_1 to a conditional distribution or set of weights $W_{2s|1i}$. Since the model is symmetric in x_1 and x_2, the roles of the two variables could be reversed without loss or gain. We wish to choose all these conditional and marginal distributions optimally. In the interests of simplicity we consider the case of one x_2-value, namely $x_{j|i}^{(2)}$ matched with each $x_i^{(1)}$ so that $j=1$, for all i. This is illustrated in Fig. 1 and leads to the following $N+1$ simultaneous multiplicative iterations:

$$W_{1h}^{(r+1)} = \frac{W_{1h}^{(r)} f(F_{1h}^{(r)}, \delta_1)}{\sum_{t=1}^{N+1} W_{1t}^{(r)} f(F_{1t}^{(r)}, \delta_1)}, \quad h=1,\ldots,N+1,$$

$$W_{2s|1i}^{(r+1)} = \frac{W_{2s|1i}^{(r)} f(F_{2s|1i}^{(r)}, \delta_2)}{\sum_{l=1}^{2} W_{2l|1i}^{(r)} f(F_{2l|1i}^{(r)}, \delta_2)}, \quad s=1,2, \;\; i=1,\ldots,N.$$

Table 2 shows the exact optimum design found for $N=6,7,8,9$. We find the same design as in Atkinson et al. (2007, p. 170). We have also explored the cases $N=10,11,12$. We note that in these cases and as illustrated in Table 2, the simplifying assumption of one x_2-value matched with one x_1-value does not prevent convergence to designs with replication of at least one value of both variables.

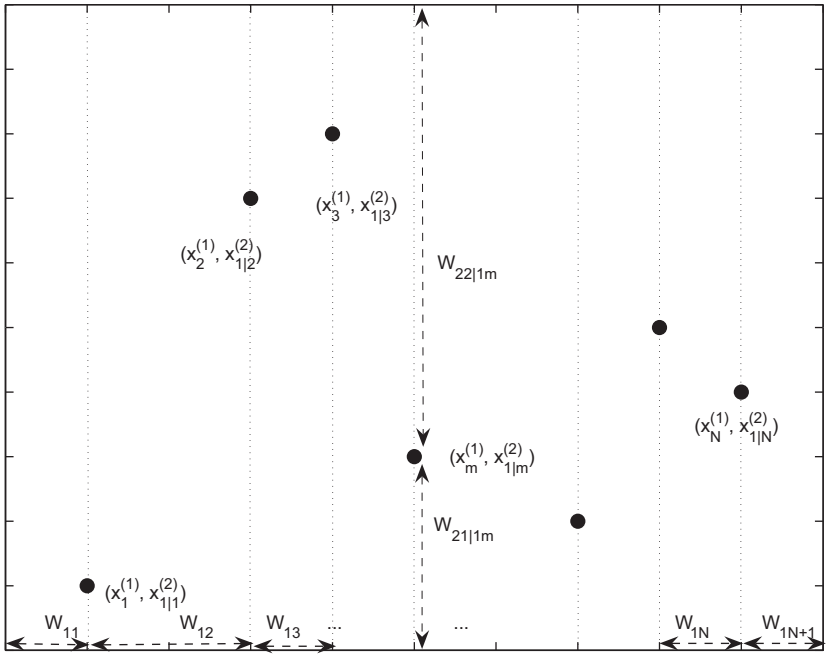

Fig. 1. Partition of the design space in models with two factors when design points are considered to be distinct.

Table 2
Optimal exact designs for second-order response surface in two factors.

N	6		7		8		9	
$\lvert M(\xi)\rvert$	5.590		6.888		7.767		8.553	
Support points	−1	1	−1	−1	−1	−1	−1	−1
	−1	−0.393	−1	−0.067	−1	1	−1	0
	−0.396	−1	−1	1	−0.215	0	−1	1
	0.131	0.131	−0.070	−1	0.082	−1	0	−1
	1	−1	0.098	0.091	0.082	1	0	0
	1	1	1	−1	1	−1	0	1
			1	1	1	0	1	−1
					1	1	1	0
							1	1

3.3. An algorithm for models with correlated observations

Methods for the construction of optimal experimental designs for models under the presence of correlated errors have been studied in the literature. Thus Brimkulov et al. (1980) proposed an exchange algorithm to compute D-optimal designs. Näther (1985) suggested some improvements. Müller and Pázman (1999) suggested an alternative algorithm based upon a new interpretation of design measures. López Fidalgo et al. (2008) modified Brimkulov's algorithm when some of the factors in the model are not under the control of the experimenter.

In this section we propose the above transformation and multiplicative algorithm in order to compute D-optimal designs in the presence of correlation.

We assume an isotropic covariance structure on the responses, say $\mathrm{Cov}(y_i, y_j) = \sigma^2 C(|x_i - x_j|, r)$. Let C be the covariance matrix of the observations with these entries. As in López Fidalgo et al. (2008) two design approaches are possible, either considering covariance parameters as a nuisance quantity, assumed known or as parameters of interest which need to be estimated. For the first approach the inverse of the Fisher information matrix has the standard meaning as being asymptotically proportional to the covariance matrix of the estimators. Nevertheless, within the second approach this property is not guaranteed in general. Thus, optimization based on the Fisher information matrix may be inappropriate.

In any case the traditional approximate design theory cannot be applied here any more. This is because we should not have more than one observation at each design point, since two such observations are perfectly correlated. So exact designs must be considered and the new approach and algorithm can again be applied.

Table 3
Number of iterations needed to achieve $\max\{F_h\} \leq 10^{-n}$ and determinant of the information matrix.

r	1			5			50		
N	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
5	67 1.9144	84 1.9247	100 1.9249	69 3.3213	98 3.4239	114 3.4242	71 3.7756	135 3.9103	315 3.9203
10	113 3.0212	303 3.1468	364 3.1472	126 7.4327	340 10.4556	360 10.4557	200 9.7064	354 10.3728	568 10.4056
15	128 3.6061	583 3.7018	736 3.7026	195 16.7357	656 18.2962	706 18.2963	150 25.0320	463 28.0349	735 28.1817
20	111 3.9030	917 3.9875	1246 3.9886	153 23.5919	596 25.9392	643 25.9522	160 46.6411	601 52.5384	902 52.8201

Assuming $\sigma^2 = 1$ and $V^T = \{\eta_j(x_i)\} = \{\partial\eta(x_i,\theta)/\partial\theta^{(j)}\}$, as in Section 1.2, the information matrices for θ only and r only are, respectively (Stehlík, 2006)

$$M_\theta(N) = VC^{-1}V^T,$$
$$M_r(N) = \frac{1}{2}\mathrm{tr}\left\{C^{-1}\frac{\partial C}{\partial r}C^{-1}\frac{\partial C}{\partial r^T}\right\},$$

while, when both parameters are of interest,

$$M(N) = \begin{pmatrix} M_\theta(N) & \mathbf{0} \\ \mathbf{0} & M_r(N) \end{pmatrix}.$$

3.3.1. Trend parameters of interest only

In this case the function to be maximized is

$$\Psi[M_\theta(N)] = \log\det\{M_\theta(N)\}.$$

As in (3) we have to compute

$$\frac{\partial\Psi}{\partial\underline{x}} = \mathrm{tr}\left(M^{-1}\frac{\partial M}{\partial\underline{x}}\right).$$

The derivatives obtained are (see Appendix)

$$\frac{\partial\Psi}{\partial\underline{x}} = 2\mathrm{Diag}(C^{-1}V^TM^{-1}D - C^{-1}V^TM^{-1}VC^{-1}G),$$

where $D = (\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_k)$ with $\underline{d}_r = \partial\underline{v}_r/\partial x_r$ and $G = (\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_k)$ with $\underline{g}_r = \partial\underline{c}_r/\partial x_r$ and $C = (\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k)$.

Example 3. A consecutive chemical reaction.

Consider the model introduced by Box and Lucas (1959) of two consecutive first-order chemical reactions $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$. Given the initial concentrations of the substances A, B and C, the concentration of B as a function of time is given by

$$\eta(t,\underline{\theta}) = \frac{\theta_1}{\theta_1-\theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}), \quad t \in T = [0,30],$$

where θ_1 and θ_2 are the rates of the reaction.

Assume, as in Uciński and Atkinson (2004), that the reaction is run only once and the measurements are taken at different times during the same reaction. Thus a correlation structure between these observations must be appropriate; we assume $\mathrm{Cov}(\eta(t_1),\eta(t_2)) = e^{-r|t_1-t_2|}$.

The algorithm was applied for the cases of 5, 10, 15 and 20 measurements, respectively, and for different values of the covariance parameter, taking 0.7 and 0.2 as initial values of θ_1 and θ_2 and $\delta = \frac{1}{10}$. In Table 3 for $n = 1,2,3$ it is shown that the larger is N the larger the number of iterations needed for convergence.

In Fig. 2 we plot the determinant of the information matrices for $1 \leq r \leq 50$ in increments of 0.1. Each was determined using the algorithm. For $N = 10, 15, 20$ we see that there is a slight but noticeable decrease in the criterion values occurring approximately at $r = 6, 7, 8$, respectively. This may be an indication that a local optimum has been found or possibly the algorithm needs to be run for longer to achieve $\max\{F_h\} \leq 10^{-n}$ for n larger than 3. We also note that we are determining exact designs of a given size.

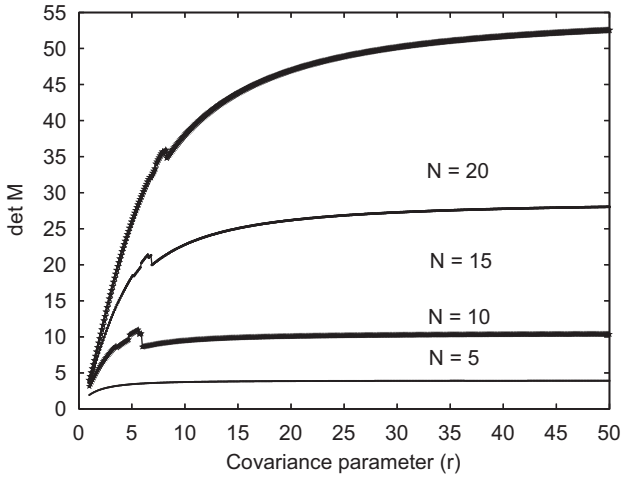

Fig. 2. Determinant of the information matrix of 5, 10, 15 and 20 points, respectively, against the values of the covariance parameter, r.

So at some correlation values there will be discrete changes to the support points. So for a given N it cannot be guaranteed that a design criterion will be a continuous increasing function of the correlation.

3.3.2. Covariance parameter of interest only

Since the covariance matrix depends on only one parameter, namely r, the only possible criterion to be maximized is

$$\Psi[M_r(L)] = \frac{1}{2}\mathrm{tr}\left\{C^{-1}\frac{\partial C}{\partial r}C^{-1}\frac{\partial C}{\partial r}\right\}.$$

Once again in order to apply the algorithm the derivatives $\partial\Psi/\partial x_r$ are required (see Appendix).

After some algebra we obtain

$$\frac{\partial\Psi}{\partial \underline{x}} = 2\,\mathrm{Diag}\left(C^{-1}\frac{\partial C}{\partial r}C^{-1}H\right) - 2\,\mathrm{Diag}\left(C^{-1}\frac{\partial C}{\partial r}C^{-1}\frac{\partial C}{\partial r}C^{-1}G\right),$$

where $H = (\underline{h}_1, \ldots, \underline{h}_k)$ with $\underline{h}_r = \partial \underline{m}_r/\partial x_r$ with $\partial C/\partial r = (\underline{m}_1, \ldots, \underline{m}_k)$ and $G = (\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_k)$ with $\underline{g}_r = \partial \underline{c}_r/\partial x_r$ and $C = (\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k)$.

Example 4. Exponential covariance structure.

Consider the model $E[y(x)] = \theta_1 + \theta_2 x,\ x \in \chi = [-1, 1]$. In this case we have computed D-optimal designs for the covariance parameter r under the exponential covariance structure $\mathrm{Cov}(y(x), y(z)) = \sigma^2 e^{-r|z-x|}$ if $z - x \neq 0$. As is shown in Fig. 3 these designs collapse to one point designs as others have experienced, e.g. Stehlík et al. (2008). One solution to this behavior is to extend the covariance function to include a nugget effect of the form

$$\mathrm{Cov}(y(x), y(z)) = \begin{cases} \sigma^2\alpha e^{-r|z-x|} & \text{if } z - x > 0, \\ \sigma^2(1-\alpha) & \text{if } z = x, \end{cases}$$

with $\sigma^2 = 1$.

3.3.3. Optimal design for both sets of parameters

We note that the criterion Ψ must depend on $M(L)$

$$M(L) = \begin{pmatrix} M_\theta(L) & \mathbf{0} \\ \mathbf{0} & M_r(L) \end{pmatrix}.$$

For the D-optimal case it is straightforward to obtain from the above results the derivatives needed for the algorithm, namely

$$\frac{\partial\Psi}{\partial x} = \frac{1}{|M_\theta(L)|}\frac{\partial|M_\theta(L)|}{\partial x} + \frac{1}{|M_r(L)|}\frac{\partial|M_r(L)|}{\partial x}.$$

Example 5. A consecutive chemical reaction.

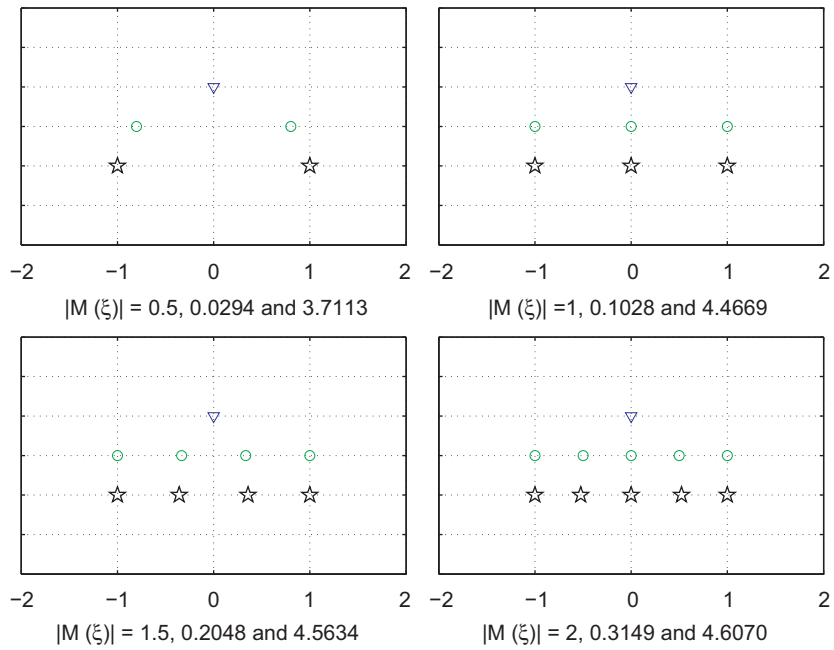

Fig. 3. Support of *D*-optimal designs when the covariance parameter is considered to be of interest without nugget effect ($\triangledown$) with nugget effect $\alpha = 0.5$ ($\circ$), and when only the parameters of the trend are of interest ($\star$), respectively, and their determinants of the information matrices.

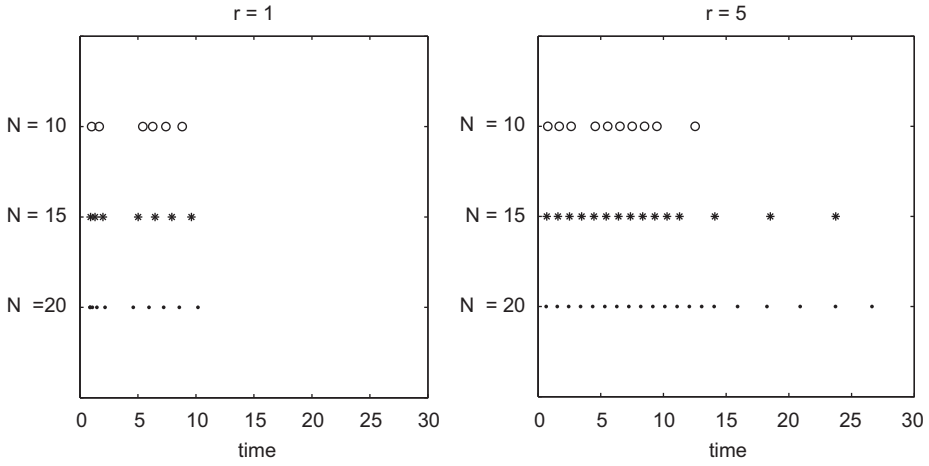

Fig. 4. Support of *D*-optimal designs when both sets of parameters are considered to be of interest for $n = 10, 15, 20$.

Consider the chemical model proposed in Example 3. with the exponential correlation structure. In this case, the purpose of the experiment is to obtain good estimates of the parameters θ_1, θ_2 and r. The algorithm was applied taking 10, 15 and 20 measurements. The nominal values chosen for the parameters of the trend were the same as in Example 3 and the values considered for r where 1 (strong correlation) and 5 (medium correlation). The algorithm readily found the optimal designs depicted in Fig. 4 using $f(F, \delta) = \Phi(\delta F)$, $\delta = 0.1$.

4. Simultaneous approach to optimal weight and support point determination

In this section we address the problem of determining both the support points and their design weights simultaneously. Although we do not know the number of support points, due to Carathedory's theorem it is known that there must exist an optimal design with at most $L = k(k+1)/2$ support points, where k is the number of parameters. So we initially assume L-support points. Under this framework we again have an optimization problem with respect to more than one distribution, one defined

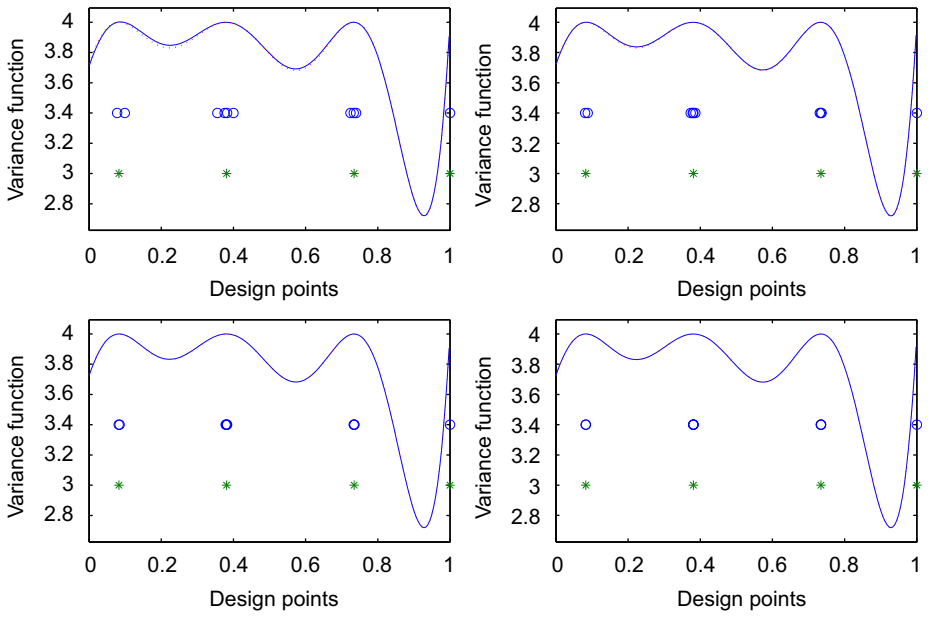

Fig. 5. Support points (*) and variance function (dotted) for D-optimal design and support points ($\circ$) and variance function (solid) for designs obtained by the algorithm when $\max\{F_h\}$ and $\max\{F_j\} \leq 10^{-n}$ for $n = 1, 2, 3$ and 4.

by the design weights and one by the W_i. We explore use of the following simultaneous iterations:

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(F_j^{(r)}, \delta_1)}{\sum_{i=1}^{J} p_i^{(r)} f(F_i^{(r)}, \delta_1)} \quad \text{where } f_p(F_p, \delta_1) = \Phi(\delta_1 F_p), \tag{4}$$

$$W_h^{(r+1)} = \frac{W_h^{(r)} f(F_h^{(r)}, \delta_2)}{\sum_{t=1}^{L+1} W_t^{(r)} f(F_t^{(r)}, \delta_2)} \quad \text{where } f_W(F_W, \delta_2) = \Phi(\delta_2 F_W). \tag{5}$$

However when partial derivatives with respect to design weights are positive we also consider $f(x, \delta_1) = f(d, \delta_1) = d^{\delta_1}$. The case $\delta_1 = 1$ leads to monotonic iterations when seeking D-optimal weights only.

Example 6. Wynn's quadratic/trigonometric example.

In order to study the performance of the simultaneous approach we study Wynn's linear model with linear, quadratic and trigonometric regression functions as in Example 1, but now we are interested not only in determining support points but also in determining design weights optimally. To determine the D-optimal design we start from an initial design of $L = 4(4+1)/2 = 10$ equidistant equally weighted support points (excluding the end points). We use the variation of iteration (4) and (5) which replaces $f_p(F_p, \delta_1)$ by $f_p(d, \delta_1) = d_p^{\delta_1}$, $\delta_1 = 1$. We run this algorithm until $\max_h \{F_h\}$ and $\max_j \{F_j\} \leq 10^{-n}$ for $n = 1, 2, 3$ and 4, respectively. The number of iterations needed to achieve this were 229, 640, 1853 and 5602, respectively.

Fig. 5 illustrates the support points for the optimal design and the design obtained after running the algorithm. Also we have plotted the variance function, $d(x, p) = \eta^T(x) M^{-1}(p) \eta(x)$ for both the optimal design equally weighted on four points and the design obtained with the algorithm. It is known that the variance function attains its maximum value (the number of the parameters) at the optimal support points. Clearly, at early iterations, the support points fall into groups. There are four "clusters of points" around the optimal design points which is consistent with the optimal design having four support points. Table 4 contains the support points and their weights.

5. Discussion

We have described a new approach to determine optimal designs, a unified one which can determine both exact and approximate optimal designs.

The approach involves transforming design point values to proportions of the design interval and choosing these proportions optimally to determine exact designs or choosing both these proportions and design weights optimally to determine approximate designs.

Table 4
Number of iterations needed to achieve $\max\{F_h, F_t\} \leq 10^{-n}$, support points, optimal weights and value of the criterion function.

n	Iterations	Design points Optimal weights										$\phi_D(\xi)$
1	229	0.0777 0.2260	0.0989 0.0242	0.3553 0.0024	0.3756 0.1383	0.3824 0.1087	0.4000 0.0005	0.7242 0.0098	0.7334 0.2280	0.7399 0.0122	1.0000 0.2500	7.0842e − 004
2	640	0.0809 0.2265	0.0880 0.0235	0.3733 0.0018	0.3793 0.1388	0.3810 0.1090	0.3863 0.0004	0.7312 0.0089	0.7340 0.2296	0.7362 0.0115	1.0000 0.2500	7.0879e − 004
3	1853	0.0820 0.2266	0.0844 0.0234	0.3786 0.0017	0.3804 0.1389	0.3809 0.1091	0.3826 0.0004	0.7334 0.0087	0.7343 0.2301	0.7350 0.0112	1.0000 0.2500	7.0883e − 004
4	5602	0.0824 0.2266	0.0832 0.0234	0.3802 0.0016	0.3807 0.1389	0.3809 0.1091	0.3814 0.0004	0.7341 0.0086	0.7343 0.2303	0.7346 0.0111	1.0000 0.2500	7.0883e − 004

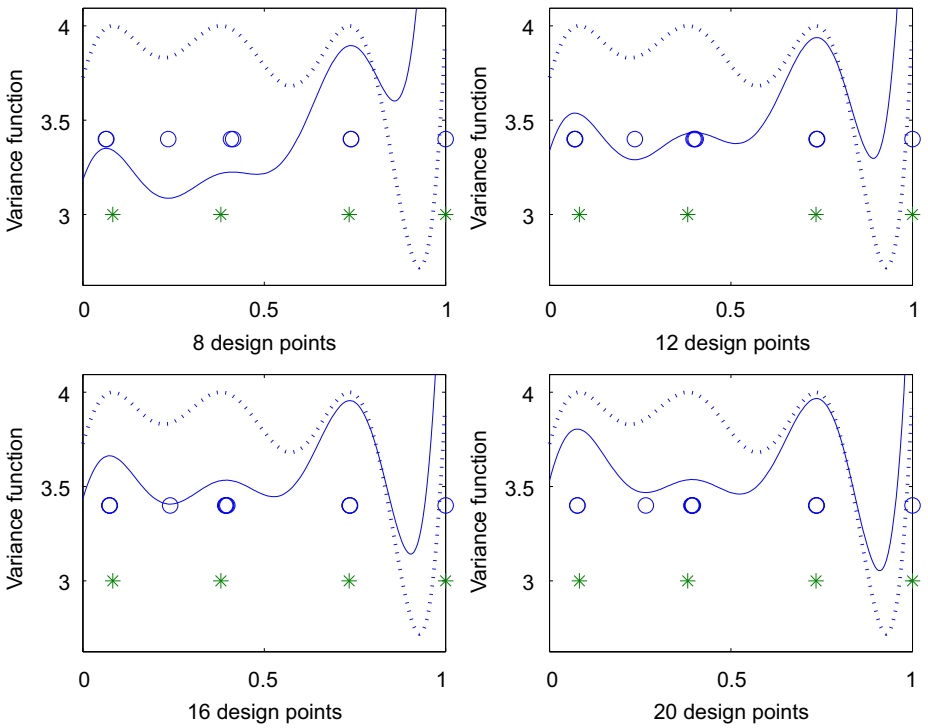

Fig. 6. Support points (*) and variance function (dotted) for D-optimal global design and support points (○) and variance function (solid) for designs obtained by the algorithm when $\max\{F_h\} \leq 10^{-4}$ for $N = 8, 12, 16$ and 20 design points.

We see one advantage of this approach being that we can use calculus to determine first-order conditions of optimality for exact designs; in effect, we are exploiting the general equivalence theorem of optimal designs (Whittle, 1973).

We see a second advantage being that we can recruit algorithms for determining optimal weights to determine optimal "design interval proportions". We have used multiplicative algorithms, which were originally developed to determine optimal design weights.

Although we have described the implementation for the determinant criteria we have also explored its application to other criteria.

We can say that we have found the optimum design in examples taken from the literature. In other cases we have found designs with high efficiencies and presumably local optima. Certainly there are instances when we have not converged to the global optimal design. We see this in Example 1 in which we have found the global D-optimal design when it has minimal equally weighted four-point support. When N is a multiple of the parameter dimension of the model the optimal exact design should replicate the approximate design. We did not converge to this design in Example 1 when $N = 8$. We explore this issue further in Fig. 6, which plots the support points and the variance function of the optimal design and of the design found by the algorithm attaining the condition $F_j \leq 10^{-4}$ in the cases $N = 8, 12, 16$ and 20. In each case the design found by the algorithm consists of four unequal sized clusters of support points in the neighborhood of each of the support points of the optimal approximate design and in the neighborhood of each their maximal turning points, plus one extra rogue design point at $x = 0.24$ which is at or close

to where all the variance functions have their first minimal turning point. This might suggest that the algorithm has converged to a stationary point with respect to the W_i, which is not a maximal one.

In both examples we can clearly draw conclusions from the turning points of these plots about the number of support points of the optimal approximate design. The size of the clusters is less informative. We do even better if we seek to determine optimal design points and weights simultaneously using iterations (4) and (5). We converge to the global optimal design in Example 6, assuming the model of Example 1, starting from an initial design of 10 equally weighted equidistant design points, 10 being the maximum number needed according to Carathéodory's Theorem.

This brings us to the issue of convergence. The rate of convergence can be fast to begin with but soon slows down, in terms of number of iterations, as happens with other algorithms such as the EM algorithm. Convergence depends, of course, on the choices of $f(x,\delta)$, x and δ. We have generally opted to take $x = F$, $\delta = 0.1$ and 0.01 and $f(F,\delta) = \Phi(\delta F)$, our primary aim being to explore the potential of our approach in a variety of problems. Run time lasts only seconds ranging from 0.01 to 1.09 s. As to proof of convergence no results are available in any context although some monotonicity results already cited may be relevant. Currently others are exploring the simple but long established case of $f(d,\delta) = d$ in the case of D-optimality; see Dette et al. (2007). The merits of our approach are that it is a unifying one for determining both approximate and exact designs and an advantage over other numerical approaches are that only first-order derivatives are needed.

Looking to the future we note the following extensions of our approach:

1. If we can identify some design points analytically e.g. endpoints as in the above example, the approach can be modified in a natural way. In general an extension would be to divide intervals between analytically established design points into "proportions" so that there would be several sets of proportions to determine optimally.
2. We have assumed the design space to be finite as is needed for standard linear models. However designs spaces need not be bounded for weighted linear and for non-linear models such as generalized linear models. Ford et al. (1992) converted design problems for generalized linear models to standardized weighted regression models. A particular class is binary regression models under which a response $y|x \sim Bi(1, F(z))$, $z = \alpha + \beta x$. In this case one possible transformation is $W_i = F(z_i) - F(z_{i-1})$. Note that the values of the $F(z_i)$ are of particular interest in binary regression.
3. There is scope for more work in the case of two or more design variables. A general approach for finding a design with no predetermined structure would be to take the design variables in some order and express a joint design as a product of a marginal one for the first variable and a sequence of conditional designs, one for each variable in the sequence; i.e., a conditional design for each variable given the values of the preceding variables in the sequence, one for each set of such values. The objective would be to determine all of these designs, marginal and conditional, optimally. Using our approach would mean finding several sets of "W_i's" optimally.
   A simpler scenario would be to restrict a design to a product one, so that the problem would reduce to finding optimal marginal designs, one for each variable. This would be appropriate if a balanced design is to be expected.
4. Our approach would also be appropriate for design problems in which support points are matrices of rank two as in the binary response design problem of Atkinson et al. (1995), in which there is only one design variable.

Appendix

Derivatives for D-optimal exact designs: Consider the criterion function $\Psi[M] = \log\det(M)$,

$$\frac{\partial \Psi}{\partial x_r} = \sum_i \sum_j M^{-1}_{ij} \frac{\partial M_{ij}}{\partial x_r} = \mathrm{tr}\left\{ M^{-1} \frac{\partial M}{\partial x_r} \right\}.$$

Since $M = VV^T$

$$\frac{\partial M}{\partial x_r} = \frac{\partial V}{\partial x_r} V^T + V \frac{\partial V^T}{\partial x_r}.$$

Further $V = (\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_k)$, where $\underline{v}_j = \underline{v}(x_j)$ i.e., only $\underline{v}_j$ depends on x_j. So

$$\frac{\partial V}{\partial x_r} = (\underline{0}, \underline{0}, \ldots, \underline{d}_r, \underline{0}, \ldots, \underline{0}) = \underline{d}_r \underline{e}_r^T \quad \text{where } \underline{d}_r = \frac{\partial \underline{v}_r}{\partial x_r}$$

and $\partial V^T / \partial x_r = \underline{e}_r \underline{d}_r^T$. Finally

$$\frac{\partial M}{\partial x_r} = \underline{d}_r \underline{e}_r^T V^T + V \underline{e}_r \underline{d}_r^T$$

and

$$\frac{\partial \Psi}{\partial x_r} = \mathrm{tr}[M^{-1}(\underline{d}_r \underline{e}_r^T V^T + V \underline{e}_r \underline{d}_r^T)] = 2(\underline{e}_r^T V^T M^{-1} \underline{d}_r).$$

Thus,

$$\frac{\partial \Psi}{\partial \underline{x}} = 2\,\mathrm{Diag}(V^T M^{-1} D),$$

where $D = (\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_k)$.

Derivatives when parameters of the trend are of interest only:

$$\frac{\partial M}{\partial x_r} = \frac{\partial V}{\partial x_r} C^{-1} V^T + V C^{-1} \frac{\partial V^T}{\partial x_r} - V C^{-1} \frac{\partial C}{\partial x_r} C^{-1} V^T,$$

$$\frac{\partial V}{\partial x_r} = (\underline{0}, \underline{0}, \ldots, \underline{d}_r, \underline{0}, \ldots, \underline{0}) = \underline{d}_r \underline{e}_r^T \quad \text{where } \underline{d}_r = \frac{\partial \underline{v}_r}{\partial x_r},$$

$$\frac{\partial V^T}{\partial x_r} = \underline{e}_r \underline{d}_r^T,$$

$$\frac{\partial C}{\partial x_r} = \underline{g}_r \underline{e}_r^T + \underline{e}_r \underline{g}_r^T \quad \text{where } C = (\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k) \text{ and } \underline{g}_r = \frac{\partial \underline{c}_r}{\partial x_r}.$$

Then we can write

$$\frac{\partial M}{\partial x_r} = \underline{d}_r \underline{e}_r^T C^{-1} V^T + V C^{-1} \underline{e}_r \underline{d}_r^T - V C^{-1} (\underline{g}_r \underline{e}_r^T + \underline{e}_r \underline{g}_r^T) C^{-1} V^T$$

and

$$\begin{aligned}\frac{\partial \Psi}{\partial x_r} &= \mathrm{tr}[M^{-1}(\underline{d}_r \underline{e}_r^T C^{-1} V^T + V C^{-1} \underline{e}_r \underline{d}_r^T - V C^{-1} (\underline{g}_r \underline{e}_r^T + \underline{e}_r \underline{g}_r^T) C^{-1} V^T)] \\ &= 2(\underline{e}_r^T C^{-1} V^T M^{-1} \underline{d}_r - \underline{e}_r C^{-1} V^T M^{-1} V C^{-1} \underline{g}_r).\end{aligned}$$

Derivatives when parameters of the covariance are of interest only:

$$\begin{aligned}\frac{\partial \Psi}{\partial x_r} &= \mathrm{tr}\left\{ C^{-1} \frac{\partial C}{\partial r} C^{-1} \frac{\partial \frac{\partial C}{\partial r}}{\partial x_r} - C^{-1} \frac{\partial C}{\partial r} C^{-1} \frac{\partial C}{\partial r} C^{-1} \frac{\partial C}{\partial x_r} \right\} \\ &= \mathrm{tr}\left\{ C^{-1} \frac{\partial C}{\partial r} C^{-1} (h_r e_r^T + e_r h_r^T) - C^{-1} \frac{\partial C}{\partial r} C^{-1} \frac{\partial C}{\partial r} C^{-1} (g_r e_r^T + e_r g_r^T) \right\} \\ &= 2 e_r^T C^{-1} \frac{\partial C}{\partial r} C^{-1} h_r - 2 e_r^T C^{-1} \frac{\partial C}{\partial r} C^{-1} \frac{\partial C}{\partial r} C^{-1} g_r.\end{aligned}$$

References

Atkinson, A.C., Donev, A.N., 1989. The construction of exact *D*-optimum experimental designs with application to blocking response surface designs. Biometrika 76 (3), 515–526.

Atkinson, A.C., Demetrio, C.G.B., Zocchi, S.S., 1995. Optimum dose levels when males and females differ in response. Applied Statistics 44 (2), 213–226.

Atkinson, A.C., Donev, A.N., Tobias, R., 2007. Optimum Experimental Designs with SAS. Oxford University Press, Oxford.

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

Brimkulov, U.N., Krug, G.K., Savanov, V.L., 1980. Numerical construction of exact experimental designs when the measurements are correlated. Zavodskaya Laboratoria (Industrial Laboratory) 36, 435–442 (in Russian).

Cook, R.D., Nachtsheim, C.J., 1980. A comparison of algorithms for constructing exact *D*-optimal designs. Technometrics 22, 315–324.

Dette, H., Pepelyshev, A., Zhigljavsky, A., 2007. Improving updating rules in multiplicative algorithms for computing *D*-optimal designs. Sonderforschungsbereich (SFB) 475 ⟨http://hdl.handle.net/2003/24796⟩.

Fedorov, V.V., 1972. In: Studden, W.J., Klimko, E.M. (Eds. and Transl.), Theory of Optimal Experiments. Academic Press, New York.

Fellman, J., 1974. On the allocation of linear observations (Thesis). Comment. Phys. Math. 44, 27–78.

Ford, I., Torsney, B., Wu, C.F.J., 1992. The use of a canonical form in the construction of locally optimal designs for non-linear problems. Journal of the Royal Statistical Society B 2, 569–583.

Haines, L.M., 1987. The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. Technometrics 29, 439–447.

Harman, R., Pronzato, L., 2007. Improvements on removing nonoptimal support points in *D*-optimum design algorithms. Statistics & Probability Letters 77, 90–94.

Kiefer, H., 1971. The role of symmetry and approximation in exact design optimality. In: Gupta, S.S., Yachel, J. (Eds.), Statistical Decision Theory and Related Topics. Academic Press, New York, pp. 109–118.

Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). Annals of Statistics 2, 849–879.

López Fidalgo, J., Martín-Martín, R., Stehlík, M., 2008. Marginally restricted *D*-optimal designs for correlated observations. Journal of Applied Statistics 35 (6), 617–632.

Mandal, S., Torsney, B., 2006. Construction of optimal designs using a clustering approach. Journal of Statistical Planning and Inference 136, 1120–1134.
Martín-Martín, R., Torsney, B., López Fidalgo, J., 2007. Construction of marginally and conditionally restricted designs using multiplicative algorithms. Computational Statistics and Data Analysis 51 (12), 5547–5561.
Meyer, R.K., Nachtsheim, C.J., 1995. The coordinate-exchange algorithm for construction exact optimal experimental designs. Technometrics 37, 60–69.
Mitchell, T.J., Miller Jr., F.L., 1970. Use of design repair to construct designs for special linear model. Mathematical Division Annual Progress Report (ORNL-4661), Oak Ridge National Laboratory, Oak Ridge, TN.
Mitchell, T.J., 1974. An algorithm for construction of *D*-optimal experimental designs. Technometrics 16, 203–210.
Müller, W.G., Pázman, A., 1999. An algorithm for the computation of optimum designs under a given covariance structure. Computational Statistics 14.
Näther, W., 1985. Effective Observations of Random Fields. Teubner/Texte zur Mathematik, Leipzig.
Pronzato, L., 2003. Removing non-optimal support points in *D*-optimum design algorithms. Statistics & Probability Letters 63, 223–228.
Schwabe, R., 1996. Optimum designs for multi-factor models. In: Lecture Notes in Statistic, vol. 113. Springer, Berlin.
Stehlík, M., 2006. Some properties of exchange design algorithms under correlation. Research Report Series no. 28, Department of Statistics and Mathematics, University of Economics and Business Administration, Vienna.
Stehlík, M., Rodríguez-Díaz, J.M., Müller, W.G., López-Fidalgo, J., 2008. Optimal allocation of bioassays in the case of parameterized covariance functions: an application to lung's retention of radioactive particles. Test 17 (1), 56–68.
Silvey, S.D., Titterington, D.M., 1973. A geometric approach to optimal design theory. Biometrika 60, 21–32.
Silvey, S.D., Titterington, D.M., Rosney, B., 1978. An algorithm for optimal designs on a finite design space. Communication in Statistics A 7, 1379–1389.
Titterington, D.M., 1976. Algorithms for computing *D*-optimal designs on a finite design space. In: Conference on Information Sciences and Systems, Department of Electrical Engineering, Johns Hopkins University of Baltimore, MD, pp. 213–216.
Torsney, B., 1977. Contribution to discussion of a paper by Dempster, Laird and Rubin. Journal of Royal Statistical Society 39, 22–27.
Torsney, B., 1983. A moment inequality and monotonicity of an algorithm. In: Kortanek, K.O., Fiacco, A.V. (Eds.), Proceedings of the International Symposium on Semi-Infinite Programming and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 215. University of Texas, Austin, pp. 249–260.
Torsney, B., 1988. Computing optimising distributions with applications in design, estimation and image processing. In: Dodge, Y., Fedorov, V.V., Wynn, H.P. (Eds.), Optimal Design and Analysis of Experiments. Elsevier Science Publishers B.V., North-Holland, Amsterdam, pp. 361–370.
Torsney, B., Alahmadi, A.M., 1992. Further developments of algorithms for constructing optimizing distributions. In: Federov, V., Muller, W.G., Vuchkov, I.N. (Eds.), Model Oriented Data Analysis, Proceedings of the second IIASA-Workshop, St. Kyrik, Bulgaria, 1900. Physica Verlag, pp. 121–129.
Torsney, B., 2007. W-iterations and ripples therefrom. In: Pronzato, L., Zhigljavsky, A. (Eds.), Optimal Design and Related Areas in Optimization and Statistics, vols. 1–12. Springer, New York, pp. 203–212.
Torsney, B., Mandal, S., 2001. Construction of constrained optimal designs. In: Atkinson, A., Bogacka, B., Zhigljavsky, A.A. (Eds.), Optimum Design 2000. Kluwer Academic Publishers, Dordrecht, pp. 141–152.
Torsney, B., Mandal, S., 2004. Multiplicative algorithms for constructing optimizing distributions: further developments. In: Bucchiano, A.D., Lauter, H., Wynn, H.P. (Eds.), Moda 7—Advances in Model Oriented Design and Analysis, Proceedings of the seventh International Workshop on Model Oriented Design and Analysis, Heeze, Netherlands, June 14–18, 2004, pp. 143–150.
Torsney, B., Mandal, S., 2006. Two classes of multiplicative algorithms for construction optimizing distributions. Computational Statistics and Data Analysis 51 (3), 1591–1601.
Uciński, D., Atkinson, A.C., 2004. Experimental design for time-dependent models with correlated observations. Studies in Nonlinear Dynamics & Econometrics 8(2), article 13.
Whittle, P., 1973. Some general points in the theory of optimal experimental designs. Journal of the Royal Statistical Society B 35, 123–130.
Wu, C.F.J., Wynn, H.P., 1978. The convergence of general step-length algorithms for regular optimum design criteria. Annals of Mathematical Statistics 6, 1273–1285.
Wu, C.F.J., 1978. Some iterative procedures for generating non/singular optimal designs. Communications in Statistics A 7, 1399–1412.
Wynn, H., 1970. The sequential generation of *D*-optimum experimental designs. Annals of Mathematical Statistics 41, 1655–1664.
Wynn, H., 1972. Results in the theory and construction of *D*-optimum experimental designs. Journal of the Royal Statistical Society B 34, 133–147.