Research on design of experiments has contributed, and continues to contribute, to the development of Statistical Sciences and to a myriad of data-based physical sciences. CSDA has long been at the forefront of publishing work in these areas. To this end a 2014 Special Issue (eds. Gilmour and Payne) was dedicated to ‘algorithms for design of experiments’, acknowledging that computational methods play a central role in the construction and validation of designs. For the current Special Issue, contributions were invited in a wide range of topics, covering a broad spectrum of design philosophies, methodologies and applications.

Some of the published papers are devoted to problems related to model uncertainty and discrimination. Dette et al. (2017a) focus on Fourier regression models. McGree (2017) uses the total entropy utility function for model discrimination and parameter estimation in Bayesian design. Designs considerations when making model selection for generalised linear models are studied by Woods et al. (2017).

Central to the study of model uncertainty is the issue of design robustness. Konstantinou et al. (2017) propose robust designs for survival trials. The robustness of various designs to missing observations is studied by Smucker et al. (2017) and by da Silva et al. (2017).

Dette et al. (2017b) consider further implications brought by dealing with correlated observations, and Rodríguez-Díaz (2017) proposes methods for constructing c-optimal designs in such situations. Atkinson and Biswas (2017) present novel ideas for adaptive designs for clinical trials with multivariate or longitudinal responses.

The work of Kobilinsky et al. (2017) provides solutions for computer generation for generalised regular factorial designs. The competitive algorithm of Masoudi et al. (2017) helps to find minimax and standardised maximin optimal designs.

Gauthier and Pronzato (2016) propose improved methods for construction of designs for estimating random-field interpolation models. Designing combined physical and computer experiments to maximise prediction accuracy is studied by Leatherman et al. (2017). An application of optimal design theory in a reliability industrial experiment is described by Rivas-López et al. (2017).

All papers present valuable contributions to the statistical theory and practice of design. They illustrate the continuing role to be played by thoughtful experimentation, at a time when mere scattershot sampling of ‘big data’ is becoming all too fashionable.

References


http://dx.doi.org/10.1016/j.csda.2017.05.012
0167-9473/© 2017 Published by Elsevier B.V.

Further Reading


The Special Issue Editors
Alexander N. Donev *
School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK
E-mail address: a.n.donev@manchester.ac.uk.

Jesús López-Fidalgo
Institute for Culture and Society (ICS), University of Navarre, Spain
E-mail address: fidalgo@unav.es.

Douglas P. Wiens
Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, T6E 2S2, Canada
E-mail address: doug.wiens@ualberta.ca.

* Corresponding editor.
T-optimal discriminating designs for Fourier regression models

Holger Dette\textsuperscript{a,}\textsuperscript{*}, Viatcheslav B. Melas\textsuperscript{b}, Petr Shpilev\textsuperscript{b}

\textsuperscript{a} Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany
\textsuperscript{b} St. Petersburg State University, Department of Mathematics, St. Petersburg, Russia

A R T I C L E   I N F O

Article history:
Received 23 December 2015
Received in revised form 13 June 2016
Accepted 20 June 2016
Available online 29 June 2016

Keywords:
T-optimal design
Model discrimination
Linear optimality criteria
Chebyshev polynomial
Trigonometric models

A B S T R A C T

The problem of constructing T-optimal discriminating designs for Fourier regression models is considered. Explicit solutions of the optimal design problem for discriminating between two Fourier regression models, which differ by at most three trigonometric functions, are provided. In general, the T-optimal discriminating design depends in a complicated way on the parameters of the larger model, and for special configurations of the parameters T-optimal discriminating designs can be found analytically. Moreover, in the remaining cases this dependence is studied by calculating the optimal designs numerically. In particular, it is demonstrated that D- and D\textsubscript{s}-optimal designs have rather low efficiencies with respect to the T-optimality criterion.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The problem of identifying an appropriate regression model in a class of competing candidate models is one of the most important problems in applied regression analysis. Nowadays, it is well known that a well designed experiment can improve the performance of model discrimination substantially, and several authors have addressed the problem of constructing optimal designs for this purpose. The literature on designs for model discrimination can roughly be divided into two parts. Hunter and Reiner (1965) and Stigler (1971) considered two nested models, where the extended model reduces to the “smaller” model for a specific choice of a subset of the parameters. The optimal discriminating designs are then constructed such that these parameters are estimated most precisely. Since these fundamental papers several authors have investigated this approach in various regression models (see Hill, 1978; Studden, 1982; Spruill, 1990; Dette, 1994, 1995; Dette and Haller, 1998, Song and Wong, 1999; Zen and Tsai, 2004; Biedermann et al., 2009 among many others). The second line of research was initialized in a fundamental paper of Atkinson and Fedorov (1975a), who introduced the T-optimality criterion for discriminating between two competing regression models. Since the introduction of this criterion, the problem of determining T-optimal discriminating designs has been considered by numerous authors (see Atkinson and Fedorov, 1975b; Ucinski and Bogacka, 2005; Dette and Titoff, 2009; Atkinson, 2010; Tommasi and López-Fidalgo, 2010 or Wiens, 2009, 2010 among others). The T-optimal design problem is essentially a minimax problem, and – except for very simple models – the corresponding optimal designs are not easy to find and have to be determined numerically in most cases of practical interest. On the other hand, analytical solutions are helpful for a better understanding of the optimization problem and can also be used to validate numerical procedures for the construction of optimal designs. Some explicit solutions of

\* Corresponding author.
E-mail address: holger.dette@rub.de (H. Dette).

http://dx.doi.org/10.1016/j.csda.2016.06.010
0167-9473/© 2016 Elsevier B.V. All rights reserved.
the $T$-optimal design problem for discriminating between two polynomial regression models can be found in Dette et al. (2012), but to our best knowledge no other analytical solutions are available in the literature.

In the present paper, we consider the problem of constructing $T$-optimal discriminating designs for Fourier regression models, which are widely used to describe periodic phenomena (see for example Lestrel, 1997). Optimal designs for estimating all parameters of the Fourier regression model have been discussed by numerous authors (see e.g. Karlin and Studden, 1966, page 347, Lau and Studden, 1985; Kitsos et al., 1988; Riccomagno et al., 1997 and Dette and Melas, 2003 among others). Discriminating design problems in the spirit of Hunter and Reiner (1965) and Stigler (1971) have been discussed by Biedermann et al. (2009) and Zen and Tsai (2004) among others, but $T$-optimal designs for Fourier regression models, have not been investigated in the literature so far. In Section 2, we introduce the problem and provide a characterization of $T$-optimal discriminating designs in terms of a classical approximation problem. Explicit solutions of the $T$-optimal design problem for Fourier regression models are discussed in Section 3. Finally, in Section 4, we provide some numerical results of these challenging optimization problems. In particular, we demonstrate that the structure (more precisely the number of support points) of the $T$-optimal discriminating design depends sensitively on the location of the parameters.

2. $T$-optimal discriminating designs

Consider the classical regression model

$$y = \eta(x) + \varepsilon,$$  \hspace{1cm} \text{(2.1)}

where the explanatory variable $x$ varies in a compact design space, say $\mathcal{X}$, and observations at different locations, say $x$ and $x'$, are assumed to be independent. In (2.1), the quantity $\varepsilon$ denotes a random variable with mean 0 and variance $\sigma^2$, and $\eta(x)$ is a function, which is called regression function in the literature (see Seber and Wild, 1989). We assume that the experimenter has two parametric models, say $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$, for this function in mind to describe the relation between predictor and response, and that the first goal of the experiment is to identify the appropriate model from these two candidates. In order to find "good" designs for discriminating between the models $\eta_1$ and $\eta_2$, we consider approximate designs in the sense of Kiefer (1974), which are probability measures on the design space $\mathcal{X}$ with finite support. The support points, say $x_1, \ldots, x_s$, of an (approximate) design $\xi$ define the locations where observations are taken, while the weights denote the corresponding relative proportions of total observations to be taken at these points. If the design $\xi$ has masses $\omega_i > 0$ at the different points $x_i (i = 1, \ldots, s)$ and $n$ observations can be made, the quantities $\omega_i n$ rounded to integers, say $n_i$, satisfying $\sum_{i=1}^s n_i = n$, and the experimenter takes $n_i$ observations at each location $x_i (i = 1, \ldots, s)$ (see for example Pukelsheim and Rieder, 1992).

For the construction of a good design for discriminating between the models $\eta_1$ and $\eta_2$, Atkinson and Fedorov (1975a) proposed in a seminal paper to fix one model, say $\eta_2$, and to determine the discriminating design such that the minimal deviation between the model $\eta_2$ and the class of models defined by $\eta_1$ is maximized. More precisely, a $T$-optimal design is defined $\xi^*$ by

$$\xi^* = \arg \max_{\xi} \int_{\mathcal{X}} \left( \eta_2(x, \theta_2) - \eta_1(x, \hat{\theta}_1) \right)^2 \xi(dx),$$

where the parameter $\hat{\theta}_1$ minimizes the expression

$$\hat{\theta}_1 = \arg \min_{\theta_1} \int_{\mathcal{X}} (\eta_2(x, \theta_2) - \eta_1(x, \theta_1))^2 \xi(dx).$$

Note that the $T$-optimality criterion is a local optimality criterion in the sense of Chernoff (1953), because it requires knowledge of the parameter $\theta_2$. Bayesian versions of this criterion have recently been investigated by Dette et al. (2013, 2015).

Remark 2.1. Consider the case where the models $\eta_1$ and $\eta_2$ are linear and nested, say

$$\eta_1(x, \theta_1) = \theta_1^T f_1(x); \quad \eta_2(x, \theta_2, b) = \tilde{\theta}_2^T f_1(x) + b^T f_2(x).$$

It was pointed out by Dette and Titoff (2009) that the function $\int_{\mathcal{X}} (\eta_2(x, \theta_2) - \eta_1(x, \tilde{\theta}_1))^2 \xi(dx)$ has the representation

$$T(\xi, b) = \int_{\mathcal{X}} (\eta_2(x, \theta_2) - \eta_1(x, \tilde{\theta}_1))^2 \xi(dx) = b^T \left( M_{22}(\xi) - M_{12}^T(\xi) M_{11}^{-1}(\xi) M_{12}(\xi) \right) b$$

where

$$M_{ij}(\xi) = \int_{\mathcal{X}} f_i(x)f_j^T(x)\xi(dx) \quad (i, j = 1, 2).$$
and $M_1^1(\xi)$ denotes a generalized inverse of the matrix $M_1(\xi)$. Thus, the $T$-optimal design maximizes the function $T(\xi, b)$ (for a fixed vector $b$). On the other hand, discriminating designs in the sense of Stigler (1971) maximize the determinant of the matrix

$$M_{22}(\xi) - M_{12}^T(\xi)M_{11}(\xi)M_{12}(\xi).$$

In the present work, we consider cases, where the competing regression functions are given by two Fourier regression models of different order, that is

$$\eta_1(x, \theta_1) = q_0 + \sum_{i=1}^{k_1} q_{2i-1} \sin(ix) + \sum_{i=1}^{k_2} q_{2i} \cos(ix)$$

and

$$\eta_2(x, \theta_2) = q_0 + \sum_{i=1}^{k_1} q_{2i-1} \sin(ix) + \sum_{i=1}^{k_2} q_{2i} \cos(ix) + \sum_{i=k_1+1}^{m} b_{2(i-k_1)-1} \sin(ix) + \sum_{i=k_2+1}^{m} b_{2(i-k_2)} \cos(ix),$$

where

$$\theta_1 = (\bar{q}_{0}, \bar{q}_{2}, \ldots, \bar{q}_{2k_2}, \bar{q}_{1}, \ldots, \bar{q}_{2k_1-1})$$

$$\theta_2 = (\bar{q}_{0}, \ldots, \bar{q}_{2k_2}, \bar{q}_{1}, \ldots, \bar{q}_{2k_1-1}, b_2, \ldots, b_{2m}, b_1, \ldots, b_{2m-1})$$

are the parameter vectors in model $\eta_1$ and $\eta_2$, respectively. Fourier regression models are widely used to describe periodic phenomena (see e.g. Mardia, 1972, or Lestrel, 1997) and the problem of designing experiments for Fourier regression models has been discussed by several authors (see the cited references in the introduction). However, the problem of constructing $T$-optimal discriminating designs for these models has not been addressed in the literature so far.

We assume that the design space is given by the interval $\chi = [0, 2\pi]$ and denote the difference $\eta_2(x, \theta_2) - \eta_1(x, \theta_1)$ by

$$\eta(x, q, \bar{b}) = q_0 + \sum_{i=1}^{k_1} q_{2i-1} \sin(ix) + \sum_{i=1}^{k_2} q_{2i} \cos(ix) + \sum_{i=k_1+1}^{m} b_{2(i-k_1)-1} \sin(ix) + \sum_{i=k_2+1}^{m} b_{2(i-k_2)} \cos(ix),$$

where $q = (q_0, q_1, \ldots, q_{2k_1-1}, q_2, \ldots, q_{2k_2})^T$, $q_i = \bar{q}_i - \bar{q}_1$, and $\bar{b} = (b_1, b_3, \ldots, b_{2(m-k_1)-1}, b_2, b_4, \ldots, b_{2(m-k_2)})^T$ denotes the vector of “additional” parameters in model (2.3). With these notations, the $T$-optimality criterion reduces to

$$T(\xi, \bar{b}) = \min_q \int_0^{2\pi} \eta^2(x, q, \bar{b})\xi(dx),$$

and a $T$-optimal design for discriminating between the models (2.2) and (2.3) maximizes $T(\xi, \bar{b})$, that is

$$\hat{\xi}^* = \arg \max_\xi T(\xi, \bar{b}).$$

The following result provides a characterization of $T$-optimal designs and is known in the literature as the equivalence theorem for $T$-optimality (see, for instance, Theorem 2.2 in Dette and Titoff, 2009).

**Theorem 2.1.** For a fixed vector $\bar{b}$, the following conditions are equivalent:

1. The design

$$\hat{\xi}^* = \begin{pmatrix} x_1^* & \ldots & x_n^* \end{pmatrix}, \quad x_i^* \in [0, 2\pi], \ i = 1, \ldots, n$$

is a $T$-optimal discriminating design for the models $\eta_1$ and $\eta_2$.

2. There exists a vector $q^*$ and a positive constant $h$ such that the function $\psi^*(x) = \eta(x, q^*, \bar{b})$ satisfies the following conditions

   (i) $|\psi^*(x)| \leq h$, for all $x \in [0, 2\pi]$.

   (ii) $|\psi^*(x_i^*)| = h$, for all $i = 1, 2, \ldots, n$.

   (iii) The support points $x_i^*$ and weights $\omega_i$ of the design $\hat{\xi}^*$ satisfy the conditions

$$\sum_{i=1}^n \psi^*(x_i^*) \cos(jx_i^*)\omega_i = 0, \ j = 0, \ldots, k_2,$$

$$\sum_{i=1}^n \psi^*(x_i^*) \sin(jx_i^*)\omega_i = 0, \ j = 1, \ldots, k_1.$$
determination of $T$-optimal discriminating designs is a very challenging problem. The complexity of the problem depends on the dimension of the vector $\mathbf{b}$. In the following Sections 3 and 4, we provide explicit and numerical solutions of this difficult optimal design problem for Fourier regression models. In particular, we will demonstrate that the structure of the $T$-discriminating design (such as the number of support points) depends on the location of the vector $\mathbf{b}$ in the $(2m - k_1 - k_2)$-dimensional Euclidean space.

3. Explicit solutions

In this section, we give some explicit $T$-optimal discriminating designs for Fourier regression models. In particular, we consider the problem of constructing $T$-optimal discriminating designs for the models (2.2) and (2.3), where

$$k_1 = k_2 = m - 1,$$  

(3.1)

$$k_1 = m - 1, \quad k_2 = m - 2$$  

(3.2)

or

$$k_1 = m - 2, \quad k_2 = m - 1.$$  

(3.3)

In the following discussion, we give an explicit solution for the case (3.1), while for the case (3.2) explicit results are provided for specific values of the parameters $b_i$ in model (2.4). Corresponding results for the case (3.3) are briefly mentioned in Remark 3.1. In general, the solution of the $T$-optimal design problem depends in a complicated way on the parameters $\mathbf{b}$, and we demonstrate numerically in Section 4 that the number of support points of the $T$-optimal discriminating design changes if the vector $\mathbf{b}$ is located in different areas of the space $\mathbb{R}^2$.

We begin with a discussion of the case $k_1 = k_2 = m - 1$ and rewrite the function in (2.4) as

$$\eta(x, \mathbf{q}, \mathbf{b}) = q_0 + \sum_{i=1}^{m-1} q_{2i-1} \sin(ix) + \sum_{i=1}^{m-1} q_{2i} \cos(ix) + b_1 \sin(mx) + b_2 \cos(mx).$$

Our first result gives an explicit solution of the $T$-optimal design problem in the case $b_1^2 + b_2^2 \neq 0$.

**Theorem 3.1.** Consider the Fourier regression models (2.2) and (2.3) with $k_1 = k_2 = m - 1$. Let $b_1^2 + b_2^2 \neq 0$, and define $x_i^* = \frac{i}{m} \alpha + \frac{(i-1)}{m} \pi \mod 2\pi$ $(i = 1, \ldots, 2m)$, where $\alpha$ is the solution of the system

$$\sin(\alpha) = \frac{b_1}{\sqrt{b_1^2 + b_2^2}}, \quad \cos(\alpha) = \frac{b_2}{\sqrt{b_1^2 + b_2^2}}.$$

Then, the design

$$\xi^* = \left( \begin{array}{cccc} x_1^* & x_2^* & \cdots & x_{2m}^* \\ 1 & 1 & \cdots & 1 \\ \frac{1}{2m} & \frac{1}{2m} & \cdots & \frac{1}{2m} \end{array} \right),$$

(3.4)

is a $T$-optimal discriminating design.

**Proof.** We consider the function

$$\psi^*(x) = \eta(x, 0, \mathbf{b}) = b_1 \sin(mx) + b_2 \cos(mx)$$

and prove that this function and the weights $\omega_i^* = \frac{1}{2m}$ and support points $x_i^*$ of the design $\xi^*$ defined in (3.4) satisfy the conditions of Theorem 2.1.

Direct calculations show for the support points of the design $\xi^*$ the identities

$$\psi^*(x_i^*) = (-1)^{i-1} \sqrt{b_1^2 + b_2^2}, \quad i = 1, \ldots, 2m.$$

Consequently, the function $\psi^*$ satisfies conditions (i)-(ii) for $h = \sqrt{b_1^2 + b_2^2}$, and it remains to show that the equations in (2.6) and (2.7) hold. In other words, we have to check that the equalities

$$\sum_{i=1}^{2m} (-1)^{i-1} \sin(jx_i^*) = 0, \quad j = 0, \ldots, m - 1,$$

(3.5)

$$\sum_{i=1}^{2m} (-1)^{i-1} \cos(jx_i^*) = 0, \quad j = 0, \ldots, m - 1,$$

are satisfied. Observing the identities $\sin(\alpha + \beta) = \sin(\alpha) \cos(\beta) + \cos(\alpha) \sin(\beta)$ and $\cos(\alpha + \beta) = \cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta)$, we can rewrite (3.5) as

$$\sum_{i=1}^{2m} (-1)^i \sin\left(\frac{(i-1)\pi}{m}\right) = 0, \quad \sum_{i=1}^{2m} (-1)^i \cos\left(\frac{(i-1)\pi}{m}\right) = 0, \quad j = 0, \ldots, m - 1.$$
These equalities are a consequence of the identity
\[
\sum_{\ell=0}^{2m-1} e^{\frac{2\pi i \ell}{m}} (-1)^j = 0 \quad j = 1, \ldots, m-1
\]
(here \(i = \sqrt{-1}\) and the case \(j = 0\) has to be considered separately), and the assertion of Theorem 3.1 now follows from Theorem 2.1.

**Example 3.1.** Suppose that \(m = 3, b_1 = b_2 = 1\) and \(k_1 = k_2 = 2\), then it follows from Theorem 3.1 that the design with equal masses at the six points \(\frac{1}{12} \pi, \frac{5}{12} \pi, \frac{3}{4} \pi, \frac{7}{12} \pi, \frac{17}{12} \pi\) and \(\frac{2}{3} \pi\) is a \(T\)-optimal discriminating design for the two Fourier regression models
\[
\begin{align*}
\tilde{q}_0 + \tilde{q}_1 \sin x + \tilde{q}_2 \cos x + \tilde{q}_3 \sin(2x) + \tilde{q}_4 \cos(2x) \\
\tilde{q}_0 + \tilde{q}_1 \sin x + \tilde{q}_2 \cos x + \tilde{q}_3 \sin(2x) + \tilde{q}_4 \cos(2x) + b_1 \sin(3x) + b_2 \cos(3x).
\end{align*}
\]

The function \(\psi^*\) of Theorem 2.1 for this design is depicted in Fig. 1.

Next we consider the case \(k_1 = m - 1, k_2 = m - 2\) and determine \(T\)-optimal discriminating designs for the trigonometric regression models
\[
\begin{align*}
\eta_1(x, \theta_1) &= \tilde{q}_0 + \sum_{i=1}^{m-1} \tilde{q}_{2i-1} \sin(ix) + \sum_{i=1}^{m-2} \tilde{q}_{2i} \cos(ix) \\
\eta_2(x, \theta_2) &= \tilde{q}_0 + \sum_{i=1}^{m-1} \tilde{q}_{2i-1} \sin(ix) + \sum_{i=1}^{m-2} \tilde{q}_{2i} \cos(ix) + b_0 \cos((m-1)x) + b_1 \sin(mx) + b_2 \cos(mx).
\end{align*}
\]

Note that the two regression models in (2.2) and (2.3) now differ by three terms when \(k_1 = m - 1, k_2 = m - 2\). In general, \(T\)-optimal discriminating designs for this case have to be determined numerically, and we will provide numerical results for \(m = 2\) and \(m = 3\) in Section 4. However, for some special configurations of the parameters, the \(T\)-optimal discriminating designs can also be found explicitly, and these cases will be discussed in the present section.

We may assume without loss of generality that \(b_0 = 1\). Indeed, if \(b_0 = 0\), the optimal designs can be obtained from Theorem 3.1. Moreover, if \(b_0 \neq 0\), the \(T\)-optimal discriminating design does not depend on the particular value of \(b_0\), since we can divide all coefficients by this parameter. After normalizing we therefore obtain
\[
\begin{align*}
\eta(x, \mathbf{q}, \mathbf{b}) &= q_0 + \sum_{i=1}^{m-1} q_{2i-1} \sin(ix) + \sum_{i=1}^{m-2} q_{2i} \cos(mx) + \cos((m-1)x) + b_1 \sin(mx) + b_2 \cos(mx).
\end{align*}
\]

We now concentrate on two special cases: \(b_1 = 0, b_2 
eq 0\) and \(b_2 = 0, b_1 
eq 0\), for which we can provide an explicit solution of the \(T\)-optimal design problem if the absolute value of the non-vanishing parameter is sufficiently large. For this purpose, we define support points and weights by
\[
\begin{align*}
\chi_i^*(b) &= \arccos(-\left(1 + \frac{1}{2mb}\right) \cos\left(\frac{(m-i+1)\pi}{m} - \frac{1}{2mb}\right)), \\
\omega_i^* &= \frac{1}{m} \cos^2\left(\frac{(i-1)\pi}{2m}\right), \quad i = 1, \ldots, m.
\end{align*}
\]

Our next result gives an explicit solution of the \(T\)-optimal design problem in the case \(b_1 = 0, b_2 \neq 0\).
Theorem 3.2. Consider the Fourier regression models (3.8) and (3.9) with $b_0 = 1$, $b_1 = 0$, $b_2 \neq 0$.

(a) If $b_2 \geq \frac{1}{2m} \cot^2 \left( \frac{\pi}{2m} \right) > 0$, then the design

$$
\xi^*_1 = \left( \frac{x_1^*(b_2)}{\omega_1^*} \cdots \frac{x_m^*(b_2)}{\omega_m^*} \ 2\pi - \frac{x_1^*(b_2)}{\omega_1^*} \cdots \frac{x_m^*(b_2)}{\omega_m^*} \right)
$$

is a $T$-optimal discriminating design, where the support points and weights are defined in (3.11) and (3.12), respectively.

(b) If $b_2 \leq -\frac{1}{2m} \cot^2 \left( \frac{\pi}{2m} \right) < 0$, then the design

$$
\xi^*_2 = \left( \frac{\pi - x_m^*(b_2)}{\omega_m^*} \cdots \frac{\pi - x_1^*(b_2)}{\omega_1^*} \ \frac{\pi + x_m^*(b_2)}{\omega_m^*} \cdots \frac{\pi + x_1^*(b_2)}{\omega_1^*} \right),
$$

is a $T$-optimal discriminating design, where the support points and weights are defined in (3.11) and (3.12), respectively.

Proof. At first we show that $\xi^*_1$ and $\xi^*_2$ are probability measures, i.e. that the weights in (3.13) and (3.14) sum up to one. Indeed it follows from the identity

$$
\frac{1}{m} + 2 \sum_{i=1}^{m-1} \cos^2 \left( \frac{i\pi}{2m} \right) = \begin{cases} 
\frac{1}{m} + 2 \sum_{i=1}^{m-1} \left( \cos^2 \left( \frac{i\pi}{2m} \right) + \sin^2 \left( \frac{i\pi}{2m} \right) \right) + \frac{1}{m} & \text{if } m \text{ is even} \\
\frac{1}{m} + 2 \sum_{i=1}^{m-1} \left( \cos^2 \left( \frac{i\pi}{2m} \right) + \sin^2 \left( \frac{i\pi}{2m} \right) \right) & \text{if } m \text{ is odd}
\end{cases}
$$

that

$$
\omega_1^* + 2 \sum_{i=2}^{m} \omega_i^* = \frac{1}{m} + 2 \sum_{i=1}^{m-1} \cos^2 \left( \frac{i\pi}{2m} \right) = 1.
$$

Next we prove that $\xi^*_1$ and $\xi^*_2$ are the in fact $T$-optimal designs. We only consider the case $b_2 \geq \frac{1}{2m} \cot^2 \left( \frac{\pi}{2m} \right) > 0$ and note that the other case follows by similar arguments. We will use Theorem 2.1 and prove the existence of a vector $q^*$, such that the function $\psi^*(x) = \eta(x, q^*, B)$ satisfies conditions (i)–(iii) in this theorem. For this purpose, let $T_m(x) = \cos(m \arccos x)$ denote the $m$th Chebychev polynomial of the first kind, then it is follows from the definition of the function $\eta(x, q, B)$ (see (3.10)) that there exists a vector $q^*$ such that the function

$$
\psi^*(x) = \eta(x, q^*, B) = (-1)^m |b_2| \left( 1 + \frac{1}{2m|b_2|} \right)^m \frac{\cos(x) - \frac{2m|b_2|}{1 + \frac{1}{2m|b_2|}}}{1 + \frac{1}{2m|b_2|}}
$$

$$
= (-1)^m |b_2| \left( 1 + \frac{1}{2m|b_2|} \right)^m 2^{m-1} \left( \frac{1 - \frac{1}{1 + \frac{1}{2m|b_2|}}}{1 + \frac{1}{2m|b_2|}} \right)^m \cos(x)^m
$$

$$
- \left( \frac{1 - \frac{1}{1 + \frac{1}{2m|b_2|}}}{1 + \frac{1}{2m|b_2|}} \right)^{m-1} \left( \frac{m}{1 + 2m|b_2|} \right) \cos(x)^{m-1} + P_{m-2}(\cos(x))
$$

$$
= \left( 2^{m-1} |b_2| \left( \frac{\cos(x)}{m} + \frac{2^{m-2} \cos(x)^{m-1}}{m} \right) + P_{m-2}(\cos(x)),
$$

(3.15)

where $P_{m-2}$ is a corresponding polynomial of degree $m - 2$. Consequently $\psi^*(x)$ is a trigonometric polynomial of degree $m$ with leading term $|b_2| \cos(mx) + \cos((m - 1)x)$ (note that the leading term of $T_m(x)$ is given by $2^{m-1}x^m$ and that $2^{m-1} \cos(x)^m = \cos(mx) + m \cos((m - 2)x) + \cdots$). Direct calculations show that the points $x_i^*(b_2)$ defined in (3.11) are the extremal points of this function, that is

$$
\psi^*(x_i^*(b_2)) = (-1)^{i-1} |b_2| \left( 1 + \frac{1}{2m|b_2|} \right)^m, \quad i = 1, \ldots, m.
$$

(3.16)

Consequently, $\psi^*$ satisfies conditions (i) and (ii) of Theorem 2.1. Finally, we prove the conditions (2.6) and (2.7). The corresponding equalities reduce to

$$
\sum_{i=1}^{m} \omega_i^* \psi^*(x_i^*(b_2)) \cos(jx_i^*(b_2)) + \sum_{i=1}^{m} \omega_i^* \psi^*(2\pi - x_i^*(b_2)) \cos(j(2\pi - x_i^*(b_2))) = 0
$$

(3.17)

($j = 0, \ldots, m - 2$), and

$$
\sum_{i=1}^{m} \omega_i^* \psi^*(x_i^*(b_2)) \sin(jx_i^*(b_2)) + \sum_{i=1}^{m} \omega_i^* \psi^*(2\pi - x_i^*(b_2)) \sin(j(2\pi - x_i^*(b_2))) = 0
$$

(3.18)
(j = 1, . . . , m - 1). By (3.11) and (3.15), we have \( \chi_i^*(b_2) = 0 \) and \( \psi^*(x) = \psi^*(2\pi - x) \). Using (3.16), we can rewrite the left hand side of (3.18) as

\[
\sum_{i=2}^{m} \omega_i^* (-1)^{j-1} (\sin(j\chi_i^*(b_1)) + \sin(j(2\pi - \chi_i^*(b_2)))) = 0
\]

where we have used the identity \( \sin(j(2\pi - x)) = -\sin(jx) \). Consequently, (3.18) is satisfied. Similarly, (3.17) is equivalent to

\[
\sum_{i=1}^{m} \overline{\omega}_i (-1)^j \cos(j\chi_i^*(b_1)) = 0, \quad j = 0, \ldots, m - 2,
\]

where we use the notations \( \overline{\omega}_1 = \omega_1^*, \overline{\omega}_i = \omega_i^*, \quad i = 0, \ldots, m - 2 \) in (3.19). Defining \( t_i = \cos(\chi_i^*(b_2)) \) we obtain for the left hand side of (3.19)

\[
\sum_{i=1}^{m} \overline{\omega}_i (-1)^j \cos(j\chi_i^*(b_2)) = \sum_{i=1}^{m} \sum_{p=0}^{m-2} \overline{\omega}_i (-1)^j a_p t_i^p = \sum_{p=0}^{m-2} a_p \sum_{i=1}^{m} \overline{\omega}_i (-1)^j t_i^p
\]

for some coefficients \( a_p \). It is proved in Appendix A.1 of Dette et al. (2012) that

\[
\sum_{i=1}^{m} \overline{\omega}_i (-1)^j t_i^p = 0, \quad p = 0, \ldots, m - 2
\]

which implies (3.17) (note that in the notation of the paper of Dette et al. (2012) we have \( m = n \), \( t_i = t_{n-i+1}^*(\frac{1}{2\pi}) \), \( \overline{\omega}_i = \frac{1}{2} \omega_{n-i+1}^* \)). The T-optimality of the design \( \xi_1^* \) now directly follows from Theorem 2.1. \( \square \)

The next theorem considers the case \( b_1 \neq 0 \), \( b_2 = 0 \), which is substantially harder. Here, we are able to determine the \( T \)-optimal discriminating designs explicitly if the degree \( m \) of the Fourier regression model is odd.

**Theorem 3.3.** Consider the Fourier regression models (3.8) and (3.9) with \( b_0 = 1 \), \( b_1 \neq 0 \), \( b_2 = 0 \), where \( m \) is odd. For \( \ell = 1, 2 \) let \( t_1^{(\ell)} \) and \( \omega_i^{(\ell)} \), denote the support points and weights of the designs \( \xi_1^* \) and \( \xi_2^* \) defined in (3.13) and (3.14) and define

\[
t_i^{(\ell)} = t_i^{(\ell)} + \frac{\pi}{2} \mod 2\pi; \quad \ell = 1, 2.
\]

(a) If \( b_1 \geq \frac{1}{2m} \cot^2 \left( \frac{\pi}{2m} \right) \), then the design

\[
\tilde{\xi}_1^* = \begin{pmatrix}
t_1^{(1)} & \cdots & t_{2m-1}^{(1)} \\
\omega_1^{(1)} & \cdots & \omega_{2m-1}^{(1)}
\end{pmatrix}
\]

is a \( T \)-optimal discriminating design.

(b) If \( b_1 \leq -\frac{1}{2m} \cot^2 \left( \frac{\pi}{2m} \right) < 0 \), then the design

\[
\tilde{\xi}_2^* = \begin{pmatrix}
t_1^{(2)} & \cdots & t_{2m-1}^{(2)} \\
\omega_1^{(2)} & \cdots & \omega_{2m-1}^{(2)}
\end{pmatrix}
\]

is a \( T \)-optimal discriminating design.

**Proof.** Assume that \( m = 2d - 1 \) for some \( d = 2, 3, \ldots \). The proof is similar to the proof of Theorem 3.2, where we use the function

\[
\psi^*(x) = \overline{\eta}(x, q^*, \overline{b}) = (-1)^{\frac{d+1}{2}} |b_1| \left( 1 + \frac{1}{2m|b_1|} \right)^m \mu \left( \frac{-\sin(x) - \frac{1}{2m|b_1|}}{1 + \frac{1}{2m|b_1|}} \right).
\]

in Theorem 2.1. This function is of the form

\[
q_0 + \sum_{i=1}^{2d-2} q_{2i-1} \sin(ix) + \sum_{i=1}^{2d-3} q_{2i} \cos(ix) + \cos((2d - 2)x) + b_1 \sin((2d - 1)x)
\]

and satisfies the assumptions of Theorem 2.1 follows from the identity

\[
\cos \left( (2d - 1) \arccos(t) \right) \equiv (-1)^{d-1} \sin \left( (2d - 1) \arcsin(t) \right), \quad t \in [-1, 1], \quad d = 2, 3, \ldots
\]

The details are omitted for the sake of brevity. \( \square \)
Example 3.2. Consider the case $m = 5$, $b_0 = 1$, $b_1 = 0$, $b_2 = 2$ and $k_1 = 4$, $k_2 = 3$. The $T$-optimal discriminating design can be obtained from Theorem 3.2 and is given by

$$
\xi^*_1 = \begin{pmatrix}
0 & 0.65 & 1.29 & 1.95 & 2.69 & 3.59 & 4.33 & 4.99 & 5.64 \\
0.20 & 0.18 & 0.13 & 0.07 & 0.02 & 0.07 & 0.13 & 0.18
\end{pmatrix}.
$$

Similarly, if $b_1 = 2$, $b_2 = 0$ the $T$-optimal discriminating design can be obtained from Theorem 3.3 and is given by

$$
\xi^*_2 = \begin{pmatrix}
1.57 & 2.21 & 2.86 & 3.52 & 4.26 & 5.16 & 5.9 & 0.28 & 0.93 \\
0.20 & 0.18 & 0.13 & 0.07 & 0.02 & 0.07 & 0.13 & 0.18
\end{pmatrix}.
$$

Note that the design $\xi^*_1$ is obtained from the design $\xi^*_1$ by the transformation $x \rightarrow x + \frac{\pi}{2}$. In Fig. 2, we display the function $\psi^*$ in the equivalence Theorem 2.1 for both cases.

Remark 3.1. In the case $k_1 = m - 2$, $k_2 = m - 1$ explicit solutions can be obtained by similar arguments as given in the proof of Theorems 3.2 and 3.3. If $m = 2d$ is even and $b_1 = 0$, then the function $\overline{\eta}$ is given by

$$
\overline{\eta}(x, q, \tilde{b}) = q_0 + \sum_{i=1}^{2d-2} q_{2i-1} \sin(\pi i x) + \sum_{i=1}^{2d-1} q_{2i} \cos(\pi i x) + \sin((2d-1)x) + b_2 \cos(2dx).
$$

If $b_2 \geq \frac{1}{2m} \cot^2(\frac{\pi}{2m})$, the $T$-optimal discriminating design for the Fourier regression models (2.2) and (2.3) is given by the design (3.20), where the support points and weights are defined by

$$
t_i^{(1)} = t_i^{(2)} + \frac{3\pi}{2} \mod 2\pi; \quad i = 1, 2, \ldots, 2m-1,
$$

and

$$
\omega_i^{(1)} = \omega_i^{(2)}; \quad i = 1, 2, \ldots, 2m-1,
$$

respectively, and $t_i^{(1/2)}$ and $\omega_i^{(1/2)}$ are the support points of the design $\xi^*_1$ in (3.13). The extremal polynomial $\psi^*$ in Theorem 2.1 is given by

$$
\psi^*(x) = \overline{\eta}(x, q^*, \tilde{b}) = (-1)^{\frac{m}{2}} |b_1| \left(1 + \frac{1}{2m|b_1|}\right)^m T_m\left(\frac{-\sin(x) + \frac{1}{2m|b_1|}}{1 + \frac{1}{2m|b_1|}}\right),
$$

where the fact that $\psi^*$ can be represented in the form (3.23) follows from (3.22). A similar result is available in the case $b_2 \leq -\frac{1}{2m} \cot^2(\frac{\pi}{2m})$ and the details are omitted for the sake of brevity.

4. Some numerical results

The results in the case $k_1 = m - 1$, $k_2 = m - 2$ are only correct if the module of $b_1$ or $b_2$ is larger or equal to some threshold. Otherwise $T$-optimal designs have a more complicated structure and have to be found numerically (see Dette et al., in press for some algorithms). In this section, we provide some more insight in the structure of $T$-optimal discriminating designs in cases, where an analytical determination of the optimal design is not possible. For this purpose, we consider the Fourier regression models (3.8) and (3.9), where $b_0 = 1$, $b_1 \neq 0$ and $b_2 \neq 0$. Recalling the representation (3.10) for the function $\overline{\eta}$ in (2.4), we see that the support points and weights of the optimal $T$-discriminating designs depend on the two parameters $b_1$, $b_2$ of the extended model. Moreover, the structure of the optimal design changes and depends on the location of the point $(b_1, b_2)$. We have calculated $T$-optimal discriminating designs for the Fourier regression models (3.8) and (3.9) for $m = 2$ and $m = 3$. It follows from Theorem 2.1 that the optimal design with support points $x_1, \ldots, x_n$ and weights $\omega_1, \ldots, \omega_n$ has
to satisfy the following system of nonlinear equations:

\[
\begin{align*}
|\psi^*(x_i^*)| &= |\psi^*(x_{i+1}^*)|, & i &= 1, \ldots, n - 1, \\
\psi^*(x_i^*)' &= 0, & i &= 1, \ldots, n, \\
\sum_{i=1}^{n} \psi^*(x_i^*) \cos(jx_i^*) \omega_j &= 0, & j &= 0, \ldots, m - 2, \\
\sum_{i=1}^{n} \psi^*(x_i^*) \sin(jx_i^*) \omega_j &= 0, & j &= 1, \ldots, m - 1, \\
\sum_{i=1}^{n} \omega_i &= 1,
\end{align*}
\]  

(4.1)

which can be solved with any optimization software (the corresponding calculations in this paper were done in Maple).

Below we consider a few examples which show how the number of support points of the \( T \)-optimal discriminating design depends on the parameters \( b_1 \), \( b_2 \). We fixed one of the parameters, say \( b_1 \), and found the \( T \)-optimal discriminating designs for different values of \( b_2 \). For each combination we solved the system (4.1) for different values of \( n \), where we start with \( n = 1 \) (for \( m = 2 \)) or \( n = 5 \) (for \( m = 3 \)). The optimality of the design calculated as the solution of the system of equations in (4.1) has been checked by an application of the equivalence Theorem 2.1. If the optimality of the design is confirmed, the procedure stops. Otherwise \( n \) is decreased (or increased) and the system (4.1) is solved again. Our numerical results, including a direct solution of the original maximin problem by means of a different optimization method (see package DirectSearch in Maple), indicate that in the examples under consideration the \( T \)-optimal discriminating design is unique. However, in general the problem of uniqueness is still open (for some examples with non-unique \( T \)-optimal discriminating designs see Dette and Titoff, 2009).

If \( m = 2 \), the \( T \)-optimal designs have either 2 or 3 support points, and the corresponding areas for the point \((b_1, b_2)\) are depicted in Fig. 3(a). For example, if \( b_1 = 0 \) and \( b_2 \geq 0.25 \) the \( T \)-optimal discriminating design has 3 support points (which coincides with the results of Theorem 3.2), while in the opposite case the optimal design is supported at only two points. This pattern does not change if \( b_1 \neq 0 \), but the threshold is slightly increasing. Numerical calculations show that the threshold converges to \( \frac{\sqrt{2}}{4} \) as \( b_1 \to \infty \).

Fig. 3(b) shows corresponding results for the case \( m = 3 \), and we see that the plane is separated into five parts. Four of them correspond to parameter configurations, where the \( T \)-optimal discriminating design is supported at 5 points. Additionally, there exists one region, where a 4-point design is \( T \)-optimal for discriminating between the two Fourier regression models. The boundaries of the corresponding areas can easily be found by adding the condition \( \omega_n(b_3) = 0 \) (for some \( k \in \{1, \ldots, n\} \)) to the system (4.1). Here \( n \) is equal to 3 (for \( m = 2 \)) or equal to 5 (for \( m = 3 \)) and \( k \) depends on \( b_1 \) and \( b_2 \). We constructed these boundaries in Fig. 3 point-by-point for a sequence of fixed values of \( b_3 \), from the corresponding interval. Consider for example the situation, where \( b_2 = 1 \) and \( b_1 \) varies in the interval \([0, 3]\). In this case, there exist two values, say \( b_1^{\text{min}} \) and \( b_1^{\text{max}} \), where the line through the point \((0, 1)\) in the direction \((1, 0)\) intersects the boundary of the fourth region (see Fig. 3(b)). If \( b_1 \in [b_1^{\text{min}}, b_1^{\text{max}}] \) the \( T \)-optimal discriminating design has 5 support points, while it has only 4 support points if \( b_1 \in [b_1^{\text{min}}, b_1^{\text{max}}] \). Finally, on the interval \([b_1^{\text{max}}, 3]\) the \( T \)-optimal discriminating design has again 5 support points. The support points and corresponding weights of the \( T \)-optimal discriminating design are shown in Fig. 4 (for the Fourier regression models (3.8) and (3.9)) as a function of the parameter \( b_1 \in [0, 3] \) where \( b_2 = 1 \). For the sake of brevity, we only consider the optimal design for nonnegative \( b_1 \), \( b_2 \). In the other cases, the structure of the optimal designs is very similar (but not necessarily symmetric). For example, in the case \( b_1 \geq 0, b_2 \leq 0 \), the optimal design has support points \( \pi - x_i(b_1) \).

We conclude this section investigating the performance of some commonly used designs with respect to the criterion of model discrimination. To be precise, recall the definition of the criterion \( T(\xi, b) \) in (2.5) and define the \( T \)-efficiency of a
Fig. 4. The support points and weights of the $T$-optimal discriminating design for the Fourier regression models (3.8) and (3.9), where $m = 3$, $b_0 = 1$, $b_2 = 1$, and $b_1 \in [0, 3]$.

Fig. 5. The $T$-efficiency of the $D$-optimal design (a) and $D_3$-optimal design (b) for discriminating between the Fourier regression models (3.8) and (3.9), where $m = 3$, $b_2 = 0, 0.5, 1, 2, 3, 5$, $b_1 \in [0, 5]$, $b_0 = 1$. The horizontal dashed line shows the upper-bound for efficiency.

given design, say $\xi$, by

$$\text{Eff}_T(\xi, b) = \frac{T(\xi, b)}{\max_\eta T(\eta, b)},$$

A frequently used design in applications is the uniform design

$$\xi^*_D = \left( \begin{array}{cccccccc} 0 & \frac{\pi}{4} & \frac{\pi}{2} & \frac{3\pi}{4} & \pi & \frac{5\pi}{4} & \frac{3\pi}{2} & \frac{7\pi}{4} \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} \end{array} \right)$$

for the extended model (2.3) (see Weber and Liebig, 1981; Aspnes and Arwin, 1983 among many others). It is well known that this design is in fact the $D$-optimal design (see Pukelsheim, 2006, Section 9.16), which justifies our notation $\xi^*_D$.

If model discrimination is a preliminary objective of the experiment one could use a discriminating design in the sense of Stigler (1971) as an alternative. This design is defined by

$$\xi^*_D = \arg \max_\xi \det(M_{22}(\xi)) = \arg \max_\xi \det(M_{22}(\xi) - M_{12}(\xi)M_{11}(\xi)M_{12}(\xi)),$$

and provides a most accurate estimation of the three highest coefficients $b_0$, $b_1$ and $b_2$ in model (3.9). The determinant of the matrix $M_{22}$ for a nonsingular matrix $M_{11}(\xi)$ can be represented as a function of the support and weights, which is then maximized numerically. The optimality of this design can be easily checked by the generalized equivalence theorem (see, e.g. Karlin and Studden, 1966, Theorem 8.3). For our case, this design is given by

$$\xi^*_D = \left( \begin{array}{cccccccc} 0 & \frac{\pi}{4} & \frac{\pi}{2} & \frac{3\pi}{4} & \pi & \frac{5\pi}{4} & \frac{3\pi}{2} & \frac{7\pi}{4} \\ 3 & 1 & 3 & 1 & 3 & 1 & 3 & 1 \\ \frac{3}{20} & \frac{10}{20} & \frac{20}{20} & \frac{10}{20} & \frac{20}{20} & \frac{20}{20} & \frac{20}{20} & \frac{20}{20} \end{array} \right)$$

and will be called $D_3$-optimal design throughout this section. The corresponding efficiencies of the $D$- and $D_3$-optimal design are shown in Fig. 5 for various values of $b_2$, where the parameter $b_1$ varies in the interval $[0, 5]$.

Both designs have rather similar $T$-efficiencies which are always smaller than 60%. This similarity can be explained by the fact that the $D$- and $D_3$-optimal design have the same support and only differ with respect to their weights. In these
examples the efficiencies are decreasing with respect to the parameter $b_2$ for most values of the parameter $b_1$. For larger values of $b_2$, the efficiencies of the D- and $D_2$-optimal design are very low. For fixed $b_2$ and larger values of $b_1$, the efficiencies do not change substantially.

Acknowledgments

The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. The work of V.B. Melas and P. Shpilev was supported by St. Petersburg State University (project “Actual problems of design and analysis for regression models”, 6.38.435.2015). This work has also been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG) and by a grant from the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors also thank two anonymous referees and the associate editor for very constructive comments on an earlier version of this paper.

References


Atkinson, A.C., Fedorov, V.V., 1975a. The designs of experiments for discriminating between two rival models. Biometrika 62, 57–70.


Developments of the total entropy utility function for the dual purpose of model discrimination and parameter estimation in Bayesian design

J.M. McGree *

School of Mathematical Sciences, Science and Engineering Faculty, Queensland University of Technology, GPO Box 2434, Brisbane, QLD, 4001, Australia

A R T I C L E   I N F O

Article history:
Received 1 December 2015
Received in revised form 24 March 2016
Accepted 4 May 2016
Available online 16 June 2016

Keywords:
Generalized linear models
Generalized nonlinear models
Optimal design
Particle filter
Sequential design
Sequential Monte Carlo

A B S T R A C T

The total entropy utility function is considered for the dual purpose of model discrimination and parameter estimation in Bayesian design. A sequential design setting is considered where it is shown how to efficiently estimate the total entropy utility function in discrete data settings. Utility estimation relies on forming particle approximations to a number of intractable integrals which is afforded by the use of the sequential Monte Carlo algorithm for Bayesian inference. A number of motivating examples are considered for demonstrating the performance of total entropy in comparison to utilities for model discrimination and parameter estimation. The results suggest that the total entropy utility selects designs which are efficient under both experimental goals with little compromise in achieving either goal. As such, for the type of problems considered in this paper, the total entropy utility is advocated as a general utility for Bayesian design in the presence of model uncertainty.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The importance of deriving experimental designs for the dual purpose of model discrimination and parameter estimation has been stressed by Hill (1978), and has been a significant and challenging problem throughout the history of model-based experimental design, see for example Hill et al. (1968), Borth (1975) and Atkinson (2008). The difficulty in deriving such designs lies in the apparent competing nature of the two experimental goals. That is, a design which is efficient for model discrimination generally has poor estimation properties, and similarly, a design which can efficiently estimate parameters generally yields minimal information about which model is true. Moreover, many of the approaches proposed to derive such designs are highly computational which has restricted the general use in a wide variety of design problems.

In this work, the total entropy utility is considered for the sequential design of experiments. Grounded in information theory, this utility addresses the dual experimental goals of model discrimination and parameter estimation through the exploitation of the additivity property of entropy for the selection of designs which are expected to minimize the sum of the entropy about the model indicator and the parameter. Unfortunately, since being proposed by Borth (1975), the computational challenges involved in evaluating total entropy have restricted the general use of this utility in Bayesian design (Ng and Chick, 2004). Through the availability of advanced Monte Carlo methods and modern computing capabilities, we show how this important utility function can be used efficiently in a wide variety of Bayesian design problems. Most
notably, our approach allows Bayesian designs to be efficiently constructed to address the dual experimental aims of model discrimination and estimation. This is an important contribution in Bayesian design as it allows more informative experimentation to be conducted when using this utility. Further, the empirical results presented suggest that these dual experimental goals are achieved for the examples considered in this paper with little to no compromise under optimal strategies to address either aim (see later).

In this paper, we consider methods for the construction of Bayesian designs, and a recent review of approaches for this purpose has been given by Ryan et al. (2016). The review discusses a wide variety of algorithms and utility functions which have been adopted in Bayesian design with a strong focus on approaches for estimation, with few approaches dealing with model discrimination. Of note was the absence of any approach since Borth (1975) which addresses the problem of deriving designs for the dual problem of model discrimination and estimation. This highlights the difficulty of addressing this design problem and the challenges involved in implementing a previously proposed methodology.

Despite the focus on Bayesian design, approaches from the (classical) design literature are of interest when designing for multiple experimental aims. Constrained and compound optimal design approaches are popular for deriving designs for multiple objectives, and have a long history (Hill et al., 1968; Lauter, 1974, 1976; Lee, 1987, 1988; Dette, 1990; Clyde and Chaloner, 1996; Huang and Wong, 1998; Wong, 1999; Dette and Franke, 2000; Tsai and Zen, 2004; Atkinson, 2008; McGree et al., 2008; Tommasi, 2009; Waterhouse et al., 2009; May and Tommasi, 2014). In the majority of cases, a weighted linear combination of the utilities is considered for selecting optimal designs. This requires the experimenter to choose a weighting parameter representing the relative importance of each experimental goal. Unfortunately, such a choice can be difficult in practice (Clyde and Chaloner, 1996). Other approaches consider finding designs which yield ‘enough’ information about a particular goal (for example, a posterior model probability greater than 0.9), and focus the remaining effort on the other experimental aim. As will be seen, this trade-off or compromise between two potentially competing utilities is naturally accommodated within total entropy for model discrimination and estimation through the additivity property of entropy. Hence, we avoid many of the difficulties in implementing alternative approaches (such as, pre-specifying a weighting parameter).

Another advantage of considering total entropy for finding designs is the avoidance of the need to pre-specify a true model for model discrimination (Atkinson and Fedorov, 1975a,b; López-Fidalgo et al., 2009; Dette and Titoff, 2009). In practice, it is generally quite difficult to pre-specify the true model as determining this is usually the reason for conducting the experiment. In using total entropy, one does not need to specify a true model, but rather one specifies a set of models with prior model probabilities. Such a framework is not dissimilar from that of Ponce de Leon and Atkinson (1991). However, implementing their approach in general (including in the examples considered in this paper) would be highly computational. Hence, using the methods proposed in this paper removes restrictions in implementing many of the proposed approaches for finding designs for model discrimination (and estimation).

Within sequential Bayesian design, sequential Monte Carlo (SMC) (Chopin, 2002; Del Moral et al., 2006) has been shown to be computationally efficient for inference and estimating Bayesian utility functions, particularly when compared to, for example, Markov chain Monte Carlo (MCMC) (Drovandi et al., 2013, 2014). As such, the SMC algorithm which handles model uncertainty in Bayesian design is adopted for the work presented here. In brief, the SMC algorithm is a particle filtering approach for approximating a sequence of target distributions. Here, this sequence of target distributions is a sequence of posterior distributions constructed through data annealing. A particle approximation for this sequence of posterior distributions is formed for each competing model and are jointly considered in finding the optimal design for the observation of the next data point. A major advantage of implementing SMC is the availability of an estimate of the evidence for a given model with minimal additional computational cost. As will be seen, this estimate of the evidence is used in efficiently estimating the total entropy utility throughout the motivating examples presented in this paper. SMC also allows particle approximations of intractable integrals to be efficiently formed, and this is also exploited in estimating total entropy.

Bayesian sequential design has seen some attention in recent times. Weir et al. (2007) and McGree et al. (2012) considered the sequential design of experiments in drug development. An MCMC algorithm was used for posterior inference, and importance sampling was adopted in design selection for computational efficiency. In terms of implementing the SMC algorithm, Drovandi et al. (2013, 2014) consider the sequential design for parameter estimation and then model discrimination in the two cited papers, respectively. The inference algorithm in the latter publication is adopted in the work presented here. Finally, McGree et al. (2015) also consider sequential design for parameter estimation. They developed an SMC algorithm for random effect models, and applied their methodology to design experiments in pharmacology.

In terms of non-sequential Bayesian design, Müller (1999) proposed exploring the joint distribution of the data, model parameter and design in locating optimal designs. This approach has been used in a variety of Bayesian design settings including experiments in epidemiology where the likelihood is intractable (Drovandi and Pettitt, 2013) and in pharmacokinetic studies (Stroud et al., 2001; Ryan et al., 2015). Amzal et al. (2006) extended upon this design approach through a particle method for the efficient exploration of the joint distribution, and concentrated the particles near the mode with simulated annealing (Corana et al., 1987). This method was demonstrated on a problem of optimal resource allocation and also in a clinical trial setting for determining the optimal treatment of patients.

This paper proceeds as follows. Initially, the inference framework within which total entropy will be developed and demonstrated will be described. In Section 3, the total entropy utility function is derived along with a utility for model discrimination and estimation. It is also shown how to estimate these utilities within the defined inference framework. In Section 4, three motivating examples are considered to demonstrate the performance of the total entropy utility in
comparison with the model discrimination utility, the estimation utility and a random design. The paper concludes with a discussion of the results and areas for future research in Section 5.

2. Inference framework

In this section, the SMC inference framework adopted in this paper is explained. The idealized SMC algorithm of Chopin (2002) and Del Moral et al. (2006) for static parameter models will be implemented for sequentially approximating posterior distributions of the parameter. The sequence of posterior distributions will be built via data annealing, and can be described as follows for a given model $m$:

$$p(\theta_m|M = m, y_t, d_t) = \frac{p(y_t|M = m, \theta_m, d_t)p(\theta_m|M = m)}{p(y_t|M = m, d_t)}, \quad \text{for } t = 1, \ldots, T,$$

where $y_t$ denotes all data observed at designs $d_t$ up until time $t$, $M \in \{1, \ldots, K\}$ describes the random variable relating to the model indicator, $p(\theta_m|M = m)$ represents the prior information about the model parameter $\theta_m$ for a given model $m$, and $p(y_t|M = m, d_t)$ denotes the model evidence, and can be used to find posterior model probabilities ($p(M = m|y_t, d_t)$) and Bayes factors. The likelihood is represented as $p(y_t|M = m, \theta_m, d_t)$, and, as independent data will only be considered throughout this paper, can be constructed for $y_t$ by taking the product of individual likelihood values for each of the $y_t$ data points. For notational convenience, $M = m$ will be abbreviated to $m$ for the remainder of the article.

SMC approximates the above sequence of posterior distributions via a particle filter using a combination of reweighting, resampling and move steps. A particle filter can approximate target $\text{distribution}$ for a given model $m$ via a particle set which consists of particles $\theta_{m,t}$ and particle weights $W_{m,t}$, and is denoted as $\{\theta_{m,t}, W_{m,t}\}_{t=1}^{N}$. Upon the observation of a new data point, the particle set needs to be updated to incorporate the information from the new data point. In SMC, this is achieved via updating the particle weights in the reweight step, and is just importance sampling with the current particle set representing the importance distribution. For a given model $m$, this can be described as follows:

$$w_{m,t+1}^i = w_{m,t}^i \frac{p(y_{t+1}|m, \theta_{m,t}^i, d_{t+1})}{\sum_{i=1}^{N} w_{m,t}^i},$$

where $w_{m,t+1}^i$ are the unnormalized particle weights for $i = 1, \ldots, N$.

Once $w_{m,t+1}^i$ are normalized to yield $W_{m,t+1}^i$, the particle set approximates target $t + 1$. The reweighting of particles reduces the effective sample size (ESS) of the particle set (the number of independent draws from the target distribution). Once the ESS (approximated by $1/\sum_{i=1}^{N} (W_{m,t+1}^i)^2$, Kitagawa (1996)) becomes undesirably small (less than $E$), the resampling and move steps are used to replenish the particle set so the ESS is approximately $N$. The resample step involves sampling (with replacement) from the particle set with probabilities proportional to the particle weights which should duplicate particles with relatively large weight and remove particles with relatively small weight. Then, the move step is used to diversify the particle set by applying an MCMC kernel $R$ times with invariant distribution proportional to $p(\theta_m|m, y_{t+1}, d_{t+1})$, and yields $\{\theta_{m,t+1}^i, 1/N\}_{i=1}^{N}$. This SMC algorithm has been implemented in sequential design by Drovandi et al. (2014) for the estimation of parameters in a number of discrete data models.

In our work, a candidate set of $K$ models will be considered where it is assumed that one of the models is responsible for data generation (so called $M$-closed perspective of Bernardo and Smith (2000)). The SMC inference algorithm just described above for model $m$ can be extended to handle model uncertainty by forming particle approximations for each target distribution for each of the $K$ models. In this way, all models will have separate reweight, resample and move steps. Model choice will be performed by the posterior model probability (normalized model evidences) with a preference for the model with the largest of such values. As shown in Del Moral et al. (2006), the model evidence can be efficiently approximated within SMC. It can be shown that the ratio of model evidences (for a given model $m$) is equivalent to the predictive distribution of the new data point $y_{t+1}$ and design $d_{t+1}$, given the current data $y_t$ (and designs $d_t$). That is,

$$\frac{p(y_{t+1}|m, d_{t+1})}{p(y_t|m, d_t)} = \int_{\theta_m} p(y_{t+1}|m, \theta_m, d_{t+1}) p(\theta_m|m, y_t, d_t) d\theta_m.$$

The above expression can be approximated via a particle approximation as follows:

$$\frac{p(y_{t+1}|m, d_{t+1})}{p(y_t|m, d_t)} \approx \sum_{i=1}^{N} W_{m,t}^i p(y_{t+1}|m, \theta_{m,t}^i, d_{t+1}),$$

where we note that $p(y_0|m, d_0) = 1$.

Hence, posterior model probabilities can be efficiently approximated each time a new data point is observed. As will be seen, this approximation is also very useful in efficiently approximating important Bayesian utility functions, see Drovandi et al. (2014).
3. Design methodology

In this section, the total entropy utility is derived. The utility is constructed by considering the change in entropy about the model indicator and the parameter upon the observation of a new data point. Separately, each of these changes in entropy form utility functions for model discrimination and estimation (respectively). Hence, in constructing the total entropy utility, a model discrimination utility and a parameter estimation utility are also derived. In all three cases, the derivations begin with initially defining entropy about the relevant random variable, then by considering the change in entropy upon observing data. In order to implement these utilities in the motivating examples that follow (Section 4), it is shown how to estimate each in the inference framework just outlined in Section 2.

3.1. Entropy

For ease of notation, define entropy for a random variable $X$ as $H(X)$. In the case when $X$ is a discrete random variable,

$$H(X) = - \sum_{x \in X} p(x) \log p(x).$$

Then, in the case where $X$ is a continuous random variable,

$$H(X) = - \int_{x \in X} p(x) \log p(x) \, dx.$$

Technically, the above expression for $H(X)$ when $X$ is a continuous random variable is not correct as a term that tends to infinity in the limiting case of a continuous density function has been omitted. This term represents the infinite amount of information needed to determine $X$ exactly, and, as this is not possible in practice, the term is omitted. In fact, this omission is of no consequence as later, when we consider the change in entropy (about the model parameter) upon the observation of additional data, the terms will cancel out when the difference is taken between the prior and posterior densities (for $\theta$). See Rosie (1966) and Borth (1975) for further discussion.

In order to understand the total entropy utility (and the discrimination and estimation utilities), it is necessary to also understand conditional entropy. Let $H(X|Y)$ denote the conditional entropy of $X$ given the discrete random variable $Y$ in $\mathcal{Y}$ be defined as

$$H(X|Y) = - \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in X} p(x|y) \log p(x|y),$$

and

$$H(X|Y) = - \sum_{y \in \mathcal{Y}} p(y) \int_{x \in X} p(x|y) \log p(x|y) \, dx,$$

for the cases where $X$ is a discrete and continuous random variable, respectively. The cases where $Y$ is a continuous random variable are omitted as they are straightforward to express.

Considering the above expressions for conditional entropy, it can be interpreted as the entropy about $X$ that is expected to remain upon the observation of $Y$. And thus, the reduction of entropy about $X$ upon the observation of $Y$ (that is, $H(X) - H(X|Y)$) can be interpreted as the information gained about $X$ given knowledge of $Y$. This difference is often referred to as the mutual information between $X$ and $Y$. It is this mutual information that has been used to derive discrimination and estimation utilities, and indeed will be used here to construct the total entropy utility.

Following the work of Borth (1975), define total entropy as the sum of the entropy about the model indicator and the parameter. This can be expressed as follows:

$$H_T(M, \theta|y_t, d_t) = H_M(M|y_t, d_t) + H_P(\theta|y_t, d_t),$$

(1)

where $H_M(M|y_t, d_t)$ measures the entropy about which model is correct and $H_P(\theta|y_t, d_t)$ measures the entropy about all parameters $\theta$ across the $K$ models (weighted by the probability that each model is correct). The unique additivity property of entropy was proved by Khinchin (1957).

Box and Hill (1967) gave an expression for the entropy of the model indicator. Following their work, define model entropy as follows:

$$H_M(M|y_t, d_t) = H(M|y_t, d_t).$$

Based on the work of Shannon (1948) and Lindley (1956), an expression for the entropy about the parameter can be defined as follows:

$$H_P(\theta|y_t, d_t) = \sum_{m=1}^{K} p(m|y_t, d_t) H(\theta_m|m, y_t, d_t),$$

which is the sum of the entropy in the posterior distribution of the parameter for each model weighted by the corresponding model probability.

Now that the entropy about the model and parameter (and hence the total entropy) have been defined, the change in entropy in each case upon the observation of a new data point is considered for the construction of utility functions.
3.2. Change in entropy—model discrimination utility

Box and Hill (1967) considered a sequential design setting to discriminate between candidate nonlinear models for a chemical reaction of the type \( A \rightarrow B \). In their work, they gave the following expression for the expected change in entropy about the model indicator upon observing future data \( z \) given a design \( d \):

\[
H_M(M|y_t, d_t) - H_M(M|z; y_t, d_t, d) = H(M|y_t, d_t) - \int H(M|z; y_t, d_t, d)p(z|y_t, d_t, d)dz,
\]

where \( p(z|y_t, d_t) = \sum_{m=1}^{K} p(m|y_t, d_t)p(z|m, y_t, d_t, d) \), and measures the total uncertainty about the outcome comprised of uncertainty about the model and the parameter, and also experimental error.

From Eq. (13) in Borth (1975), the above expression can be simplified to

\[
H(Z|y_t, d_t, d) - \sum_{m=1}^{K} p(m|y_t, d_t)H(Z|m, y_t, d_t, d).
\]

Interpreting this representation, one seeks a design which maximises the total uncertainty in the outcome \( Z \) minus the uncertainty in \( Z \) conditional on the model.

Drovandi et al. (2014) derived a slightly different formulation of this utility, and extended it for use within an SMC algorithm for sequential design with applications to discrete data settings. In their work, the utility is given by

\[
\sum_{m=1}^{K} p(m|y_t, d_t) \sum_{z \in Z} p(z|m, y_t, d_t, d) \log p(m|y_t, d_t, z, d),
\]

where \( z \in Z \) represents all possible discrete values of the future outcome and \( p(z|m, y_t, d_t, d) \) is formed by integrating out the parameters from the likelihood for model \( m \).

Under this formulation, one seeks to find the design \( d \) which maximizes the expectation of the logarithm of the posterior model probability. A Bayes risk type formulation of this utility was given by Cavagnaro et al. (2010), and applied to experiments in cognitive science.

3.3. Change in entropy—estimation utility

In terms of the change in entropy about the parameter, we consider the change in entropy from the prior to the posterior distribution of the parameter upon the observation of an additional data point (Lindley, 1956). From Eq. (19) in Borth (1975), the change in entropy for the parameter can be expressed as follows:

\[
H_P(\theta|y_t, d_t) - H_P(\theta|Z; y_t, d_t, d) = \sum_{m=1}^{K} p(m|y_t, d_t) \left[ H(Z|m, y_t, d_t, d) - \int p(\theta_m|m, y_t, d_t, d)H(Z|m, y_t, d_t, d)d\theta_m \right].
\]

Then, in using this utility, the selected designs will be those which are in the area of the largest expected uncertainty about the outcome \( Z \) conditional on the model after removing the uncertainty due to experimental error.

Lindley (1956) provided many alternative formulations of this utility which in this framework include the following:

\[
\sum_{m=1}^{K} p(m|y_t, d_t) \int \int p(z, \theta_m|m, z, y_t, d_t, d) \log \left[ \frac{p(z, \theta_m|m, z, y_t, d_t, d)}{p(z|m, \theta_m, y_t, d_t, d)p(\theta_m|m, y_t, d_t)} \right] dz d\theta_m.
\]

This formulation shows the symmetry between the outcome \( z \) and \( \theta_m \), and shows the invariance of this utility to re-parameterizations of \( \theta_m \).

Lindley also showed that any design will be informative (in expectation). However, despite the gain in information being positive in expectation, there is the possibility that an experiment may reduce the amount of information gained. This could occur upon the observation of ‘surprising’ values of the data resulting in the experimenter being less sure about the parameter values.

The utility can be formulated as the sum of the Kullback–Liebler divergences (Kullback and Leibler, 1951) between the prior and posterior distributions of the parameter weighted by the model probability. The utility can be formulated under the framework of this paper as follows:

\[
\sum_{m=1}^{K} p(m|y_t, d_t) \int \int p(\theta_m|m, z, y_t, d_t, d) \log \frac{p(z|m, \theta_m, y_t, d_t, d)}{p(z|m, y_t, d_t)} d\theta_m dz.
\]

The above utility has been applied within an SMC framework for estimation of a single model by Drovandi et al. (2013).
Singly, the model discrimination and estimation utilities work as intended for experiments with only one experimental aim. However, it must be said that, if one thinks about running an actual experiment, it seems reasonable that one would seek to obtain the most information possible from each data point, and therefore focusing only on either model discrimination or estimation may not be the most sensible experimental aim. In fact, each utility will prefer the largest reduction in uncertainty about the respective random variables with total disregard for other potential gains in information. In practice, one may prefer slightly reduced information about a given random variable for potentially large gains in information about another random variable. This leads to the consideration of the total entropy utility.

3.4. Change in entropy—total entropy utility

From Eq. (1), combining the change in entropy about the model (Eq. (3)) and parameter (Eq. (4)) yields the change in total entropy which can be expressed as follows:

\[ H_T(M, \theta | y_t, d_t) = H_T(M, \theta | y_t, y_t, d_t) = H(Z | y_t, d_t) - \sum_{m=1}^K p(m | y_t, d_t) \int p(\theta_m | m, y_t, d_t) H(Z | \theta_m; m, y_t, d_t) d\theta_m. \]  

(5)

see Eq. (20) in Borth (1975).

Hence, this utility will prefer designs for which the outcome \( Z \) is most uncertain after removing the uncertainty due to experimental error. Such designs are considered most informative under this utility.

3.5. Estimating the utility functions

It is now shown how to estimate the model, parameter and total entropy utilities in an SMC framework. Assume a set of weighted particles \( \{(\theta_{m,t}^i, W_{m,t}^i)\}_{i=1}^N \) approximates the current target distribution at time \( t \), for \( m = 1, \ldots, K \). Upon the proposal of a new design point \( d \), let \( z \) denote a possible value for the future data point \( y_{t+1} \).

3.5.1. Model discrimination utility

The first term in Eq. (3) is:

\[ - \int p(z | y_t, d_t) \log p(z | y_t, d_t) dz. \]  

(6)

Note that \( p(z | m, y_t, d_t) \), the predictive probability of \( z \), can be approximated for model \( m \) as follows:

\[ \hat{p}(z | m, y_t, d_t) = \sum_{i=1}^N W_{m,t}^i p(z | m, \theta_{m,t}^i; y_t, d_t). \]

Then \( \hat{p}(z | y_t, d_t) \) can be obtained by averaging the above over the models as follows:

\[ \hat{p}(z | y_t, d_t) = \sum_{m=1}^K \hat{p}(m | y_t, d_t) \hat{p}(z | m, y_t, d_t), \]

where \( \hat{p}(m | y_t, d_t) \) is the estimated model probability.

Then, Eq. (6) can be approximated as follows:

\[ - \sum_{z \in Z} \hat{p}(z | y_t, d_t) \log \hat{p}(z | y_t, d_t), \]

where \( z \in Z \) represents all possible values of \( z \) in a discrete data setting.

Similarly, the last term in Eq. (3) can be approximated as follows:

\[ - \sum_{z \in Z} \hat{p}(z | m, y_t, d_t) \log \hat{p}(z | m, y_t, d_t). \]

Then finally the model discrimination utility (Eq. (3)) can be approximated as such:

\[ - \sum_{z \in Z} \hat{p}(z | y_t, d_t) \log \hat{p}(z | y_t, d_t) + \sum_{m=1}^K \hat{p}(m | y_t, d_t) \sum_{z \in Z} \hat{p}(z | m, y_t, d_t) \log \hat{p}(z | m, y_t, d_t). \]

3.5.2. Estimation utility

Using the same argument as above, the estimation utility (Eq. (4)) can be approximated as follows:

\[ - \sum_{m=1}^K \hat{p}(m | y_t, d_t) \left( \sum_{z \in Z} \hat{p}(z | m, y_t, d_t) \log \hat{p}(z | m, y_t, d_t) + \sum_{i=1}^N W_{m,t}^i \left( \sum_{z \in Z} p(z | m, \theta_{m,t}^i; y_t, d_t) \log p(z | m, \theta_{m,t}^i; y_t, d_t) \right) \right), \]
where
\[
\sum_{i=1}^{N} W_{m,t}^{i} \left( - \sum_{z \in Z} p(z|m, \theta_{m}^{i}, \mathbf{y}_{t}, \mathbf{d}_{t}) \log p(z|m, \theta_{m}^{i}, \mathbf{y}_{t}, \mathbf{d}_{t}) \right)
\]
is the particle approximation for discrete data of the expected entropy
\[
\int p(\theta_{m}|m, \mathbf{y}_{t}, \mathbf{d}_{t}) H(Z|m, \mathbf{y}_{t}, \mathbf{d}_{t}) \theta_{m}.
\]

3.5.3. Total entropy utility

From the above results for discrete data, the total entropy utility (Eq. (5)) may be approximated as follows:
\[
- \sum_{z \in Z} \hat{p}(z|\mathbf{y}_{t}, \mathbf{d}_{t}) \log \hat{p}(z|\mathbf{y}_{t}, \mathbf{d}_{t}) + \sum_{m=1}^{K} \hat{p}(m|\mathbf{y}_{t}, \mathbf{d}_{t}) \sum_{i=1}^{N} W_{m,t}^{i} \left( \sum_{z \in Z} p(z|m, \theta_{m}^{i}, \mathbf{y}_{t}, \mathbf{d}_{t}) \log p(z|m, \theta_{m}^{i}, \mathbf{y}_{t}, \mathbf{d}_{t}) \right).
\]

4. Examples

In the three examples that follow, competing models for binary and count response data will be considered in a sequential design setting. Three utilities for design selection will be considered: total entropy (Eq. (5)); the estimation utility (Eq. (4)); and the model discrimination utility (Eq. (3)). Results will also be compared to the random design where designs are selected completely at random. In each case, an optimal design will be found for the observation of the next data point, and, once this data point has been observed, all prior information will be updated. To do this, a model from the candidate set used to generate data (with parameter values being the mode of the relevant prior information about the parameter) in all examples, each model in the candidate set was considered for data generation. However, only a selection of results will be presented.

At each iteration of each simulation study, the posterior model probability and the posterior variance–covariance of the parameter for the model responsible for data generation will be recorded. As each study will be simulated a large number of times, a distribution of posterior model probabilities and, say, determinants of the posterior variance–covariance matrix will be available, and this is what will be used to compare the utilities. The selected designs over the whole simulation study will also be recorded and used to compare the utilities.

In order to explore the performance of each utility across a wide range of problems, three scenarios were considered in each example. These scenarios differ in the variability assumed in the prior information about the parameter, and will consequently be termed informative, vaguely informative and highly informative scenarios. Specifically, independent normal distributions will be assumed about each parameter (in a given model) with a variance of 4, 25 and 0.25 for the above scenarios, respectively. In particular for total entropy, it is thought that investigating scenarios ranging from vaguely to highly informative prior information may provide the opportunity for the utility to trade-off learning about the model responsible for data generation and estimating the parameters. For example, in the highly informative scenario, total entropy may focus on model discrimination as the parameters are already quite certain. Similarly, in the vaguely informative scenario, total entropy may focus on estimation (rather than discrimination). Further, these scenarios may also reveal if total entropy can efficiently learn about both the model and the parameter or if a trade-off significantly compromises information about one experimental aim for another. Again, although all three scenarios were considered in the simulation studies, only a selection of results will be presented.

In all three examples, the models are constructed through the generalized linear modeling framework (McCullagh and Nelder, 1989), with the exception of Example 1 where a generalized nonlinear model is also considered. Of note is the recent interest in the optimal design for such models, for example, see Woods et al. (2006), Russell et al. (2008), Dror and Steinberg (2008), Beidermann and Woods (2011), Yang et al. (2011), Wu and Stufken (2014) and Atkinson and Woods (2015). However, the developments presented in this paper are not restricted to such models. Implementations for more complex models such as those with random effects are possible, and this is discussed as an area for further research in Section 5.

The examples begin with two competing models for binary response data being motivated by clinical trials in drug development. The second example considers three competing Poisson regression models for count data. And lastly, the third example considers a screening-type scenario for binary data where some or all covariates may significantly influence the probability of success. For design selection in all three examples, a discrete set of possible designs is proposed, and an exhaustive search is used to determine which is selected as optimal. In all simulation studies, the following values of the tuning parameters in the SMC algorithm were chosen: \( N = 1000, R = 10 \) and \( E = 750 \), and each model in a given candidate set was assumed to be equally likely \emph{a priori}. The Matlab code for running these simulation studies is available from the author upon request.
Fig. 1. The 5th (--), 50th (–) and 95th (--) percentiles of the prior predictive distribution of data from the (a) logistic and (b) emax model for Example 1 in the highly informative scenario.

4.1. Example 1

Motivated by the work of Dette et al. (2008), Pavan Kumar and Duffull (2011) and Drovandi et al. (2014), consider the following competing dose–response models in drug development:

Logistic model: \[ \log \left( \frac{p_t}{1 - p_t} \right) = \theta_0 + \theta_1 d_t \]

Emax model: \[ \log \left( \frac{p_t}{1 - p_t} \right) = \theta_0 + \theta_1 d_t / (\theta_2 + d_t) \]

where \( d_t \in \{0.1, 0.2, \ldots, 5\} \).

For this example, the prior means for the logistic and emax model were selected to be \((-5, 5\) and \((-5, 10, 1\) respectively, and the results for the highly informative scenario were selected for demonstration and comparison of the performance of the utilities. Fig. 1 shows the 5th, 50th and 95th percentiles of the prior predictive distributions of data from the logistic and emax model for the range of designs considered in this example. As can be seen, the logistic model reaches a median probability of success of 1 for doses of around 2. However, for the emax model, the median actually never reaches a probability of success of 1. Thus, larger doses may be useful for discrimination. Further, there appears to be more variability associated with the predicted responses for the emax model when compared to the logistic model.

For the highly informative scenario, Fig. 2 shows the 5th, 50th and 95th percentiles of the distribution of the posterior model probabilities at each iteration of 500 simulated trials each yielding 100 data points. The results shown in the first row are for when the logistic model was responsible for data generation. Here, total entropy appears to perform similarly to the model discrimination utility but is less efficient and slightly more variable. Further, total entropy performs better than the random design. The estimation utility yields designs which are somewhat able to discriminate between the two models reaching a median posterior model probability of around 0.8 after all data have been observed. The second row of Fig. 2 shows results for when the emax model was responsible for data generation. In this case, the estimation utility performs poorly with no indication of being able to discriminate between the two models. In terms of the median posterior model probability, total entropy and the random design perform well when compared to the model discrimination utility. It seems reasonable for the random design to discriminate well as it will eventually select designs which cover the entire design region, and may do so rather quickly in this one dimensional setting. For total entropy, the performance is quite variable until all data have been observed.

Fig. 3 shows the 5th, 50th and 95th percentiles of the distribution of the logarithm of the determinant of the posterior variance–covariance matrix for the model responsible for data generation at each iteration of 500 simulated trials with highly informative prior information. The results shown in the first and second row are for when the logistic and emax models were responsible for data generation, respectively. The performance of total entropy appears to be very similar to the estimation utility, particularly when the emax model was responsible for data generation. When the logistic model was generating data, total entropy is slightly more variable in performance than the estimation utility. We note that the estimation utility actually performed quite poorly in terms of model discrimination when either model was generating data. Hence, the slightly more variable performance of total entropy may indicate the utility is selecting designs which can discriminate between the two models, particularly as the parameters are relatively quite certain in this scenario. The performance of the random design seems to depend upon which model was generating data. That is, it performed reasonably well when the emax model was generating data, but performed poorly when the logistic model was generating data.

Across all scenarios, the model discrimination and estimation utilities were not consistent in either being able to discriminate between competing models or providing efficient parameter estimates. The random design performs well but is inefficient (particularly for estimation). In contrast, total entropy was the only utility which performed well for model discrimination and estimation across all scenarios. Moreover, in cases where either the discrimination or estimation
utility performed poorly, total entropy was able to perform well under both experimental aims with very little compromise between the two objectives.

Fig. 4 shows all selected doses in the 500 simulations for the three utilities and the random design in Example 1 with highly informative prior information. The plot shows that there are notable differences between the selected designs across all utilities. For total entropy, a bimodal plot is observed. This is also observed in a similar area of the design space in the designs selected by the model discrimination utility. However, for the estimation utility, only a single mode appears in part of this design region. Hence, there appears to be an area of the design space which is preferred by both the model discrimination and estimation utilities, and this is exploited by total entropy in yielding designs to address each objective. The second modal region in the designs selected by total entropy only appears in the designs for discrimination showing a preference for such designs throughout the simulation study. Moreover, the avoidance of this region by the estimation utility may explain why this utility was the worst performer in terms of discrimination. In the informative and vaguely informative scenarios, the designs selected by total entropy appeared very similar to those selected by the estimation utility. As such, in cases where the parameters are relatively uncertain, total entropy may be focussing more on estimation. However, the selection of additional design points in regions of preference for the model discrimination utility reveals how the utility is also addressing this aim.

4.2. Example 2

Poisson regression is considered in this example where it is thought that at least one of two covariates influences the average of the count data. As such, the candidate set of models can be described as follows:

Model 1: \( \log \lambda = \theta_0 + \theta_1 d_{t,1} \)
Model 2: \( \log \lambda = \theta_0 + \theta_2 d_{t,2} \)
Model 3: \( \log \lambda = \theta_0 + \theta_1 d_{t,1} + \theta_2 d_{t,2} \),

where \( d_{t,1}, d_{t,2} \in \{-1, -0.75, \ldots, 0.75, 1\} \) for \( t = 1, \ldots, T \).

The array of potential designs is then constructed by considering all possible combinations of \( d_{t,1} \) and \( d_{t,2} \). Hence, a total of 81 potential designs will be considered for design selection.
Fig. 3. The 5th (--), 50th (–) and 95th (--) percentiles of the distribution of the logarithm of the determinant of the posterior variance–covariance matrix for the model responsible for data generation over 500 simulated trials with different utility functions for Example 1 in the highly informative scenario when the logistic model was generating data (first row) and when the emax model was generating data (last row).

Fig. 4. Selected doses over 500 simulated trials with different utility functions for Example 1 in the highly informative scenario when the logistic model was generating data (first row) and when the emax model was generating data (second row).
The results for the cases where Model 1, 2 and 3 were responsible for data generation are presented. In this case, the following prior means were assumed: \( \theta = (\theta_0, \theta_1, \theta_2)' = (-1, 2, 2)' \), where relevant. As in the previous example, the informative, vaguely informative and highly informative scenarios were assumed. That is, independent normal prior distributions with variances 4, 25 and 0.25. However, only results for the highly informative case will be shown.

Fig. 5 shows the 5th, 50th and 95th percentiles of the posterior model probability distributions at each iteration of 500 simulated trials each with 50 data points in the highly informative scenario. The first, second and third rows correspond to when Model 1, 2 and 3 were responsible for data generation, respectively. Here, total entropy and the model discrimination utility are the best performers requiring very few data points to determine which model was generating data, regardless of whether this was Model 1, 2 or 3. In contrast, the random design in general requires 10 to 15 data points to determine which model is true. However, the performance is quite variable as the 5th percentile of the distribution does not reach 1 when Model 1 and 2 were generating data. Moreover, in these two cases, the estimation utility yields little to no information about which model was generating data until around 20 data points have been observed. After this point, the utility appears to select designs which yield such information as the posterior model probabilities rapidly increase towards 1.

In the scenarios with informative and vaguely informative prior information, total entropy and the model discrimination and estimation utilities perform similarly well for discrimination each with an initial steep incline towards a posterior model probability of 1 across all three models. This initial steep incline was particularly observed in the cases when Model 3 was generating data. For the random design, the performance for discrimination was relatively inefficient for discrimination overall.

Fig. 6 shows the 5th, 50th and 95th percentiles of the distribution of the logarithm of the determinant of the posterior variance–covariance matrix for the model responsible for data generation for different utility functions at each iteration of the simulation study in the highly informative scenario. Interestingly, total entropy and the estimation utility perform similarly. If the discrimination results are recalled (Fig. 5), the estimation utility initially struggled to determine which model was generating data while total entropy was able to determine this rather quickly. Hence, the ability to efficiently discriminate between competing models did not compromise the performance of total entropy in estimation. Thus, total entropy may be exploiting a commonality in the design regions for estimation and model discrimination to select designs which perform well under both utilities. This certainly appears reasonable when inspecting the selected designs (see Fig. 7, which will be discussed later).

The performance of the random design for estimation was more variable than the utility functions and seems to depend upon which model was generating data. Specifically, it performed reasonably well when Model 1 and 2 were generating data, but performed rather poorly when Model 3 was generating data. As such, the generally strong performance of the random design seen in Example 1 is not seen here. Indeed, it seems quite reasonable that the random design will become more and more inefficient as the dimension of the design space increases.

Fig. 7 shows all selected design points over 500 simulated studies for Example 2 with highly informative prior information. Note that a small amount of noise was added to the selected designs to make replicate points more visible. It would appear as though the selected designs under total entropy are very similar to those selected by the estimation utility. Additionally, when Model 1 and 2 were generating data, selected designs appear to overlap with design regions preferred by the discrimination utility. This may seem a little strange as recall that, in this case, the estimation utility yielded little to no information about which model was generating data until about the 20th observation. However, the plots do not show the time ordering of the selected designs, which is important here. This is demonstrated with an additional plot of selected designs over 500 trials but only for the first 15 iterations of the sequential design process, see Fig. 11 in the Appendix. Clearly, designs in the region of \( d_1 = 1 \) and \( d_2 = -1 \) are essentially avoided by the estimation utility in the first 15 iterations, but frequently visited by total entropy and the discrimination utility. It is only after the first 15 observations that the estimation utility selects many design points in this region (which would appear to yield information about which model was generating the data). It would also appear that this region of the design space yields information about the parameter values, and thus total entropy was able to exploit this to determine which model was generating data early on in the trial while still providing precise estimates of parameters. Similar statements can be made when Model 2 was generating data.

4.3. Example 3

In this example, the logistic model is again considered. However, here four covariates were considered where it is known that at least one covariate influences the probability of success. As such, there is a total of 15 models in the candidate set (4 one and three covariate models, 6 two covariate models and the full main effects model). Hence, there is considerably more uncertainty about the model in this example when compared to the previous examples. The full model is given below:

\[
\log \left( \frac{p_t}{1-p_t} \right) = \theta_0 + \theta_1 d_{t,1} + \theta_2 d_{t,2} + \theta_3 d_{t,3} + \theta_4 d_{t,4}, \quad \text{for } t = 1, \ldots, T,
\]

where \( d_{t,1}, d_{t,2}, d_{t,3}, d_{t,4} \in \{-1, -0.75, \ldots, 0.75, 1\} \).
Again, the array of potential designs is constructed by considering all possible combinations of each covariate. Hence, a total of 6561 potential designs will be considered for design selection. To reduce the computing time in selecting the optimal design, all proposed designs were evaluated in parallel (using 8 computer processing units).

For all simulation studies, the following prior means were assumed for each model: \( \theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)' = (0, -3, 3, -3, 3)' \), where relevant. The prior means suggest that, if a covariate is influential, then it will have a reasonable effect on the probability of success (odds ratio of about 20 for a unit change in the covariate). As in the previous two examples, informative, vaguely informative and highly informative prior information about the parameter values were assumed.

The results selected for demonstration and comparison of the performance of the utilities relate to the above full main effects model without \( d_{1,3} \) being included in the model (denoted as Model 12). Hence, for the results presented here, a three factor main effects model was responsible for generating data throughout the simulation studies.

Fig. 8 shows the 5th, 50th and 95th percentiles of the posterior model probability distributions at each iteration of 500 simulated trials each with 150 data points when Model 12 was generating data across the informative, vaguely informative and highly informative scenarios. In the first two cases, all utilities appear quite similar. The model discrimination utility provides the least variable results followed by total entropy, then the estimation utility and then the random design.
However, in the highly informative scenario, the estimation utility generally has difficulty in determining which model was responsible for data generation. In fact, the random design outperforms the estimation utility in this regard, and both total entropy and the discrimination utility perform well requiring only around 15 data points to determine the true model.

Fig. 9 shows the 5th, 50th and 95th percentiles of the distribution of the logarithm of the determinant of the posterior variance–covariance matrix for Model 12 for different utility functions at each iteration of the simulation study with informative, vaguely informative and highly informative prior information. The results from total entropy and the estimation utility appear very similar, with the discrimination utility and the random design being less efficient and more variable. Hence, total entropy is able to perform very well for discrimination and estimation despite the model discrimination and estimation utilities performing relatively poorly under each objective. Further, this performance is under considerable uncertainty about the model responsible for data generation.

Fig. 10 shows plots of the covariate levels selected in the case of highly informative prior information by total entropy (a to f), the estimation utility (g to l) and the model discrimination utility (m to r) when Model 12 was responsible for
Fig. 7. Selected designs over 500 simulated trials for Example 2 in the highly informative scenario when Model 1 was generating data (first row), when Model 2 was generating data (second row) and when Model 3 was generating data (third row).

data generation. Again, a small amount of noise was added to the selected designs for plotting purposes. Firstly, the plots may be a little misleading as they do not show the full picture in terms of the designs selected but rather just show the pairwise relationships between each covariate, regardless of the levels of the other covariates. Nonetheless, there are clear similarities between plots (b), (d) and (f) and plots (n), (p) and (r) in Fig. 10 where designs appear to contrast the response against extreme values of $d_3$ at a variety of values of all other covariates. Interestingly, $d_3$ is the only covariate that does not appear in Model 12. Hence, it seems reasonable that contrasting the response against extreme values of $d_3$ will yield data from which one could determine whether $d_3$ should be in the model or not. The other plots in Fig. 10 appear to show designs selected at random, and may be a result of any design in this pairwise comparison yielding little information for discrimination once there is certainty about whether the covariate should be included in the model. Presumably, early on in the trial, designs which contrast the response against extreme levels of these covariates would be observed but are eventually masked in these plots once certainty about inclusion is determined. Further, there are clear similarities between plots (a), (c) and (e) and plots (g), (i) and (k) in Fig. 10. In each of the plots, there is a preference for designs on the boundaries of the covariate space of $d_1$, $d_2$ and $d_4$ which are the covariates that appear in Model 12. Hence, this suggests total entropy is selecting these designs for the estimation of parameters which actually appear in the true model (Model 12).

The fact that the selected designs by total entropy have features seen in designs selected by the estimation and model discrimination utilities suggests that the utility is addressing both experimental goals, and this is certainly seen in the estimation and discrimination results. Recall that in this case total entropy performed similarly to the model discrimination utility for discrimination, and also similarly to the estimation utility for estimation (despite each utility selecting designs which are relatively inefficient under the other utility).

5. Conclusion

In this work, the total entropy utility function was considered for the dual experimental goals of model discrimination and parameter estimation in Bayesian design. The utility simultaneously addresses these two experimental goals through the unique additivity property of entropy about the relevant random variables. A framework for efficient Bayesian inference and design was also proposed such that total entropy could be implemented in a computationally feasible manner. This utility was demonstrated through a number of motivating examples where it appeared that total entropy selected designs which were highly efficient for both model discrimination and parameter estimation when compared to the optimal designs under utility functions for either experimental goal. Indeed, this high efficiency appeared to come with very little compromise in performance under either experimental goal. The empirical evidence presented in this paper is by no means a proof, but given that it appears possible to yield such efficient designs under both experimental goals, it seems reasonable to advocate using total entropy in design for the range of problems considered in this paper.
In this work, the design space was discretized, and, when needed, utilities were evaluated in parallel. This worked well for the design problems considered in this paper. However, further developments are needed for the consideration of more covariates and a continuous design space, and will unfortunately impose further computational challenges. Indeed, such challenges are highlighted as areas of future research in a recent review of Bayesian design (Ryan et al., 2016). General purpose optimization approaches such as nature inspired metaheuristic algorithms (Yang, 2013) have been successful in solving a wide variety of real and complicated optimization problems, and may prove useful in efficiently locating optimal designs. Recently, such an algorithm has been used to locate optimal designs for mixture models (Wong et al., 2011), and general use within Bayesian design is an area for further research. Alternatively, the approximate coordinate exchange algorithm proposed by Overstall and Woods (2015) could also be considered. In essence, the algorithm is a coordinate exchange algorithm where marginal utility evaluations are approximated via a Gaussian Process such that a wide variety of designs can be efficiently considered.

One might also consider extensions to more complex models, including, for example, random effect models. McGree et al. (2015) extend the use of the SMC algorithm to random effect models via unbiasedly approximating the likelihood in the reweight and move steps, and also in approximating the expectation of the utility of a given design. Such an algorithm could be adopted here as an inference framework for, for example, generalized linear mixed models, and one could develop an approximation to the total entropy utility function in a similar manner to what is shown in this paper.

In the presence of model and parameter uncertainty, sequential designs are generally more efficient than static (non-sequential) designs in achieving experimental goals as the design choice can be updated as more information is accrued.
Then, in regards to efficiency, it is worth considering potential stopping rules for sequential designs. The exact form of the stopping rule will depend upon the context of the experiment. For example, in clinical trials, stopping rules may involve lack of safety and efficacy considerations (Stallard et al., 2001) and/or be based on futility studies (Lachin, 2005). For total entropy more generally, Borth (1975) noted the difficulty in defining a stopping rule but suggested that insight into whether future experimentation is worthwhile could be provided by comparing the entropy about the response given the optimal design and the entropy about the experimental error distribution.

Of course, total entropy need not be limited to sequential design settings, and could be considered for the location of static designs. The major challenge here will be handling the generally large distance between the prior and posterior distributions. Hence, efficient posterior approximations are required. For this purpose, it might be useful to consider deterministic approximations such as the Laplace approximation, the integrated nested Laplace approximation (Rue et al., 2009) or Variational Bayes (Beal, 2003). In particular, as the Laplace approximation is a Gaussian distribution, the total entropy utility (and the discrimination and estimation utilities) may have an analytic form.
In reviewing the results from implementing the total entropy utility, there were occasions where an initial dip in the posterior model probabilities was observed. This occurred in Example 1 when the emax model (the more complex model) was responsible for data generation (results omitted). This may be a consequence of the posterior model probabilities favoring the simpler model when only a small number of data points have been observed (and of course a consequence of the assumed prior information). Despite this not being observed in the other examples considered here, one wonders whether this could be observed in other examples, and further, if it leads to sub-optimality. If this is the case, one could consider initially maintaining the prior model probabilities until a reasonable number of data points have been collected.

Alternatives to the total entropy utility for the construction of designs for dual experimental goals appear in the classical design literature, and were discussed in the introduction to this paper. Although one could implement one of these approaches as a comparator to the total entropy utility, this was not pursued as none of the cited approaches exploit the unique additivity property of entropy, and hence do not trade-off the aims of model discrimination and estimation in a very natural way. A further reason for not undertaking such comparisons is that many of the cited approaches require the choice of a weighting parameter which focuses design selection on both or either of the two utilities. As noted in Clyde and Chaloner...
(1996) and McGree et al. (2008), because of the scaling of the utilities, an appropriate choice of this weighting parameter can be quite difficult in general. Indeed, if one chooses this parameter incorrectly, then too much emphasis could inadvertently be placed on either utility limiting the potential information gain from the experiment. When implementing total entropy, there is an implicit weighting of utilities based solely on the prior information, and thus this weighting will generally be informed by expert elicited data and/or previously collected data. This could be explored a priori by considering the relative entropy of the model probabilities and model parameters. Such exploration may be informative in determining the expected performance of the total entropy utility function when implemented in, for example, large-scale design problems.

Lastly, as an interesting aside to this research, the designs selected in Example 2 by the estimation utility appear to resemble features of corresponding optimal pseudo-Bayesian designs found by maximizing the determinant of the expected Fisher information matrix as given by the analytic solutions of Russell et al. (2009) and McGree and Eccleston (2012). In particular, if such methodology is applied here, the optimal design points would be located at either 1 or around 0 for factors included in the model. Such features appear in all selected designs across the three models, see Fig. 7. This suggests that it might be worth exploring whether an analytic solution for Bayesian optimal designs for parameter estimation based on the Kullback–Liebler divergence could be derived in settings similar to that considered in this research.

Appendix. Additional figure for Example 2

See Fig. 11.

Fig. 11. Selected designs over the first 15 iterations of the sequential design process over 500 simulated trials for Example 2 in the highly informative scenario when Model 1 was generating data (first row), when Model 2 was generating data (second row) and when Model 3 was generating data (third row).

References

Atkinson, A.C., Fedorov, V.V., 1975a. The design of experiments for discriminating between two rival models. Biometrika 62, 57–70.


Model selection via Bayesian information capacity designs for generalised linear models

David C. Woods\textsuperscript{a,}\textsuperscript{*}, James M. McGree\textsuperscript{b}, Susan M. Lewis\textsuperscript{a}

\textsuperscript{a} University of Southampton, UK
\textsuperscript{b} Queensland University of Technology, Australia

\textbf{A R T I C L E I N F O}

\textbf{Article history:}
Received 29 January 2016
Received in revised form 22 October 2016
Accepted 26 October 2016
Available online 2 November 2016

\textbf{Keywords:}
Bayesian D-optimality
Factorial experiments
Generalised information criterion
Screening

\textbf{A B S T R A C T}

The first investigation is made of designs for screening experiments where the response variable is approximated by a generalised linear model. A Bayesian information capacity criterion is defined for the selection of designs that are robust to the form of the linear predictor. For binomial data and logistic regression, the effectiveness of these designs for screening is assessed through simulation studies using all-subsets regression and model selection via maximum penalised likelihood and a generalised information criterion. For Poisson data and log-linear regression, similar assessments are made using maximum likelihood and the Akaike information criterion for minimally-supported designs that are constructed analytically. The results show that effective screening, that is, high power with moderate type I error rate and false discovery rate, can be achieved through suitable choices for the number of design support points and experiment size. Logistic regression is shown to present a more challenging problem than log-linear regression. Some areas for future work are also indicated.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

\textbf{1. Introduction}

An important problem in scientific discovery is to find those variables (or factors) that have a substantive influence on an observed response through experiments on a possibly large set of potentially important variables. There has been much research into such variable screening, or model selection, focussed on the design and analysis of experiments in which the response variable is adequately approximated by a linear model (see Draguljić et al., 2014 and Woods and Lewis, 2016, and references therein). Such experiments are used increasingly in scientific research and product development, for example, in the pharmaceutical and chemical industries.

In many practical applications, for example when binary or count data are observed, a generalised linear model (GLM; McCullagh and Nelder, 1989) may be needed to describe a response. Previous research on designs for model selection for GLMs has focussed on experiments involving only a few variables through pairwise comparisons of a small number of models (see, for example, López-Fidalgo et al., 2007 and Waterhouse et al., 2008). Hence, such methods are not applicable to, or easily generalisable for, the screening problem. In the literature, the majority of multi-variable experimentation with GLMs has employed (fractional) factorial designs, including examples on solder-joint defects (Hamada and Nelder, 1997), windshield moulding, non-conforming tiles and semi-conductor defects (see Lewis et al., 2001). Although such designs are effective for both model selection and estimation for normal-theory linear models, they have been shown to be inefficient for experiments that provide non-normal data (Woods et al., 2006).
In common with other non-linear models, for the GLMs considered in this paper the performance of a design depends on the unknown values of the parameters in the model. One approach to overcoming this problem is to assume a particular value for each parameter and hence obtain a “locally optimal” design; that is, a design that is optimal under a given criterion provided the assigned parameter values are correct. We adopt the alternative approach of making the less stringent assumption of a priori distribution for each model parameter from which we obtain a “pseudo-Bayesian” design (Atkinson and Woods, 2015).

In this paper, we investigate variable screening for GLMs with \( q \) independent variables, labelled \( x_1, \ldots, x_q \). In the \( j \)th run \((j = 1, \ldots, N)\) of the experiment, a treatment or combination of variable values \( x_j = (x_j, \ldots, x_q)^T \) is applied to an experimental unit and a univariate response, \( y_j \), is observed. We assume that \(|x_j| \leq 1\) for \( i = 1, \ldots, q; j = 1, \ldots, N \).

The aim of the experiment is to identify those active variables having a substantial effect on the response variable and to estimate efficiently a GLM involving those variables alone. For \( j = 1, \ldots, N \), the \( y_j \) have independent exponential family distributions with expectation \( \mu_j \) related to a linear predictor \( \eta_j = f(x_j)^T \beta \) via a link function, \( g(\mu_j) = y_j \). The vectors \( f(x) \) and \( \beta \) are \( p \times 1 \) vectors of known functions of \( x \) and unknown model parameters, respectively. We also assume that the experimental units are exchangeable, in the sense that the distribution of the response to a treatment does not depend on the unit to which the treatment is applied.

For canonical link functions, the log-likelihood may be written as

\[
  l(\beta; y) = \sum_{j=1}^{N} [y_j \eta_j - b(\eta_j) + c(y_j)],
\]

where \( b(\cdot) \) and \( c(\cdot) \) are known functions of the linear predictor and response, respectively. For the binomial distribution and the logistic link, \( b(\eta_j) = -\eta_j \log(1 + e^{\eta_j}) \) and \( c(y_j) = \log(y_j!/(\eta_j!(\eta_j - y_j)!)) \), with \( \eta_j \) the number of Bernoulli trials made at the \( j \)th run. For the Poisson distribution and the log link, \( b(\eta_j) = e^{\eta_j} \) and \( c(y_j) = -\log(y_j!) \).

Maximum likelihood estimators (MLEs) \( \hat{\beta} \) can be found via (numerical) maximisation of (1). For small data sets, however, the MLEs may have considerable bias. For sparse data, such as binomial data with small numbers, \( \eta_j \), of trials for each run, one or more maximum likelihood estimates may be infinite, for example, as the result of separation of the responses into zeros and ones via a hyperplane in the linear predictor (Silvapulle, 1981). To remove this bias and guarantee the existence of estimates for GLMs with a canonical link function, Firth (1993) defined penalised maximum likelihood estimators \( \tilde{\beta} \) as maximisers of

\[
  l'(\beta; y) = l(\beta; y) + \frac{1}{2} \log \det \{X^TWX\},
\]

where \( X \) is the \( N \times p \) model matrix with \( j \)th row \( f(x_j)^T \) and \( W = \text{diag}(\text{var}(y_j)) \) (see also Kosmidis and Firth, 2009). This estimation procedure is equivalent to finding the posterior mode of \( \beta \) assuming the Jeffreys prior distribution.

The information matrix \( X^TWX \), which is the asymptotic inverse variance–covariance matrix for both \( \hat{\beta} \) and \( \tilde{\beta} \), is used to define the \( D \)-optimality criterion. This criterion specifies selection of a design that maximises the objective function

\[
  \phi_D(\xi) = \frac{1}{\beta} \log \det \{X^TWX\},
\]

where \( \xi = \{x_1, \ldots, x_n\} \) and \( \omega_1, \ldots, \omega_n \) are the distinct treatments in the design (assumed, without loss of generality, to be applied to the first \( n \) runs of the experiment), \( \omega_k > 0 \) is \( \mathbb{N} \), and \( \sum_{k=1}^{n} \omega_k = N \), the total number of runs. For the GLMs considered in this paper, (3) depends on \( \beta \) through the matrix \( W \) and hence selection of a \( D \)-optimal design requires knowledge of the values of these parameters. Thus a locally optimal design is obtained.

The relative performance of two designs, \( \xi_1 \) and \( \xi_2 \), under \( D \)-optimality may be assessed using relative \( D \)-efficiency, defined as

\[
  \text{DEff}(\xi_1, \xi_2) = \exp \{ \phi_D(\xi_1) - \phi_D(\xi_2) \},
\]

where \( 0 \leq \text{DEff}(\xi_1, \xi_2) \). If \( \xi_2 \) is a \( D \)-optimal design that maximises (3), then (5) provides an absolute measure of the performance of design \( \xi_1 \).

In this paper, we address the screening problem of model selection and estimation of parameters in the selected model. We define, in Section 2, a Bayesian information capacity criterion that generalises \( D \)-optimality to provide model-robust designs for GLMs. We also present and discuss a model selection strategy that uses all-subsets regression and suitable penalties for model complexity. Sections 3 and 4 describe simulation studies of logistic and log-linear regression modelling, respectively, which demonstrate and assess the effectiveness of the methods. In Section 5, we present some avenues for future work to further develop methodology for screening experiments with non-normal data.
2. Information capacity designs and model selection

Consider a set $\mathcal{M}$ of $M = |\mathcal{M}|$ distinct candidate models, each of which have the same link function. The linear predictor for the $m$th model and $j$th run is given by

$$\eta_{ij}^m = \beta_{0m} + \sum_{i=1}^{q} \beta_{im} x_{ij} I(i, m),$$

where the $\beta_{im}$ are the values of the parameters in model $m$, $I(i, m) = 1$ if variable $i$ is in model $m$, and $I(i, m) = \beta_{im} = 0$ otherwise. Hence, the number of parameters in model $m$ is $p_m = 1 + \sum_{i=1}^{q} I(i, m)$.

2.1. Bayesian information capacity

Information capacity (IC) was introduced as a linear-model design selection criterion by Sun (1993). It has been further developed and applied by, for example, Wu (1993) (supersaturated designs), and Li and Nachtsheim (2000) (model-robust factorial designs). In essence, this criterion seeks a design whose projections onto subsets of the variables produce
sub-designs having good estimation properties for the corresponding submodels. This is achieved by selecting a design that maximises a weighted average of the $D$-criterion objective function for each submodel.

For GLMs, Woods (2010) employed the criterion of Woods et al. (2006) to find locally optimal information capacity designs for an example having five variables. Designs were found that maximised

$$\Psi(\xi) = \sum_{m=1}^{M} \frac{1}{p_m} \log \det \{X_m^T W_m X_m\},$$

where $X_m$ and $W_m$ are the respective model and weight matrices for the $m$th model in $\mathcal{M}$.

We define the Bayesian IC criterion which incorporates into (7) uncertainty in the parameter values assumed for each model. This criterion selects a design that maximises the objective function

$$\Phi(\xi) = \sum_{m=1}^{M} \frac{1}{p_m} \int_{\mathcal{B}_m} \log \det \{X_m^T W_m X_m\} \pi_m(\beta_m) \, d\beta_m,$$

where $\mathcal{B}_m \subset \mathbb{R}$ is the parameter space for model $m$, $\beta_m = (\beta_{0m}, \ldots, \beta_{qm})^T$, and $\pi_m(\beta_m)$ is the prior distribution for $\beta_m$. The choice of $\pi(\beta_m)$ and $\mathcal{B}_m$, and the evaluation of (8), are discussed in Section 3.1 for logistic regression. For log-linear
Fig. 3. Average power, type I error rate and FDR for logistic regression with $\kappa = 1$ and $N = 30, 50, 80, 100$ runs.

regression, we make use of results in the literature that enable analytical construction of minimally-supported $D$- and Bayesian $D$-optimal designs, see Section 4.1.

2.2. Model selection

A variety of model selection procedures exist for determining the most appropriate GLM from a set of models, including Bayesian (Chen et al., 2008) and shrinkage methods (Park and Hastie, 2007). To focus investigations on the impact of design selection, we restrict attention to all-subsets regression and use an information criterion to adjust for the bias inherent from in-sample estimation of the prediction error (see Burnham and Anderson, 2002, ch. 2). When maximum likelihood estimation is employed, we use the Akaike information criterion (AIC; Akaike, 1974) as the model selection criterion, and choose a model that minimises

$$AIC(m; \hat{\beta}) = -2l_m(\hat{\beta}; y) + 2p_m,$$

where $l_m(\cdot; \cdot)$ is the log-likelihood function (1) for model $m$.

When $\beta$ is estimated via penalised maximum likelihood (see (2)), AIC is no longer an appropriate criterion. This is because the effective number of parameters is reduced, equivalent to the inclusion of prior information (Gelman et al., 2014). The reduction depends on the number $(N)$ of runs, with a smaller number of effective parameters for smaller $N$. Hence when $N$ is small, use of AIC will over-penalise larger models. To avoid this problem, we use a generalised information criterion (GIC; Konishi and Kitagawa, 1996) that relaxes the assumptions of (i) estimation via maximum likelihood, and (ii) inclusion of the true model in $M$. Hence, we select the model that minimises

$$GIC(m; \tilde{\beta}) = -2l_m(\tilde{\beta}; y) + 2tr \left\{ J^{-1}(\tilde{\beta}) I(\tilde{\beta}) \right\},$$

where

$$J(\tilde{\beta}) = -\frac{1}{N} \frac{\partial^2 l_m(\beta; y)}{\partial \beta \partial \beta^T} \bigg|_{\tilde{\beta}}, \quad I(\tilde{\beta}) = \frac{1}{N} \frac{\partial^2 l_m(\beta; y)}{\partial \beta \partial \beta^T} \bigg|_{\tilde{\beta}},$$
with \( l_\beta^m(\cdot; \cdot) \) the penalised log-likelihood function (2) for model \( m \); see also Murata et al. (1994) and Zhang et al. (2010). The evaluation of \( J(\hat{\beta}) \) and \( I(\hat{\beta}) \) is straightforward for the GLMs and penalised likelihood estimation method used in this paper. The performance of the GIC is investigated in Section 3.2.

Following analysis of the data from an experiment, those variables found to be involved in the selected model are deemed to be active. In simulation studies to assess the performance of the model selection strategies, we use three summary measures: (i) \textit{power}: the proportion of truly active variables that are correctly identified as active by the model selection strategy; (ii) \textit{type I error rate}: the proportion of inactive variables (i.e. those not included in the true model) that are incorrectly identified as active by the model selection strategy; and (iii) \textit{false discovery rate} (FDR): the proportion of variables identified as active by the model selection strategy that are truly inactive (i.e. not in the true model).

3. Designs and model selection for binomial response and logistic regression

To investigate the performance of the methodology for logistic regression we study a five variable example, with linear predictors of the form (6). We assume that any subset of these variables may be the set of active variables. Therefore there are 31 possible models. The models are ordered lexicographically within each model size and assigned labels 1, . . . , 31. Models 1, . . . , 5 have linear predictors that contain a single variable, 1, . . . , 5, respectively; models 6, . . . , 15 have two variables, 1, 2; 1, 3; . . . , 4, 5. Similarly, models 16, . . . , 25 are three variable models, 26–30 are four-variable models and model 31 contains all five variables.

To find optimal designs and perform subsequent simulation studies, the model parameters \( \beta_{im} \) are assumed to have independent prior distributions of the form

\[
\beta_{im} \sim \begin{cases} 
\text{Uniform}(\kappa, 5) & \text{for } i = 1, 3, 4 \text{ and } l(i, m) = 1, \\
\text{Uniform}(-5, -\kappa) & \text{for } i = 2, 5 \text{ and } l(i, m) = 1,
\end{cases}
\]

where \( \kappa = 1, 2, 3 \) and we assume \( \beta_{0m} = 0 \) and \( \beta_{im} = 0 \) if \( l(i, m) = 0 \) for \( m = 1, \ldots, M \). The adoption of bounded uniform prior distributions prevents the occurrence of parameter vectors in the support of the prior for which no design has a non-singular information matrix (cf. Waite, 2015).
Fig. 5. Average power, type I error rate and FDR for logistic regression with $\kappa = 1, N = 30, 50, 80, 100$ runs and a locally $D$-optimal design for the maximal model.

3.1. Information capacity designs

We relax the assumption in (4) that $\omega_k$ is integer ($k = 1, \ldots, n$) and consider approximate designs (e.g. Atkinson et al., 2007, ch. 9). An approximate Bayesian IC design for logistic regression maximises

$$
\Phi^\dagger(\xi) = \sum_{m=1}^M \frac{1}{p_m} \int \log \det \left\{ \sum_{k=1}^n \tilde{\omega}_k \var(y_k)f_m(x_k)f_m(x_k)^T \right\} \pi_m(\beta_m) \, d\beta_m,
$$

where $0 < \tilde{\omega}_k = \omega_k/N \leq 1$ and $f_m(x_k)^T \beta_m$ is the linear predictor for the $m$th model. Clearly, an optimal choice of approximate Bayesian IC design can be made independently of the total experiment size $N$. Finding approximate designs also substantially reduces the computational burden of the design optimisation. We found designs using simulated annealing (Haines, 1987) where the integral in (12) was evaluated numerically as a summation across a quasi-Monte Carlo sample (Lemieux, 2009, ch. 5). The simulated annealing algorithm employed was a cyclic descent algorithm that proposed, evaluated and accepted moves for one coordinate of the design at a time; see Woods (2010). Such “coordinate exchange” algorithms are standard in the design of experiments, and solve difficult, high-dimensional, optimisation problems via a series of one-dimensional optimisations. Use of a stochastic optimisation algorithm such as annealing has the advantage of helping to escape local optima which we have found is a particular issue when each coordinate can take values in a continuous range.

Figs. 1 and 2 (left columns) summarise the $D$-efficiencies of the Bayesian IC designs for $n = 6$ and $n = 30$ support points, respectively. For each choice of prior distribution for $\kappa = 1, 2, 3$ from (11), the plots are obtained using (i) 500 random draws of $\beta_m$ values from distribution (11); (ii) the locally $D$-optimal design for each value of $\beta_m$, again found using simulated annealing; and (iii) calculation of efficiency (5) to compare the Bayesian IC design with the locally $D$-optimal design. In general, the efficiencies decrease with model size. For $n = 6$, the efficiencies are highly variable between models, even for models of the same size. It is not uncommon for a Bayesian design for logistic regression to require a large number of support points (see Chaloner and Larntz, 1989, and Woods and Lewis, 2011). This variability is also evident in the simulation results for model selection. Hence, in the next section, we present only assessments of the designs with $n = 30$ support points.
Fig. 6. Average power, type I error rate and FDR for logistic regression with $\kappa = 3$, $N = 30, 50, 80, 100$ runs and a locally $D$-optimal design for the maximal linear model.

For comparison, Figs. 1 and 2 (right columns) also present the $D$-efficiencies of locally $D$-optimal designs for the maximal model containing all five variables (model 31) and with each parameter set to its prior expectation. With the obvious, and expected, exception of model 31, the $D$-efficiencies are more variable and generally lower than those obtained from the Bayesian IC design.

3.2. Model selection results

To assess the performance of the designs for model selection, a simulation study was performed for each Bayesian IC design in which (i) each of the models, in turn, was used as the true model for the data generating process; (ii) 1000 data sets were generated independently by simulating values of $\beta_m$ from the prior distribution, followed by simulation of responses $y$ from a Binomial distribution; (iii) for each data set, each of the models in $\mathcal{M}$ was fitted using maximum penalised likelihood, and the model selected that minimises GIC (10); and (iv) power, type I error rate and FDR were calculated for each simulated data set. The optimal approximate designs were converted into exact designs via rounding $\omega_k$ to the nearest integer. Results are provided for $N = 30, 50, 80, 100$ in Figs. 3 and 4, for $\kappa = 1, 3$ respectively. The results for $\kappa = 2$ (not shown) are similar to those for $\kappa = 3$.

For all three prior distributions, high power is achieved for all experiment sizes: greater than 80% for $N = 30, 50$, and greater than 90% for $N = 80, 100$. Generally, there is a slight downward trend in power as the size of the true model increases. With the exception of the model including all five variables, a similar trend was observed for $D$-efficiency. The type I error rate is, unsurprisingly, an increasing function of the number of variables in the data-generating (true) model, as there are fewer inactive variables (smaller denominator) for larger true models. The maximum type I error rate of about 0.5 occurs for those true models involving four variables and corresponds to identifying, on average, less than one additional active variable. In contrast, FDR is a decreasing function of true model size, as again there are fewer inactive variables for larger true models. The maximum FDR of approximately 0.5 occurs when $N = 30$ for true models that contain a single variable, and corresponds to identifying one fewer additional active variable on average.

The results are fairly similar for the different prior distributions. The major difference is lower type I error rates when $\kappa = 3$, where there is a greater distinction between the sizes of the model coefficients for active and inactive variables. For
Fig. 7. Boxplots of the GIC penalty from (10) with $\kappa = 1$ and data generated from models 1 (variable 1 only) and 16 (variables 1 and 2 only).

comparison, we present results obtained from using the locally $D$-optimal design for model 31 (Fig. 5 for $\kappa = 1$ and Fig. 6 for $\kappa = 3$). In general, the average power is lower and average Type I error and false discovery rates higher for this design than for the Bayesian IC designs. For $\kappa = 3$, the locally optimal design has particularly poor performance for some models having three or four variables.

Another obvious comparator for the Bayesian IC designs are locally optimal IC designs, i.e. designs that maximise (7) with parameters $\beta_m$ set equal to the mean of prior distribution (11) ($m = 1, \ldots, M$). For this example, the Bayesian IC and locally optimal IC designs have almost identical $D$-efficiency distributions and model selection results, with the Bayesian IC designs displaying very slightly less variable $D$-efficiencies when there are $n = 6$ support points. These results (which are not shown) further illustrate the generally good performance of IC designs for model selection. We anticipate greater differences between Bayesian and locally optimal IC designs for examples where the prior distribution does not specify a known sign for each model parameter.

A key determinant of the model selection findings is the size of the GIC penalty term in (10). Our numerical studies have shown that this depends not only on the size, $p_m$, of the model but also on the estimated model parameters and the goodness of fit, with better-fitting models having a smaller penalty. Fig. 7 shows the distributions of the penalties obtained when model 1 (variable 1) and model 16 (variables 1 and 2) are true for $N = 30$, 100 and $\kappa = 1$. In general, the penalty is somewhat less than $2p_m$, although it increases with $N$ and hence does not penalise larger models to the same degree as AIC. Models that include the correct variables have smaller penalty than other models. Further research on the use of this penalty is needed.

4. Design and model selection for Poisson response and log-linear regression

To investigate the performance of the methodology for log-linear regression, simulation studies were performed for two examples. Both assume the log link, $g(\mu) = \eta$, and linear predictors of the form (6). In the first example, there are again $q = 5$ variables that may affect the response (31 possible models) and prior distribution (11) is assumed. In the second example, there are $q = 10$ variables but, in line with factor sparsity (Box and Meyer, 1986), we consider only linear predictors including at most three active variables (175 possible models). Prior distributions for $\beta_{im}$ are given by (11) for
Fig. 8. Boxplots of D-efficiencies for robust designs for log-linear regression: five variables (a) \( \kappa = 1 \), (b) \( \kappa = 2 \) and (c) \( \kappa = 3 \); 10 variables (d) \( \kappa = 1 \), (e) \( \kappa = 2 \) and (f) \( \kappa = 3 \). For the 10 variable case, only results for every sixth model are displayed.

Again, \( \beta_{0m} = 0 \) and for the remaining parameters by

\[
\beta_{im} \sim \begin{cases} 
    \text{Uniform}(\kappa, 5) & \text{for } i = 6, 8, 9 \text{ and } I(i, m) = 1, \\
    \text{Uniform}(-5, -\kappa) & \text{for } i = 7, 10 \text{ and } I(i, m) = 1.
\end{cases}
\]

Again, \( \beta_{0m} = 0 \), \( \beta_{im} = 0 \) if \( I(i, m) = 0 \) and \( \kappa = 1, 2, 3 \).

### 4.1. Minimally-supported designs

We restrict attention to designs that are minimally supported with respect to the maximal model, that is, where the number, \( n \), of distinct support points is \( q + 1 \). For this class of designs, Russell et al. (2009) and McGree and Eccleston (2012) presented analytical design construction methods. Atkinson and Woods (2015) showed that for these designs with \(-1 \leq x_{ij} \leq 1 \) and \( E(\beta_{im}) \geq 1 \), for any \( m = 1, \ldots, M \),

\[
\int_{B_M} \log \det \left\{ X_m^T W_m X_m \right\} \pi_m(\beta_m) \, d\beta_m = \log \det \left\{ X_M^T W_M^* X_M \right\},
\]

where \( W_m^* = \text{diag}(w_{jm}^*) \) with \( w_{jm}^* = f(x_j)^T \beta_m^* \) and \( \beta_m^* = E(\beta_m) \), the prior expectation of \( \beta_m \). Hence, numerical integration is no longer required for design evaluation. To exploit the available theory, we find designs that maximise

\[
\Phi_D(\xi) = \log \det \left\{ X_M^T W_M^* X_M \right\}.
\]
Fig. 9. Boxplots of $D$-efficiencies for fractional factorial designs for log-linear regression: five variables (a) $\kappa = 1$, (b) $\kappa = 2$ and (c) $\kappa = 3$; 10 variables (d) $\kappa = 1$, (e) $\kappa = 2$ and (f) $\kappa = 3$. For the 10 variable case, only results for every sixth model are displayed.

Maximisation of (13) defines a (pseudo-) Bayesian $D$-optimality criterion for the maximal model. A heuristic justification for using this criterion to find model-robust designs was given by McGree and Eccleston (2012) who pointed out that, assuming common prior distributions, the levels included for each variable in the minimally-supported Bayesian $D$-optimal design for each individual model $m$ are the same. Only the numbers of replications of each variable value differ between the designs. Hence the sub-designs defined as projections of the minimally-supported design for the maximal model into a subset of the variables will contain the same values of the variables as a minimally-supported optimal design for that subset of variables but with different replication. Typically, designs defined in this way display less balance in the variable levels than the $D$-optimal designs for the sub-models. The advantage of maximising (13) is that no numerical optimisation is required for design selection, and hence large examples (e.g. 10 variables) can be investigated.

We replicate minimally-supported Bayesian $D$-optimal designs that maximise (13) to obtain designs with $N = 16$ runs for five variables and $N = 32$ runs for 10 variables. Fig. 8 summarises, for each model, the $D$-efficiencies (5) for the five variable and 10 variable designs, calculated as described in Section 3.1 except that, in step (ii), the locally $D$-optimal designs obtained from the theorem of Russell et al. (2009) are used. In general, the efficiencies are somewhat higher than those achieved by the equivalent five variable designs for logistic regression.

There are three main points of interest: (1) the $D$-efficiencies are higher for the five variable design due to the smaller number of variables in the maximal model leading to less imbalance in the variable values in the sub-designs; (2) for both the five-variable and 10-variable designs, the $D$-efficiency increases with the size of the model, reflecting the construction method of maximising (13) for the maximal model; and (3) the spread of the $D$-efficiencies decreases as $\kappa$ increases, making
the prior distribution more concentrated. In both examples and for all $\kappa$ values, the minimum efficiency is greater than $\sim 0.4$, and the mean efficiency is greater than $\sim 0.55$. For smaller $\kappa$ and models with larger numbers of variables, the designs often have much higher $D$-efficiencies.

For this example, we also assess the performance of minimum aberration fractional factorial designs of resolution V (with $N = 16$ runs for five variables) and resolution IV (with $N = 32$ runs for 10 variables), see Fig. 9. Although these designs are $D$-optimal for the linear model, they perform uniformly poorly under log-linear regression, and much worse than the robust minimally-supported designs. Their efficiencies are particularly low for the larger values of $\kappa$, where the variance of the response is least constant.

4.2. Model selection results

Simulations to assess the performance of the designs for model selection were conducted as described in Section 3.2 except that, in step (iii), the model parameters were estimated using maximum likelihood and a model was chosen using AIC. Fig. 10 shows the results for five variable and 10 variable studies with $\kappa = 1$. Results in both cases are very encouraging, with almost uniformly high power and low type I error rates ($< 0.2$). For data-generating models with only one active variable (models 1–5 for the five variable experiment and models 1–10 for the 10 variable experiment), the truly active variable is occasionally missed, and another variable is identified as active. These errors lead to slightly lower power for these models, and non-zero FDR. For models with larger numbers of variables, all active variables are successfully identified (power equal to 1). For the five variable study, the FDR is consistently just below 0.3, corresponding to a maximum of about one non-active variable being incorrectly identified as active. For the 10 variable study, no screening errors are made for models containing three active variables (model 56 onwards). For both studies, the somewhat counter-intuitive result that performance improves for true models containing more active variables is explained by the construction method of the design (see Section 4.1), which focusses on the model containing the maximum number of variables.

For this example, the model selection performance of the two fractional factorial designs was very similar to that of the robust designs. Hence, the factorial designs would be effective for discrimination between the competing models but would provide poor estimation of the selected model.

5. Discussion and further research

This paper provides the first investigation of designs for screening variables under a generalised linear model. The results demonstrate that effective screening (high power with only moderate type I error rate and FDR) is achievable. For a binomial response and logistic regression, both design and model selection are more challenging than for a Poisson response, and larger designs are required to achieve good model selection results. For a binomial response, the results presented here can easily be extended to linear predictors that include products of variables representing interactions.

Future work is needed to investigate in more detail the use of the GIC penalty with maximum penalised likelihood. In some experiments, it may also be necessary to choose the link function in addition to the linear predictor. Compromise designs for this situation were found by Woods et al. (2006). In this paper, we have restricted the size of the model space under consideration by applying the principle of factor sparsity or by restricting the number of variables. For larger model spaces, the curse of dimensionality may prevent an all-subsets approach to model selection and alternative methods, such as sampling the model space (Smucker and Drew, 2015) or shrinkage regression (Friedman et al., 2010), would then need to be employed. Clearly, the choice of design, and any resultant “confounding” of model effects, will have an impact on any model selection procedure, including Bayesian and shrinkage methods. Investigations into the performance of these methods is another area for future work.
In the binomial and Poisson examples, we chose designs that ensured all models were estimable. This strategy is in contrast to the use of a design criterion tailored to model discrimination alone such as $T$-optimality (Atkinson and Fedorov, 1975), where the requirement of model estimability is often not met for nested models. An alternative approach is to generalise to multiple models those design selection methods that focus on both estimation and discrimination, such as the use of compound criteria (Atkinson, 2008) or hybrid designs (Waterhouse et al., 2008).

**Acknowledgement**

D. C. Woods was supported by Fellowship EP/J018317/1 from the UK Engineering and Physical Sciences Research Council. The authors thank an Associate Editor and two referees for comments that improved the paper.

**References**


Model robust designs for survival trials

Maria Konstantinou a, Stefanie Biedermann b,*, Alan Kimber b

a Ruhr-Universität Bochum, Fakultät für Mathematik, Germany
b University of Southampton, Statistical Sciences Research Institute, UK

ARTICLE INFO

Article history:
Received 28 February 2016
Received in revised form 25 July 2016
Accepted 14 October 2016
Available online 24 October 2016

Keywords:
Proportional hazards models
Minimax optimal designs
D-optimality
c-optimality
Type-I censoring

ABSTRACT

The exponential-based proportional hazards model is often assumed in time-to-event experiments but may only approximately hold. Deviations in different neighbourhoods of this model are considered that include other widely used parametric proportional hazards models and the data are assumed to be subject to censoring. Minimax designs are then found explicitly, based on criteria corresponding to classical c- and D-optimality. Analytical characterisations of optimal designs are provided which, unlike optimal designs for related problems in the literature, have finite support and thus avoid the issues of implementing a density-based design in practice. Finally, the proposed designs are compared with the balanced design that is traditionally used in practice, and recommendations for practitioners are given.

© 2016 Published by Elsevier B.V.

1. Introduction

Optimal experimental designs are often constructed assuming that the model generating the data is known, up to the values of the parameters involved. In many practical situations, however, the proposed parametric model may only be approximately true and thus may cause the vector of parameter estimators to be biased. As illustrated by Box and Draper (1959) for the case of a linear regression model, the advantages of using an optimal design that minimises just the variance are lost even if the deviations from the assumed model are small.

Following Box and Draper (1959), robust designs for approximately linear regression have been constructed by Wiens (1992) based on classical optimality criteria but involving the mean squared error matrix. He finds minimax designs which are optimal in that they minimise the criteria functions for the worst possible deviation from the linear regression model. Prediction and extrapolation problems with possible heteroscedasticity are studied by Wiens (1998) and Fang and Wiens (1999) respectively among others. Sinha and Wiens (2002) consider the construction of sequential designs which are robust against model uncertainty for nonlinear models. Further results on misspecified nonlinear regression include Woods et al. (2006), Wiens and Xu (2008) and Xu (2009a) for prediction and extrapolation problems.

However, none of these authors considers the case where the data are subject to censoring. This arises in many time-to-event experiments when a particular event of interest is not observed for some of the subjects utilised in the experiment. Censoring is often a result of the fact that the experiments are not run as long as necessary in order to obtain complete data, that is, event times for all the subjects, because of time and cost limitations. Therefore, it is of interest to find optimal designs which are robust to misspecifications of the assumed model and which allow for the possibility of the data being censored.
The available literature on model robust designs for time-to-event data is focused on accelerated life tests for which the subjects are put under extreme conditions in order for the event of interest to occur sooner than under normal circumstances. In this case, extrapolation to lower covariate values and prediction problems is often of interest; see, for example, Pascual and Montepiedra (2003), Xu (2009b) and McGree and Eccleston (2010).

An alternative class of models used for the modelling of time-to-event data is studied, namely that of proportional hazards models. Such models satisfy the proportional hazards assumption of constant hazard ratio over time and are frequently used in practice because of the simple interpretation of the regression coefficients in terms of hazard ratios. When a specific distribution is assumed for the event times, the resulting parametric models are referred to as distribution-based proportional hazards models. Cox's proportional hazards model, on the other hand, leaves the underlying distribution unspecified and therefore inference is based on the partial likelihood function (see Collett, 2003 for further details).

Konstantinou et al. (2015) consider Cox's model and show that in the presence of Type-I censoring an exponential distribution can be assumed without greatly affecting the optimal choice of design for partial likelihood estimation. They also find that the full and partial likelihood approaches result in very similar designs for the same assumed model.

Following these findings, small deviations in a neighbourhood of the exponential-based proportional hazards model are considered. The model uncertainty is formulated via a contamination function and the data are assumed to be subject to Type-I censoring. Then following along the same lines as in Xu (2009b), both censoring and model uncertainty are incorporated to obtain the asymptotic properties of the maximum likelihood estimator. Based on the asymptotic mean squared error matrix, minimax optimal designs for full likelihood estimation are constructed which protect against the worst possible misspecification of the assumed exponential model. Note that Xu (2009b) considers various prediction and extrapolation problems for normally distributed data and investigates the construction of designs that are continuous with respect to the Lebesgue measure. However, the focus of the present paper is on designs with finite support. This allows for explicit solutions to be obtained and then compared with the corresponding results of Konstantinou et al. (2014) for the case of the assumed model being true.

In Section 2 the assumed and true models considered are introduced and two different classes of contamination functions are defined to account for the various forms of the true distribution for the data. Then in Section 3 the asymptotic properties of the maximum likelihood estimator for the parameter vector are derived under model uncertainty and Type-I censoring. Analytical characterisations of minimax c- and D-optimal designs are given in Section 4. These designs are found using criteria corresponding to the classical c- and D-optimality criteria but are based on the mean squared error matrix rather than just the information matrix. In Section 5 the behaviour of the proposed designs is illustrated and they are compared with the balanced design traditionally used in practice. Finally, the main conclusions are discussed in Section 6.

2. Models and contamination functions

Time-to-event experiments are usually conducted in order to evaluate a particular intervention or treatment. Therefore, in what follows the focus is on models that involve one explanatory variable \( x \). The mean squared error matrix for general designs is derived, and then design search is illustrated for the situation in which \( x \) takes values in the binary design space denoted by \( \mathcal{X} = \{0, 1\} \), corresponding, for example, to a placebo and an active treatment in a clinical trial.

The aim of the experiment is assumed to be the estimation of one or both of the two model parameters. Let \( c \) be the predetermined duration of the experiment at which point the observations of subjects for which the event of interest has not occurred are said to be right-censored. Possibly censored data are summarised mainly using the hazard function which expresses the risk of the event of interest occurring at any time after the commencement of the experiment (Collett, 2003).

Consider the situation where the experimenter assumes the exponential-based proportional hazards model specified by the hazard function

\[
h_1(t) = \exp(\alpha + \beta x), \quad t > 0, \ x \in \mathcal{X} \subseteq \mathbb{R},
\]  

where \( \alpha \) and \( \beta \) are real parameters, when in fact this is only an approximation to the true underlying model. Denote the hazard function of the unknown true model by

\[
h_2(t) = \exp \left\{ \alpha + \beta x + \frac{g(t)}{\sqrt{n}} \right\}, \quad t > 0, \ x \in \mathcal{X} \subseteq \mathbb{R}, \ g(t) \in \mathcal{G},
\]

where \( n \) denotes the sample size. The function \( g(t) \) represents uncertainty about the exact form of the underlying distribution for the data and, following the literature, it is called the contamination function or just the contaminant. It is assumed that \( g(t) \) is unknown and ranges in a neighbourhood specified by the class \( \mathcal{G} \).

The parametrisation in (2.2) allows one to remain within a proportional hazards framework and ensures that the model parameters are well defined. In particular, unlike the existing literature, see, for example, Wiens (1992), the contamination function is independent of the covariate value \( x \). Therefore, the parameter \( \beta \) corresponds to the effect of the explanatory variable. For identifiability reasons it is further required that \( g(t) \) does not involve an additive constant. If this were not the case, the constant term would be absorbed in the quantity \( \exp(\alpha) \) that represents the baseline hazard for model (2.1), that is, the hazard function for a subject with \( x = 0 \).

The factor \( n^{-1/2} \) is included so that the deviations are of the order \( O(1/\sqrt{n}) \), resulting in models that are in a neighbourhood of the exponential model (2.1). At the same time, the dependence of \( g \) on the time \( t \) ensures that the general
form of the true model includes widely used parametric proportional hazards models, such as, for example, the Weibull and Gompertz distributions with known shape parameter \( \gamma \). These distributions correspond to the cases where \( g(t) \) is equal to \( (\gamma - 1) \log t \) and \( \gamma t \) respectively (Collett, 2003). Therefore, assuming that model (2.1) is correct, the corresponding log-likelihood function of the random variables \( c \) by zero, and so the censoring time \( c \) considered, Type-I censoring corresponds to the case where all the subjects enter the experiment at the same time, indicated to that used in Xu (2009b) is adopted in order to incorporate both censoring and model misspecification in the maximum likelihood estimation method and to obtain the asymptotic properties of the resulting estimator vector.

Now consider the case of unbounded contamination functions such as \( g(t) = (\gamma - 1) \log t \) for which \( \lim_{t \to 0} g(t) = -\infty \). A class that can be used to include such contaminants is

\[
g_2 = \left\{ g : \int_0^c e^{-te^{\alpha + \beta x}} g(t) dt \leq c_2, \forall x \in \mathcal{X} \text{ and } \int_0^c g(s) ds < \infty \right\},
\]

where, as before, \( c_2 \) is a specified positive constant. This class includes contamination functions \( g(t) \) which are bounded on the time interval \([0, c] \) and is also used in Li and Notz (1982). They, however, considered extrapolation and interpolation problems for linear regression models with complete data.

### 3. Estimation under model uncertainty and Type-I censoring

For the estimation of the model parameters the full likelihood approach is adopted since the assumed parametric model is completely specified as the exponential-based proportional hazards model. Throughout this section a similar procedure to that used in Xu (2009b) is adopted in order to incorporate both censoring and model misspecification in the maximum likelihood estimation method and to obtain the asymptotic properties of the resulting estimator vector.

Let \( T_1, \ldots, T_n \) be the independent random variables indicating the times to the occurrence of the event of interest for the \( n \) subjects utilised in the experiment with corresponding observed values \( t_1, \ldots, t_n \). Under the right-censored data scenario considered, Type-I censoring corresponds to the case where all the subjects enter the experiment at the same time, indicated by zero, and so the censoring time \( c \) is common for all the subjects. This situation occurs commonly in reliability applications. Alternatively, as is more common in clinical studies, subjects may enter the study at different calendar times but each be followed up for \( c \) time units. Therefore, in the presence of Type-I censoring, what is actually observed are the values \( y_j \) of the random variables \( Y_j = \min\{T_j, c\}, j = 1, \ldots, n \). This is formulated using an indicator variable \( \delta_j \) that is equal to unity if observation \( j \) is an event time and zero if it is a right-censored observation. That is,

\[
\delta_j = \begin{cases} 1, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c. \end{cases}
\]

Note that \( \delta_j \sim Bin(1, P_j) \), where \( P_j = P(\delta_j = 1) = P(Y_j = T_j) \).

The likelihood function for possibly censored data involves the survivor and probability density functions which are defined in terms of the hazard function as

\[
S_k(y) = \exp \left\{ -\int_0^y h_k(s) ds \right\} \quad \text{and} \quad f_k(y) = h_k(y)S_k(y), \quad y \in [0, c]; \quad k = 1, 2,
\]

respectively (Collett, 2003). Therefore, assuming that model (2.1) is correct, the corresponding log-likelihood function of the \( j \)th observation \( y_j \) with covariate value \( x_j \) is given by

\[
l = l(x_j, \alpha, \beta) = \delta_j \log f_1(y_j) + (1 - \delta_j) \log S_1(c) = \delta_j (\alpha + \beta x_j - ye^{\alpha + \beta x_j}) - (1 - \delta_j)ce^{\alpha + \beta x_j}.
\]

To find the limiting properties of the maximum likelihood estimator for the vector of model parameters requires the evaluation of the asymptotic expectation and variance–covariance matrix of the score function where

\[
\frac{\partial l}{\partial \alpha} = \delta_j \left( 1 - ye^{\alpha + \beta x_j} \right) - (1 - \delta_j)ce^{\alpha + \beta x_j}, \quad \frac{\partial l}{\partial \beta} = x_j \frac{\partial l(x_j, \alpha, \beta)}{\partial \alpha}.
\]

for the \( j \)th observation and also the calculation of the asymptotic information matrix involving the second order derivatives

\[
\frac{\partial^2 l}{\partial \alpha^2} = -e^{\alpha + \beta x_j} \left[ \delta_j y_j + c(1 - \delta_j) \right], \quad \frac{\partial^2 l}{\partial \alpha \partial \beta} = x_j^2 \frac{\partial^2 l}{\partial \alpha^2}, \quad \frac{\partial^2 l}{\partial \beta^2} = x_j^2 \frac{\partial^2 l}{\partial \alpha^2}.
\]
At this stage the fact that the true model is actually specified by (2.2) must be taken into account. Assuming this true model and observing that the above expressions involve only two random quantities via \( \delta_j \) and \( \delta_j Y_j \), gives

\[
E \left[ \frac{\partial l}{\partial \alpha} \right] = e^{\alpha + \beta x_i} \int_0^c e^{-y_j e^{\alpha + \beta x_i}} g(y_j) \frac{\sqrt{n}}{y_j} dy_j + o \left( \frac{1}{\sqrt{n}} \right),
\]

\[
\text{Var} \left( \frac{\partial l}{\partial \alpha} \right) = 1 - e^{-c e^{\alpha + \beta x_i}} + e^{\alpha + \beta x_i} e^{-c e^{\alpha + \beta x_i}} \int_0^c \frac{g(s)}{\sqrt{n}} ds
- (e^{\alpha + \beta x_i})^2 \int_0^c 2 \frac{y_j g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha + \beta x_i}} dy_j + o \left( \frac{1}{\sqrt{n}} \right),
\]

\[
E \left[ -\frac{\partial^2 l}{\partial \alpha^2} \right] = 1 - e^{-c e^{\alpha + \beta x_i}} + e^{\alpha + \beta x_i} e^{-c e^{\alpha + \beta x_i}} \int_0^c \frac{g(s)}{\sqrt{n}} ds
- e^{\alpha + \beta x_i} \int_0^c \frac{g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha + \beta x_i}} dy_j + o \left( \frac{1}{\sqrt{n}} \right),
\]

(3.1)

using Taylor expansions. The calculations for the derivation of the set of expressions (3.1) can be found in the Appendix.

Now let \( \theta = (\alpha, \beta)^T \) be the vector of model parameters and \( \theta_0 \) the vector of their true values. Also let

\[
\xi = \left\{ x_1, \omega_1, x_2, \omega_2, \ldots, x_m, \omega_m \right\}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^m \omega_i = 1,
\]

(3.2)

where \( x_1, \ldots, x_m \) are the distinct experimental points where observations are taken and \( \omega_1, \ldots, \omega_m \) represent the relative proportions of observations taken at the corresponding point \( x_i \). Using the expressions in (3.1) gives the asymptotic information matrix of \( \theta_0 \)

\[
M(\xi) = M(\xi, \theta_0) = \lim_{n \to \infty} \frac{1}{n} \left[ -\sum_{j=1}^n \frac{\partial^2 l}{\partial \theta \partial \theta^T} \right] = \sum_{i=1}^m \omega_i (1 - e^{-c e^{\alpha + \beta x_i}}) \left( \begin{array}{c} x_i x_i^2 \\ x_i \end{array} \right),
\]

the asymptotic expectation of the score function evaluated at \( \theta_0 \)

\[
\tilde{b}(\xi, g) = \tilde{b}(\xi, g, \theta_0) = \frac{1}{\sqrt{n}} \lim_{n \to \infty} \frac{1}{n} \left[ \sqrt{n} \sum_{j=1}^n \frac{\partial l}{\partial \theta} \right]_{\theta=\theta_0} := \frac{1}{\sqrt{n}} b(\xi, g),
\]

and finally the asymptotic variance–covariance matrix of the score function evaluated at \( \theta_0 \) which is given by

\[
C(\xi) = C(\xi, \theta_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^n \text{Cov} \left( \frac{\partial l}{\partial \theta} \right)_{\theta=\theta_0} = \sum_{i=1}^m \omega_i (1 - e^{-c e^{\alpha + \beta x_i}}) \left( \begin{array}{c} x_i x_i^2 \\ x_i \end{array} \right).
\]

Note that the asymptotic matrices \( M(\xi) \) and \( C(\xi) \) are identical. Now expanding the score function \( s(\theta) \) around \( \theta_0 \) gives

\[
s(\theta) = s(\theta_0) + s'(\theta_0)(\theta - \theta_0) + \cdots,
\]

and, since the maximum likelihood estimate \( \hat{\theta} \) is a root of the score function,

\[
0 \approx s(\theta_0) + s'(\theta_0)(\hat{\theta} - \theta_0)
\]

\[
(\hat{\theta} - \theta_0) \approx M^{-1}(\xi, \theta_0)s(\theta_0).
\]

Now \( \sqrt{n} s(\theta_0) \sim AN(b(\xi, g), C(\xi)) \) and therefore the asymptotic distribution of the maximum likelihood estimator is described by

\[
\sqrt{n}(\hat{\theta} - \theta_0) \sim AN \left( M^{-1}(\xi)b(\xi, g), M^{-1}(\xi) \right).
\]

(3.3)
4. Minimax optimal designs

The optimal planning of time-to-event experiments is concerned with finding the experimental points and the number of subjects that should be assigned to each point so that the parameters are estimated as precisely as possible. To illustrate the proposed methodology, the binary design space $X = \{0, 1\}$ is considered, corresponding, for example, to placebo and active treatment, respectively. Following Kieler (1974), the problem is formulated through an approximate design of the form (3.2) with support points 0 and 1 and corresponding weights $\omega$ and $1 - \omega$. In practice, if an approximate design is available and a total number of $n$ observations can be taken, the quantities $\omega n$ and $(1 - \omega) n$ are rounded to integers using an efficient rounding procedure in order for the design to be used (see Pukelsheim and Rieder, 1992).

As mentioned in the introduction and as can be seen in (3.3), fitting the exponential-based proportional hazards model given in (2.1) when in fact the true underlying model is specified by (2.2) adds a bias to the maximum likelihood estimator and an efficient rounding procedure in order for the design to be used (see Pukelsheim and Rieder, 1992).

Furthermore, the minimax approach is adopted in order to find designs that ensure precise parameter estimation for the worst possible contamination function in the class of contamination functions (either $g_1$ or $g_2$).

4.1. Optimality criteria

The optimality criteria used correspond to classical optimality criteria but they are based on the mean squared error matrix rather than just the information matrix. The resulting minimax optimal designs minimise the corresponding criteria with respect to the design, for the worst possible contamination function $g$.

The first criterion to be studied corresponds to the $c$-optimality criterion for estimating only the parameter $\beta$, treating $\alpha$ as a nuisance parameter. This is often the case in time-to-event experiments since $\beta$ represents the explanatory variable effect and is therefore of primary interest. A design $\xi^*$ is a minimax $c$-optimal design for estimating $\beta$ if (0 1)$^T$ is in the range of $\text{MSE}(\xi^*, g)$ and

$$
\xi^* = \arg \min_{\xi} \max_{g \in \{g_1, g_2\}} (0 1) \text{MSE}(\xi, g) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
$$

The case corresponding to $D$-optimality is also considered, that is, when one is interested in estimating both model parameters $\alpha$ and $\beta$. A design $\xi^*$ is minimax $D$-optimal if

$$
\xi^* = \arg \min_{\xi} \max_{g \in \{g_1, g_2\}} \det \{\text{MSE}(\xi, g)\}.
$$

Note that under both optimality criteria the resulting optimal designs depend on the parameter values and therefore, following Chernoff (1953), these are referred to as locally optimal designs. The corresponding locally optimal designs for the case of the exponential-based proportional hazards model being the true model are readily available in Konstantinou et al. (2014).

4.2. Minimax $c$-optimal designs for $\beta$

Design search is illustrated through the special case of a binary design space $X = \{0, 1\}$. Hence a candidate design for estimating the parameter $\beta$ can have either one or two support points. However, in the former case the mean squared error matrix cannot be defined since the information matrix $M(\xi^*)$ is singular. Therefore, the designs must be supported at both 0 and 1, and let $\omega$ and $1 - \omega$ be their corresponding weights. The following theorem gives the minimax $c$-optimal design for estimating $\beta$ for both the cases of $g \in \{g_1, g_2\}$ (see Appendix for a proof).

**Theorem 1.** Regardless of whether the contamination function belongs in $g_1$ or $g_2$, the minimax $c$-optimal design for estimating $\beta$ on $X = \{0, 1\}$ allocates a proportion of $\omega^*$ of observations at point $x = 0$, where

$$
\omega^* = \frac{\sqrt{1 - e^{-c\omega + \beta}}}{\sqrt{1 - e^{-c\omega}} + \sqrt{1 - e^{-c(1-\omega) + \beta}}.}
$$

The minimax $c$-optimal weight given in (4.4) is the same as the $c$-optimal weight for estimating $\beta$ when the exponential-based proportional hazards model is true (see Konstantinou et al., 2014). Therefore, the contamination function $g$ does not affect the minimax $c$-optimal design for $\beta$ and the exponential distribution can be assumed without loss of generality. This result is in line with the findings of Konstantinou et al. (2015) for partial likelihood estimation.
4.3. Minimax D-optimal designs

To allow estimation of both parameters a design must have at least two support points. For $\mathcal{X} = \{0, 1\}$ this means that both points 0 and 1 must be support points of the minimax D-optimal design. However, now the choice of contamination class and therefore the worst possible contaminant affects the optimal choice of design. Theorems 2 and 3 provide analytical characterisations of the minimax D-optimal designs when $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$ respectively and are proven in the Appendix.

**Theorem 2.** Let $g \in \mathcal{G}_1$. The minimax D-optimal design on $\mathcal{X} = \{0, 1\}$ allocates a proportion of $\omega^*$ observations at point $x = 0$, where

$$\omega^* = \sqrt{c_1^2 (1 - e^{c_1 \alpha}) + 1} \left[ \sqrt{c_1^2 (1 - e^{c_1 \alpha}) + 1} + \sqrt{c_1^2 (1 - e^{c_1 \beta}) + 1} \right] \over c_1^2 (e^{-c_1 \beta} - e^{-c_1 \alpha}).$$  \hspace{1cm} (4.5)

**Theorem 3.** Let $g \in \mathcal{G}_2$. The minimax D-optimal design on $\mathcal{X} = \{0, 1\}$ allocates a proportion of $\omega^*$ observations at point $x = 0$, where

$$\omega^* = \sqrt{c_2^2 (1 - e^{c_2 \alpha} + 1 \left[ \sqrt{c_2^2 (1 - e^{c_2 \alpha}) + 1} + \sqrt{c_2^2 (1 - e^{c_2 \beta}) + 1} \right] \over c_2^2 (e^{c_2 \beta} - e^{c_2 \alpha})}. \hspace{1cm} (4.6)$$

Note that the D-optimal design when model (2.1) is true allocates equal proportions of observations at points 0 and 1. Furthermore, it is easy to check that both minimax D-optimal weights have limiting values as $c_1$ or $c_2$, increases. These are

$$\lim_{c_1 \to \infty} \omega^* = \frac{\sqrt{1 - e^{c_1 \alpha} + 1}}{\sqrt{1 - e^{c_1 \alpha} + \sqrt{1 - e^{c_1 \beta}}}}, \text{ when } g \in \mathcal{G}_1,$$

and

$$\lim_{c_2 \to \infty} \omega^* = \frac{\sqrt{1 - e^{c_2 \alpha} + 1}}{\sqrt{1 - e^{c_2 \alpha} + \sqrt{1 - e^{c_2 \beta}}}}, \text{ when } g \in \mathcal{G}_2.$$  \hspace{1cm} (5.1)

Also note that the minimax D-optimal weight for $g \in \mathcal{G}_1$ given in (4.5) tends to the $c$-optimal weight for $\beta$ when the exponential-based proportional hazards model is true.

5. Numerical results

For time-to-event experiments comparing two treatments, or equivalently a placebo with an active treatment, practitioners traditionally use the balanced design allocating equal proportions of observations at the two treatments. The aim is to illustrate the theoretical results on minimax optimal designs found in the previous section and also to examine the efficiency of the balanced design in the presence of model uncertainty and possibly censored data.

5.1. Minimax c-optimal designs for $\beta$

As shown in Section 4.2, the minimax c-optimal design for estimating $\beta$ does not depend on the contamination function $g$ but is locally optimal through the parameter values (see Theorem 1). To illustrate this parameter dependence $\beta$-values corresponding to small, moderate and large covariate effects are used along with various proportions of censored observations. Following Kalish and Harrington (1988), the proportion of censoring is characterised as the overall probability of censoring for model (2.1) had a balanced design been used. That is,

$$\text{proportion of censoring} = 1 - 0.5 \left( 1 - e^{-c_1 \alpha} \right) - 0.5 \left( 1 - e^{-c_1 \beta} \right). \hspace{1cm} (5.1)$$

Setting $\alpha = 0$, for illustration purposes, this equation provides the value of the censoring time $c$ for a given combination of $\beta$-value and censoring proportion.

Two different contamination scenarios are considered. For $g \in \mathcal{G}_1$, the Gompertz distribution is selected for which $g(t) = \gamma t$, where $\gamma$ is the shape parameter. A value of $\gamma = 0$ would correspond to the exponential regression model. For the class $\mathcal{G}_2$ of possibly unbounded contamination functions the Weibull distribution with shape parameter $\gamma$ is studied for which $g(t) = (\gamma - 1) \ln t$, so a value of $\gamma = 1$ would correspond to the exponential distribution.

The case of $\gamma = 1$ for the Gompertz model and $\gamma = 2$ for the Weibull model is presented. For both contamination types, various different values for $\gamma$ gave similar results, and are thus omitted. Table 1 shows the minimax c-optimal design...
weights at point \( x = 0 \) for several combinations of \( \beta \)-values and proportions of censoring. For each combination, the value of \( c \) is determined using Eq. (5.1) by setting \( \alpha = 0 \). Note that the minimax c-optimal weights are the same regardless of whether the Gompertz or the Weibull distribution is used.

It can be observed that the minimax c-optimal design for \( \beta \) allocates more observations at point \( x = 0 \) when \( \beta > 0 \), that is, when the probability of occurrence of the event increases with \( x \), and less when \( \beta < 0 \). Therefore, the design puts more weight at the experimental point where censoring is more likely.

The efficiencies of the balanced design for the Gompertz and the Weibull models under consideration are given in Table 2 for the various explanatory effects and proportions of censoring scenarios.

It turns out that the balanced design is highly efficient for small proportions of censoring whereas its efficiency drops below 90% for absolute \( \beta \)-values of 2.3 or more and proportion of censoring of 50% or more. Furthermore, the efficiencies are almost identical for both contamination functions. This can be explained by the form of the objective function defined in (4.2) which is given by

\[
\frac{e^{\beta/\alpha} \left( \int_0^c e^{-y e^{\beta+\alpha} g(y_j) dy_j} - \int_0^c e^{-y e^{\beta} g(y_j) dy_j} \right)^2 + \frac{1}{\omega(1 - e^{-c e^{\alpha}}) + (1 - \omega)(1 - e^{-c e^{\beta+\alpha}})}}{1 - e^{-c e^{\alpha}}} \]

for a design of the form \( \hat{\xi} = \{0, 1; \omega, 1 - \omega\} \). The above expression shows that the objective function is dominated by the terms involving \( \omega \) but not \( g \).

### 5.2 Minimax D-optimal designs

Besides the parameter dependence, the minimax D-optimal designs also depend on the choice of contamination class and therefore on the values of the positive constants \( c_1 \) or \( c_2 \) (see Theorems 2 and 3). In order to illustrate the contaminant dependence a numerical example is used that is based on the study reported by Freireich et al. (1963), for which the maximum likelihood estimates are \( \hat{\alpha} = -2.163 \) and \( \hat{\beta} = -1.526 \) with approximately 30% of the observations right-censored. Using this proportion of censoring and the estimates \( \hat{\alpha} \) and \( \hat{\beta} \) for the \( \alpha \) and \( \beta \) values, the value \( c = 30 \) is obtained from the characterisation of the proportion of censoring defined in Section (5.1). Figs. 1 and 2 illustrate the limiting behaviour of the minimax D-optimal weights \( \omega^* \) on \( x = 0 \) given in (4.5) and (4.6) respectively as \( c_1 \) and \( c_2 \) increase.

For both cases of \( g_1 \) and \( g_2 \), the weight at point \( x = 0 \) is smaller than 0.5 (the D-optimal weight when model (2.1) is true) and its value decreases as \( c_1 \) and \( c_2 \) increase, with the limiting weight for \( g \in g_1 \) being larger than that for \( g \in g_2 \).

To investigate the performance of the D-optimal minimax design with respect to contamination functions \( g \in g_1 \) and \( g_2 \), the Gompertz and the Weibull distributions respectively are considered again. In this situation the results turn out to be

### Table 1

Minimax c-optimal weights \( \omega^* \) at point \( x = 0 \).

<table>
<thead>
<tr>
<th>prop. of cens.</th>
<th>( e^\beta/\alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3.51)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.47</td>
</tr>
<tr>
<td>0.3</td>
<td>0.39</td>
</tr>
<tr>
<td>0.7</td>
<td>0.25</td>
</tr>
<tr>
<td>0.9</td>
<td>0.15</td>
</tr>
</tbody>
</table>

### Table 2

Efficiency, in percent, of the balanced design, for the Gompertz model with \( \gamma = 1 \) and the Weibull model with \( \gamma = 2 \) (in brackets).

<table>
<thead>
<tr>
<th>prop. of cens.</th>
<th>( e^\beta/\alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3.51)</td>
</tr>
<tr>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>0.3</td>
<td>99.4</td>
</tr>
<tr>
<td>0.5</td>
<td>76.5</td>
</tr>
<tr>
<td>0.7</td>
<td>70.1</td>
</tr>
<tr>
<td>0.9</td>
<td>67.6</td>
</tr>
</tbody>
</table>

(99.9) (99.8) (99.8) (99.9) (99.9) (99.8) (99.8) (99.9)
(97.2) (96.1) (97.5) (99.2) (99.2) (97.5) (96.1) (97.2)
(76.7) (87.3) (94.7) (98.6) (98.6) (94.7) (87.3) (76.7)
(70.1) (82.6) (92.5) (98.0) (98.0) (92.5) (82.6) (70.1)
(67.6) (79.8) (90.8) (97.4) (97.4) (90.8) (79.8) (67.7)
less clear cut than for $c$-optimality. In particular, the relative efficiencies of the balanced design are close to 1 in all scenarios, and sometimes even exceed 1. It is not surprising that an optimal minimax design can be less efficient than a non-optimal design in some scenarios, since the minimax designs protect against a whole class of contamination functions whereas each scenario is characterised by just one function from this class. This phenomenon is similar in nature to situations where parameter robust designs over a range of values are outperformed for specific values in this range.

As proposed by a referee, the case of the continuous design space $X = [0, 1]$ is also investigated, for the same combinations of $\beta$-values and censoring proportions as in Table 1. Using either the $c$- or $D$-optimality criterion, the support points of the resulting minimax optimal designs for either contamination class are found to always be the points 0 and 1. Furthermore, the corresponding minimax optimal weights are the same as in the case of the binary design space $X = \{0, 1\}$ described above.

Overall, the conclusion is that if estimation of both parameters, $\alpha$ and $\beta$, is of interest, the balanced design is highly efficient, and can be used in practice. If, however, the main focus is on the treatment effect $\beta$, then the minimax $c$-optimal designs are recommended.

6. Conclusions

In practice when parametric models are used for time-to-event experiments, often the exponential distribution is naturally assumed for the event times along with the proportional hazards assumption. However, this assumed model may only be an approximation of the true underlying parametric proportional hazards model.

Following this practical scenario, deviations in a neighbourhood of the exponential-based proportional hazards model are considered which are specified by a contamination function $g$. Two different classes of contamination functions are defined which can be used to include various forms of $g$ but most importantly they include the commonly used parametric proportional hazards models based on the Weibull and Gompertz distributions.

Since time-to-event experiments are usually conducted in order to evaluate a particular intervention or treatment, the focus is on models involving one explanatory variable. Nevertheless, the model misspecification introduced in Section 2 and the results of Section 3 regarding the maximum likelihood estimator can be easily generalised to models with more than one explanatory variables.

Assuming that the time-to-event data are subject to Type-I censoring, the construction of designs which are robust to model misspecifications is investigated. Following Wiens (1992), optimality criteria corresponding to the classical $c$- and
Fig. 2. Minimax $D$-optimal weight at point 0 for $g \in G_2$.

*D*-optimality criteria but based on the mean squared error matrix are used and minimax optimal designs are constructed which guard against the worst possible deviation from the assumed model. Therefore, previous results on minimax optimal designs are extended by considering both possibly censored data as well as the class of proportional hazards models. However, the choice of contamination classes enables the use of designs with finite support and therefore analytical characterisations of minimax $c$- and $D$-optimal designs are provided.

This framework is established for general designs, hence optimal designs on continuous design spaces $\mathcal{X}$ corresponding to, for example, doses of a drug can easily be found numerically. The design search is illustrated for the important special case of a binary design space, and some analytical results have been presented.

The results on minimax $c$-optimal designs for estimating the covariate coefficient $\beta$, show that the deviations from the exponential distribution do not affect the optimal choice of design if one remains in a proportional hazards framework. This is in accordance with the result for partial likelihood estimation, stating that under Type-I censoring the exponential distribution can be assumed for design search without loss of generality (see Konstantinou et al., 2015).

If estimation of both parameters is required, that is, if $D$-optimality is the desired criterion, then Theorems 2 and 3 give the minimax optimal weights for deviations in the classes $G_1$ and $G_2$ respectively. Both of these weights have limiting values if we allow the deviations to become large and in particular when $g \in G_1$ the minimax $D$-optimal weight tends to the $c$-optimal weight corresponding to the case of the assumed model being true, as $c_1 \to \infty$. This again highlights the importance of the latter design in a model uncertainty situation.

The analytical characterisations of minimax optimal designs and the numerical results of Section 5 suggest that if the main interest is in estimating the treatment effect one has to move away from the traditional balanced design to guard against misspecifications of the assumed exponential model. A suitable candidate for practical use would appear to be the classical $c$-optimal design for estimating the covariate effect assuming the exponential-based proportional hazards model. It is minimax $c$-optimal for both contamination classes, is (in the limit) minimax $D$-optimal for $G_1$, and is also highly efficient if Cox’s proportional hazard model is fitted via partial likelihood estimation (see Konstantinou et al., 2015). An analytical characterisation of the locally $c$-optimal design for estimating $\beta$ is given in Konstantinou et al. (2014).

The designs derived are locally optimal hence, while being robust against model misspecifications, they depend on the values of the unknown model parameters. Finding designs which are robust to both sources of uncertainty is an interesting area of future research. A promising starting point for such an investigation could be the parameter robust designs derived in Konstantinou et al. (2014). They show that for a binary design space and a given range of possible $\beta$-values, the optimal weight at point 0 of the standardised maximin $c$-optimal design for $\beta$ (for parameter robustness) is the average of the two
locally $c$-optimal weights at point 0 corresponding to the end-points of the given interval of $\beta$-values. As it is shown in Section 4, the (locally) minimax $c$- and $D$-optimal weights (for model robustness) are the same or tend, respectively, to the locally $c$-optimal weight. Therefore, it is of interest to investigate whether such an averaging of the locally optimal weights result also holds both for minimax $c$- and $D$-optimal designs for model robustness.

Acknowledgement

The work of M. Konstantinou has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilproject C2) of the German Research Foundation (DFG).

Appendix

A.1. Proof of set of expressions in (3.1)

The true underlying model specified by (2.2) has corresponding probability density function given by

$$f_z(y_i) = \exp \left\{ \alpha + \beta x_j + \frac{g(y_j)}{\sqrt{n}} \right\} \exp \left\{ -e^{\alpha + \beta x_j} \int_0^T e^{s(s)/\sqrt{n}} \, ds \right\}, \quad j = 1, \ldots, n.$$  

Taking this into account gives

$$E(\delta_j) = P_j = P(Y_j = T_j) = \int_0^C f_z(y_j) \, dy_j = 1 - \exp \left\{ -e^{\alpha + \beta x_j} \int_0^C e^{s(s)/\sqrt{n}} \, ds \right\}.$$  

Since small deviations from the exponential-based proportional hazards model are considered, the Taylor expansion of $e^{g(s)/\sqrt{n}}$ around $g(s) = 0$ can be used. Then the above expression becomes

$$E(\delta_j) = 1 - \exp \left\{ -e^{\alpha + \beta x_j} \left[ 1 + \frac{g(s)}{\sqrt{n}} + o \left( \frac{g(s)}{\sqrt{n}} \right) \right] \right\}$$  

$$= 1 - \exp \left\{ -ce^{\alpha + \beta x_j} \left[ \int_0^C \frac{g(s)}{\sqrt{n}} \, ds + o \left( \frac{1}{\sqrt{n}} \right) \right] \right\}.$$  

By further expanding around $\int_0^C \frac{g(s)}{\sqrt{n}} \, ds + o \left( \frac{1}{\sqrt{n}} \right) = 0$, the expectation of the random variable $\delta_j$ becomes

$$E(\delta_j) = 1 - e^{-ce^{\alpha + \beta x_j}} + e^{\alpha + \beta x_j} e^{-\alpha x_j} \int_0^C \frac{g(s)}{\sqrt{n}} \, ds + o \left( \frac{1}{\sqrt{n}} \right),$$  

where the first term, $- e^{-ce^{\alpha + \beta x_j}}$, corresponds to the expectation if the assumed exponential model was in fact the true model. Using this expression the variance of $\delta_j$ can be found without making any further calculations and is given by

$$\text{Var}(\delta_j) = P_j(1 - P_j) = e^{-ce^{\alpha + \beta x_j}} (1 - e^{-\alpha x_j}) + e^{\alpha + \beta x_j} e^{-\alpha x_j} (2e^{-\alpha x_j} - 1) \int_0^C \frac{g(s)}{\sqrt{n}} \, ds + o \left( \frac{1}{\sqrt{n}} \right).$$

Note that

$$\delta_j Y_j = \begin{cases} Y_j, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c. \end{cases}$$  

Following along the same lines as for the random quantity $\delta_j$, that is, using two consecutive Taylor expansions, the following expressions can be obtained

$$E(\delta_j Y_j) = \frac{1 - e^{-ce^{\alpha + \beta x_j}}}{e^{\alpha + \beta x_j}} - ce^{-ce^{\alpha + \beta x_j}} + e^{-ce^{\alpha + \beta x_j}} (ce^{\alpha + \beta x_j} + 1) \int_0^C \frac{g(s)}{\sqrt{n}} \, ds$$  

$$- \int_0^C \frac{g(y_j)}{\sqrt{n}} e^{-\alpha x_j} \, dy_j + o \left( \frac{1}{\sqrt{n}} \right),$$  

$$\text{Var}(\delta_j Y_j) = -c^2 e^{-ce^{\alpha + \beta x_j}} (1 + e^{-\alpha x_j}) + \frac{1 - e^{-2ce^{\alpha + \beta x_j}}}{(e^{\alpha + \beta x_j})^2} - \frac{2ce^{-2ce^{\alpha + \beta x_j}}}{e^{\alpha + \beta x_j}} + e^{-ce^{\alpha + \beta x_j}}$$  

$$\times \left( c^2 e^{\alpha + \beta x_j} + 4ce^{-ce^{\alpha + \beta x_j}} + 2e^{-ce^{\alpha + \beta x_j}} e^{-\alpha x_j} + 2e^{-ce^{\alpha + \beta x_j}} \int_0^C \frac{g(s)}{\sqrt{n}} \, ds \right)$$  

$$- \int_0^C 2e^{-\alpha x_j} e^{\alpha + \beta x_j} \left( y_j + \frac{e^{-ce^{\alpha + \beta x_j}}}{e^{\alpha + \beta x_j}} + ce^{-ce^{\alpha + \beta x_j}} \right) \frac{g(y_j)}{\sqrt{n}} \, dy_j + o \left( \frac{1}{\sqrt{n}} \right).$$
Taking the first order derivative of this expression with respect to \( \omega \):

\[
\begin{align*}
\text{Cov}(\delta_j, \delta_j Y_j) &= e^{-ce^{x+\beta x}} \frac{1 - e^{-ce^{x+\beta x}}}{c^{x+\beta x} - ce^{-2c^{x+\beta x}}} \\
&\quad + e^{-ce^{x+\beta x}} \left( 2c^{x+\beta x} e^{-ce^{x+\beta x}} + 2e^{-c^{x+\beta x}} - 1 \right) \int_0^c \frac{g(s)}{\sqrt{n}} ds \\
&\quad - e^{-ce^{x+\beta x}} \int_0^c \frac{g'(y)}{\sqrt{n}} e^{-y^{x+\beta x}} dy + o \left( \frac{1}{\sqrt{n}} \right). 
\end{align*}
\]

The set of expressions given in (3.1) then follows since

\[
E \left( \frac{\partial l}{\partial \alpha} \right) = E(\delta_j) - e^{x+\beta x} E(\delta_j Y_j) - ce^{x+\beta x} E(1 - \delta_j),
\]

\[
\text{Var} \left( \frac{\partial l}{\partial \alpha} \right) = \text{Var}(\delta_j) + e^{2(x+\beta x)} \text{Var}(\delta_j Y_j) + c^2 e^{2(x+\beta x)} \text{Var}(\delta_j) - 2e^{x+\beta x} (1 + ce^{x+\beta x}) \text{Cov}(\delta_j, \delta_j Y_j),
\]

\[
E \left( -\frac{\partial^2 l}{\partial \alpha^2} \right) = -e^x + \beta x \left[ E(\delta_j Y_j) + cE(1 - \delta_j) \right].
\]

A.2. **Proof of Theorem 1**

Let \( \xi = \{0, 1, \omega, 1 - \omega\} \). The objective function defined in (4.2) becomes

\[
\begin{align*}
\left[ e^{x+\beta y} \left( 1 - e^{-c^{x+\beta x}} \right) \int_0^c e^{-y^{x+\beta x}} g(y) dy_j - \frac{e^x}{1 - e^{-c^{x}} \left( 1 - e^{-x^{x+\beta x}} \right)} \int_0^c e^{-y^{x+\beta x}} g(y) dy_j \right]^2 \\
+ \frac{1}{\omega (1 - e^{-c^{x}})} + \frac{1}{(1 - \omega)(1 - e^{-x^{x+\beta x}})}. 
\end{align*}
\]

The minimax \( c \)-optimal design for \( \beta \) is found by minimising the above expression with respect to \( \omega \) for the worst possible contaminant. Observe that the term involving the contamination function \( g \) is independent of the weight \( \omega \) and therefore, it is enough to minimise

\[
\frac{1}{\omega (1 - e^{-c^{x}})} + \frac{1}{(1 - \omega)(1 - e^{-x^{x+\beta x}})},
\]

which gives the optimal weight

\[
\omega^* = \frac{\sqrt{1 - e^{-c^{x+\beta x}}}}{\sqrt{1 - e^{-c^{x}}} + \sqrt{1 - e^{-x^{x+\beta x}}}}.
\]

A.3. **Proof of Theorem 2**

For \( \xi = \{0, 1, \omega, 1 - \omega\} \) the determinant of the mean squared error matrix is given by

\[
\begin{align*}
\frac{1}{\omega (1 - e^{-c^{x}})} (1 - e^{-c^{x}}) (1 - e^{-x^{x+\beta x}}) \left[ 1 + \omega \left( \int_0^c e^{-y^{x+\beta x}} g(y) dy_j \right)^2 \right] \\
+ (1 - \omega) \left( \int_0^c e^{-y^{x+\beta x}} g(y) dy_j \right)^2.
\end{align*}
\]

Since \( g \in \mathcal{G}_1 \), then \( \max_{y \in [0, c]} |g(y)| \leq c_1 \forall j = 1, \ldots, n \) and so

\[
\int_0^c e^{-y^{x+\beta x}} g(y) dy_j \leq c_1 (1 - e^{-c^{x+\beta x}}) / e^{x+\beta x}, \quad \forall x \in [0, 1].
\]

Therefore, for contamination functions \( g \) in the class \( \mathcal{G}_1 \) the maximum value of the determinant of the mean squared error matrix is given by

\[
\frac{c_1^2}{\omega (1 - e^{-c^{x}})} + \frac{c_1^2}{(1 - \omega)(1 - e^{-x^{x+\beta x}})} + \frac{1}{\omega (1 - e^{-c^{x}}) (1 - e^{-x^{x+\beta x}})}.
\]

Taking the first order derivative of this expression with respect to \( \omega \) and equating it to zero gives

\[
c_1^2 \omega^2 (1 - e^{-c^{x}}) - c_1^2 (1 - \omega)^2 (1 - e^{-x^{x+\beta x}}) - (1 - 2\omega) = 0
\]

\[
\iff \omega_{1,2} = -\frac{-c_1^2 (1 - e^{-c^{x+\beta x}}) + 1}{c_1^2 (1 - e^{-c^{x}}) + 1 \sqrt{c_1^2 (1 - e^{-x^{x+\beta x}}) + 1}}.
\]
When \( \beta \) is positive, it is easy to see that both the numerator and the denominator of the above expression are non-positive. The negative root of the numerator is rejected since
\[
- c_1^2 (1 - e^{-c e^{\alpha + \beta} + \beta}) - 1 - \sqrt{c_1^2 (1 - e^{-c e^{\alpha + \beta}} + 1) c_1^2 (1 - e^{-c e^{\alpha + \beta} + \beta}) + 1} < - c_1^2 (1 - e^{-c e^{\alpha + \beta}} + \beta) c_1^2 (1 - e^{-c e^{\alpha + \beta}}) = c_1^2 (e^{-c e^{\alpha + \beta}} - e^{-c e^{\alpha}}),
\]
and the weight must be always less than or equal to unity. In the case of negative \( \beta \)-values the denominator is positive and since \( \omega > 0 \), again the positive root is accepted.

Therefore, whatever the sign of the parameter \( \beta \), the minimax \( D \)-optimal weight at point 0 is given by (4.5).

A.4. Proof of Theorem 3

Since \( g \in g_2 \), then
\[
\int_0^N e^{-y e^{\alpha + \beta} + \beta} g(y) dy \leq c_2 \quad \forall x \in \{0, 1\}. \]
Therefore, for a fixed design \( \xi \) supported at 0 and 1 with corresponding weights \( \omega \) and \( 1 - \omega \) the determinant of the mean squared error matrix is smaller than or equal to
\[
\frac{(c_2 e^{\alpha + \beta})^2}{(1 - e^{-c e^{\alpha}})} - (1 - \omega)^2 - (1 - 2 \omega) \omega = 0
\]
Taking the first order derivative of this expression with respect to \( \omega \) and equating it to zero gives
\[
\omega_{1,2} = \frac{\beta e^{\alpha + \beta}}{(1 - e^{-c e^{\alpha}}) - (1 - \omega)(1 - e^{-c e^{\alpha}})}
\]
When \( \beta \) is positive, it is easy to check that both the numerator and the denominator of the above expression are non-positive. The negative root of the numerator is rejected since
\[
- \frac{(c_2 e^{\alpha + \beta})^2}{(1 - e^{-c e^{\alpha}})} - 1 - \sqrt{\frac{(c_2 e^{\alpha + \beta})^2}{(1 - e^{-c e^{\alpha}})} + 1} < \frac{(c_2 e^{\alpha + \beta})^2}{(1 - e^{-c e^{\alpha}})} + 1
\]
and the weight must be always less than or equal to unity. In the case of negative \( \beta \)-values the denominator is positive and since \( \omega > 0 \), again the positive root is accepted.

Therefore, whatever the sign of the parameter \( \beta \), the minimax \( D \)-optimal weight at point 0 is given by (4.6).

References

Robustness of classical and optimal designs to missing observations

Byran J. Smucker\textsuperscript{a,}\textsuperscript{*}, Willis Jensen\textsuperscript{b}, Zichen Wu\textsuperscript{a}, Bo Wang\textsuperscript{a}

\textsuperscript{a}Department of Statistics, Miami University, Oxford, OH, United States
\textsuperscript{b}W.L. Gore & Associates, Flagstaff, AZ, United States

\textbf{Abstract}

Missing observations are not uncommon in real-world experiments. Consequently, the robustness of an experimental design to one or more missing runs is an important characteristic of the design. Results of an evaluation of the robustness of classical and optimal designs to missing observations are presented, and optimal designs fare relatively well in terms of robustness compared to classical designs. Additionally, a modified version of an existing robustness criterion is used to construct designs that are robust to missing observations.

1. Introduction

When designing experiments, there is a tension between robustness and optimality. Optimality suggests a single, best design while robustness implies quality under a broad range of conditions. Robustness is crucial in the real world of experimentation, as noted generally by Wendelberger (2010). There is also a growing acknowledgment that multiple criteria should be used when constructing optimal designs (e.g. Lu et al., 2011; Gilmour and Trinca, 2012). In the face of many different kinds of variability and uncertainty, it is important for designs that are optimal for estimation or prediction to be robust in other ways. Since there is generally not a single design that is optimal in all aspects, an acceptable compromise is to consider appropriate tradeoffs between key characteristics. Though Box and Draper (1975) gave a list of fourteen attributes that generally relate to a design’s robustness and practicality, one aspect not specifically mentioned is robustness to missing values. Box and Draper do mention outliers or “wild observations”, and to the extent that these observations can be omitted from the analysis in a principled way they would fall under the purview of our work. However, we discuss in particular the situation where no value is obtained for a particular run. The primary goal of our work is to consider some common classical and optimal designs in terms of their robustness to missing observations. Secondarily, we adapt an existing criterion to construct some new designs that are missing-robust.

Several robust-to-missing-observations criteria have been proposed in the literature. Box and Draper (1975) pointed out the connection between the diagonal values of the hat matrix and a design’s robustness and applied their related criterion to central composite designs. Siddiqi (2011) applied a similar criterion to the smaller response surface designs of Draper and Lin (1990). Herzberg and Andrews (1976) considered the probability that a design will not estimate the desired model,
and Andrews and Herzberg (1979) suggested maximizing the expected value of the determinant of the information matrix under possible missing observations. Akhtar and Prescott (1986) developed a criterion that minimizes the maximum loss due to missing observations and applied it to the evaluation and generation of central composite designs, and Ahmad and Gilmour (2010) used this measure to study the robustness of so-called subset designs (Gilmour, 2006). Herzberg et al. (1987) proposed equi-information designs, which retain equal information when up to two design points are missing. Imhof et al. (2002) presented results based on a different maximin criterion assuming continuous designs.

Missing from the literature is (1) a systematic comparison of classical and optimal designs in terms of their robustness to missing observations; and (2) a general algorithmic approach to generating designs that are robust to missing observations. We note that Hackl (1995) did generate robust designs for small response surface experiments, though they have not been formally published. Furthermore, a referee pointed us to a recent paper (da Silva et al., 2016) that incorporates a leverage-based missing-robustness criterion used within a larger compromise criterion that balances pure error and lack-of-fit degrees of freedom (see Gilmour and Trinca, 2012).

The rest of the paper unfolds as follows. We first motivate our work by explaining why missing runs occur in practical situations. We then outline several evaluative measures of missing-robustness. Next we demonstrate the impact of missing runs on some standard designs, including fractional factorial and D-optimal designs for screening experiments, and central composite and I-optimal designs for prediction/optimization. Included in these comparisons are some new designs constructed using a version of a robustness criterion from the literature. We conclude with a discussion and some thoughts on future research.

2. Motivation

A common solution in the statistical literature for dealing with missing observations is imputation. Imputation methods fill in the missing values based on the other data points, to allow for model fitting. Though imputation methods are common for large observational datasets, we believe these methods are problematic in an experimental design situation because of the relatively small number of runs. Thus, we do not consider this a viable solution to many experimental, missing-observation problems.

Obviously, if possible, the preferred option is to redo the runs to fill in the missing observations. However, if an experimental run fails during an experiment, it is sometimes difficult or impossible to redo. Consider, as an example, a manufacturing process experiment that uses actual production equipment. Laboratory or pilot plant experiments may be useful in the early phases of experimentation to suggest acceptable factor ranges and a rough idea of the optimal factor region. Ultimately a final experiment using the production equipment will be necessary to determine the best operating conditions for that equipment, because the results from the laboratory or pilot plant are not typically directly generalizable due to varying conditions in the respective facilities. In order to maximize manufacturing throughput and minimize production costs, it is not unusual to have production equipment running constantly. It becomes very expensive – though necessary – to pause production to execute a series of experimental runs on that equipment. As soon as the experimental runs are completed, production resumes and the experimental runs are tested and analyzed. If a run was inadvertently missed, there will be strong resistance to shutting down production again for a single run. In this environment, testing errors made on the experimental runs could render one or more of the runs unusable. These missing values are generally not tied to one of the experimental factors of interest and can thus be treated as missing at random.

Missing runs can also occur in new product development. With proper planning, an adequate amount of custom-made raw material can be obtained from a supplier for use in an experiment. In some situations, however, the material is limited because of cost or scarcity. In this case, a failed run would require additional material which is likely to be costly and time-inefficient to obtain. A probable outcome is that the run is simply lost to the experiment.

Complex processes can also cause missing runs. In today’s global and complex manufacturing environment, it is not unusual to have processes whose steps span multiple, possibly far-flung, locations. In these situations, it may be difficult or impossible to do a full experiment across all process steps involving all factors so experiments will be performed on localized portions of the process. However, in order to obtain the final measurements on the experimental runs it may be necessary to continue processing them for multiple steps beyond those within the scope of the experimental factors of interest. Inadvertent events in downstream processing may result in a loss of some of the experimental runs and it is difficult to redo the runs without a significant time lag or change in processing conditions.

As mentioned previously, an assumption throughout is that any missing observations are missing at random. There are certainly many cases for which the assumption does not hold—for instance, runs near the edge of the experimental region may have a higher probability of yielding unusable or no information. This can occur with a poor selection of factor levels and can be prevented with more thorough exploration of factors ranges prior to formal experimentation. We concentrate on those cases where the missing observations are independent of the factor levels used.

Industrial experience suggests that the number of missing observations is generally low (<20%). If the percentage of missing observations is large, it is clear that little information may be obtained from the experimental results and it is likely that more fundamental issues need to be addressed before attempting to execute the experiment. Here, we focus on situations in which an observation is missing with a low probability. This is important to note because the evaluations we perform implicitly or explicitly make this low-probability assumption.
3. Evaluating the impact of missing observations

In preparation for the evaluation of designs in terms of their robustness to missing observations, we establish the linear model setting within which we are working, and present and motivate several ways to summarize a design’s missing-robustness.

Throughout, we assume that the typical linear regression model will be fit:

\[ y = X\beta + \epsilon, \]

with \( \epsilon \sim N(0, \sigma^2 I) \), where \( y \) and \( \epsilon \) are \( n \times 1 \) vectors, \( X \) is the \( n \times p \) expanded design matrix, and \( \beta \) is a \( p \times 1 \) vectors of unknown parameters. The least squares estimates of \( \beta \) are \( \hat{\beta} = (X^T X)^{-1} X^T y \) which implies that \( \text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \). We consider two particular model forms. First, for screening applications in which the goal is to find the subset of active factors, the model includes just the main effects. Referring to the linear model (1), let \( f^T(x) \) be a row of \( X \) so that

\[ f^T(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i, \]

where \( x = \{x_1, x_2, \ldots, x_k\} \) and \( k \) is the number of factors. For response surface experiments, the usual model is

\[ f^T(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{j=1}^{k} \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_i x_i^2. \]

In what follows, we will emphasize two sorts of criteria. First, we will describe the evaluation of design efficiency when one, two, or three runs are missing. For screening designs we focus on D-efficiency, and for response surface designs we examine both I- and D-efficiencies. Secondly, we consider two summative missing-robust criteria. The first relates to leverage values (Box and Draper, 1975) and the second is an adaptation of the criterion of Andrews and Herzberg (1979), which is intuitively appealing because it balances D-optimality with robustness against missing values.

3.1. Evaluation of D-efficiencies when \( m \) runs are missing

D-optimal designs take their cue from the variance of the least-squares estimators, and include design points such that \( |(X^T X)^{-1}| \) is minimized. In order to assess the impact of missing runs, Herzberg and Andrews (1976) and Andrews and Herzberg (1979) used a criterion based upon the loss of D-efficiency. We adopt a similar approach but consider the performance of a design when any set of \( m \) runs are missing. Because there are \( \binom{n}{m} \) combinations of \( m \) missing runs, we calculate \( D_T(i, m) \) as the efficiency of a design, with respect to the \( n \)-run optimal design, when the \( i \)th combination of \( m \) runs is missing:

\[ D_T(i, m) = \left( \frac{X_{n-m,i}^T X_{n-m,i}}{|(X^*)_n^T(X^*)_n|} \right)^{1/p} , \quad i = 1, 2, \ldots, \binom{n}{m}, \]

where \( X_{n-m,i} \) is the \( (n-m) \times p \) design submatrix and \( (X^*)_n \) is the corresponding D-optimal design with \( n \) runs. This metric gives a sense, in an absolute way, of how much information is being lost when \( m \) runs are missing. We are interested in the distributions of \( D_T \) for various designs and values of \( m \). Those designs with relatively large medians, small variances, and small reductions in efficiency as \( m \) increases are desirable as robust to missing observations.

3.2. Evaluation of I-efficiencies when \( m \) runs are missing

Though there are many optimal design criteria, I-optimal designs are appropriate for response surface models because they seek to minimize the average prediction variance across the design space. The variance of the fitted value at \( x = (x_1, x_2, \ldots, x_n) \) is \( V(\hat{Y} | x) = \sigma^2 f^T(x)(X^T X)^{-1} f(x) \). The I-criterion for a design is

\[ I = \int_R f^T(x)(X^T X)^{-1} f(x) \, dx, \]

where \( R \) represents the design region. This criterion can be computed as \( I = \text{trace}[(X^T X)^{-1} A] \) where \( A = \int_R f(x)f^T(x) \, dx \) is the so-called moments matrix which does not depend upon the design. To facilitate straightforward comparisons between the designs, we limit ourselves to cuboidal design regions. For this design region and the full quadratic model, the I-criterion can be computed in closed-form (see Goos and Jones, 2011). This design region also means that the central composite designs that we study are face-centered. The I-efficiency of a design with respect to another design with \( I^* \) can be defined, then, as \( I_{eff} = I^* / I \).
Analogous to the D-efficiency criterion, we define \( I_f(i, m) \) as the l-efficiency of a design with respect to the \( n \)-run l-optimal design, when the \( i \)th combination of \( m \) runs is missing:

\[
I_f(i, m) = \frac{I^n}{I^{n-m,i}}, \quad i = 1, 2, \ldots, \binom{n}{m},
\]

where \( I^{n-m,i} \) is the \( i \)th \( n-m \)-run design evaluated according to the l-criterion and \( I^n \) is the \( n \)-run l-optimal design. As before, we are interested in the distribution of these l-efficiencies for various designs and values of \( m \).

3.3. Leverage

Based on the regression model (1) and the least squares estimates \( \hat{\beta} = (X'X)^{-1}X'y \), we have \( \hat{y} = X(X'X)^{-1}X'y = Hy \), where \( H = \{h_{ij}\} \) is the so-called hat matrix and its diagonals the leverage values.

To justify the use of leverage values in evaluating the missing-robustness of a design, we can make an argument similar to Box and Draper (1975). Since \( \hat{y}_i = h_{11}y_1 + h_{22}y_2 + \cdots + h_{nn}y_n \), if the \( u \)th observation is missing, the change in the \( i \)th predicted value is simply \( \delta_i = -h_{uu}y_u \). If we wish to aggregate the effect of the missing observation \( u \) over all observations, we might define

\[
\sum_{i=1}^n \delta_i^2 = y_u^2 \sum_{i=1}^n h_{uu}^2 = y_u^2 h_{uu}.
\]

where the last equality follows from the fact that \( H \) is idempotent. Since \( \sum_{i=1}^n h_{ii} = p \), we cannot minimize the sum of the leverage values. Instead, as Box and Draper suggest, we might desire the leverage values to be as uniform as possible. This can be accomplished by minimizing the leverage criterion \( h = \sum_{i=1}^n h_{ii} \). We compare designs in Sections 4 and 5 using this criterion.

3.4. The truncated Herzberg–Andrews criterion

We now present a final criterion that we will use to evaluate designs in terms of missing-robustness. However, here we are slightly more ambitious: we will also use this criterion to construct missing-robust designs.

3.4.1. The full criterion

Herzberg and Andrews (1976) introduced a quantity that reduces to the determinant of the reduced information matrix when any set of \( i (i = 1, 2, \ldots, n) \) observations are missing. Andrews and Herzberg (1979) went further, proposing a design construction criterion based on the expected value of a related quantity. That is, they constructed a criterion that serves as a measurement of the expected precision after taking into account all the possible ways observations could be missing. The larger this expectation, the greater the robustness.

Consider the value of \( |X'D^2X|^\frac{1}{2} \), where \( p \) is the number of model parameters and \( D^2 \) is a diagonal matrix with \( d_i^2 \) on the diagonal, such that \( d_i^2 \) is:

\[
d_i^2 = \begin{cases} 0 & \text{with probability } \alpha(x), \\ 1 & \text{with probability } 1 - \alpha(x). \end{cases}
\]

The value of \( d_i^2 \) indicates whether observation \( i \) is missing. For instance, if \( d_i^2 = d_j^2 = 0 \), then \( |X'D^2X|^\frac{1}{2} = |X'_jX'|^\frac{1}{2} \) with \( X_j \) the design matrix resulting when the \( i \)th and \( j \)th points are omitted. If all diagonal entries in \( D^2 \) are 1, then this quantity corresponds to a version of the D-criterion, \( |X'D^2X|^\frac{1}{2} \). Here \( \alpha(x) \) specifies the probability of losing the unit that received treatment \( x \) in the design, and we note that in general its value could vary depending on where the points are in the design space.

The criterion which we call HA, studied by Andrews and Herzberg (1979), is defined as the expectation of \( |X'D^2X|^\frac{1}{2} \). If we assume that missing a particular run is independent of missing any other run, then \( \alpha(x) = \alpha \) for every design point and an expression for the HA criterion is:

\[
E \left( |X'D^2X|^\frac{1}{2} \right) = (1 - \alpha)^n |X'X|^\frac{1}{2} + \alpha(1 - \alpha)^{n-1} \sum_{i=1}^n |X'_iX'_i|^\frac{1}{2} + \alpha^2(1 - \alpha)^{n-2} \sum_{i < j} |X'_iX'_j|^\frac{1}{2} + \cdots + \alpha^n = \sum_{\ell=0}^n \alpha^\ell(1 - \alpha)^{n-\ell} T_\ell,
\]

where \( T_\ell \) is a function of the design matrix.
where conceptually $T_i$ stands for the sum of all the possible $|X^T X|^{1/p}$ when there are $i$ observations missing. Though there could be a non-zero probability that more than $n - p$ observations will go missing, this would result in a singular system and the element in the criterion would be taken as 0.

The HA criterion is a composite of design efficiency and robustness since it incorporates the $D$-criterion not only when all observations have been retained, but also when one or more are missing. The main challenge associated with this criterion is its computational difficulty. Andrews and Herzberg (1979) showed that $|X^T D^T X|^{1/p}$ can be simplified to $|X^T X| (1 - \alpha)^p$ when we disregard the $(1/p)$th-root and let $\alpha(x) = \alpha$, which suggests that by this measure $D$-optimal designs may be relatively missing-robust. However, one can imagine the complexity of the criterion in general since to evaluate it involves the computation of up to $\sum_{i=0}^{n-p} (^n_i)$ determinants of $(p - i) \times (p - i)$ matrices.

### 3.4.2. The truncated criterion

We wish to evaluate and construct designs using the HA criterion, so it is important that the criterion is computationally feasible. To ameliorate the computational difficulty, we propose two simplifications. First, we truncate the criterion so that we only include the first four terms of the original. Secondly, we take advantage of some previously developed algebraic simplifications that reduce the computational complexity of the truncated quantity. The truncated HA criterion is as follows:

$$
THA = \mathbb{E} \left( |X^T X|^{1/p} \right) = (1 - \alpha)^n |X^T X|^{1/p} + \alpha(1 - \alpha)^{n-1} \sum_{i=1}^{n} |X_i^T X_i|^{1/p} + \alpha^2 (1 - \alpha)^{n-2} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i^T X_j|^{1/p} + \alpha^3 (1 - \alpha)^{n-3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} |X_i^T X_j X_k|^{1/p}.
$$

Under the assumption that the number of missing observations is binomially distributed, the mean number of unusable observations out of $n$ trials is $n\alpha$. In industrial applications, both $n$ and $\alpha$ are typically small. In fact, based on experience we believe that the interval $(0, 0.2)$ is a reasonable range for $\alpha$. This implies that the distribution for the number of runs missing is right-skewed, suggesting that focusing on the first four terms will not compromise the full criterion heavily in many realistic situations. However, we acknowledge that this does limit somewhat the applicability of this criterion and the designs constructed from it.

The truncated criterion can be simplified in a way that will reduce computational demands. More specifically, we can rewrite the criterion in terms of $|X^T X|$ and elements of the hat matrix, $H = [h_{ij}]$. These simplifications were first introduced by Andrews and Herzberg (1979), but we reproduce them here for clarity and completeness.

Consider the second term of the truncated criterion, $\alpha(1 - \alpha)^{n-1} \sum_{i=1}^{n} |X_i^T X_i|^{1/p}$. This corresponds to the case where exactly one observation is missing. We focus on the determinant part of the term, $|X_i^T X_i|$, in which we assume that the $i$th observation, represented by the $1 \times p$ expanded design point $x_i$, is missing. Because $X^T X = \sum_{i=1}^{n} f(x_i) f^T(x_i)$, we have that $X_i^T X = X^T X - f(x_i)f^T(x_i)$. So, using standard determinant results, $|X_i^T X - f(x_i) f^T(x_i)| = |X^T X| R_i$ where $R_i = 1 - h_{ii}$.

Thus, the summation in the second term in (5) can be simplified to $|X^T X|^{1/p} \sum_{i=0}^{n-1} R_i^{1/p}$. Similarly, when two observations are missing, $|X^T X|$ becomes $|X_{ij}^T X_{ij}| = |X^T X| R_{ij}$ where $R_{ij} = (1 - h_{ii})(1 - h_{jj}) - h_{ik}^2(1 - h_{ij}) - h_{ik}^2(1 - h_{ij}) - h_{ik}^2(1 - h_{ij}) - h_{ik}^2(1 - h_{ij}) - 2h_{ij}h_{ik}h_{jk}$.

Putting these together, we can compute the truncated criterion as

$$
THA = |X^T X|^{1/p} \left( (1 - \alpha)^n + \alpha(1 - \alpha)^{n-1} \sum_{i=0}^{n-1} R_i^{1/p} + \alpha^2 (1 - \alpha)^{n-2} \sum_{i,j}^{n} R_{ij}^{1/p} + \alpha^3 (1 - \alpha)^{n-3} \sum_{i,j,k}^{n} R_{ijk}^{1/p} \right).
$$

### 3.4.3. Constructing THA-optimal designs

In this section, we discuss how the coordinate exchange algorithm (Meyer and Nachtsheim, 1995) can be adapted to use the THA criterion. The basic coordinate exchange algorithm starts with a randomly generated initial design and considers exchanges of the first element in the first row, comparing the corresponding criterion values and choosing the exchange which results in the largest increase in the criterion value. Then, it proceeds to consider exchanges with the second element of the first row and finds its best level. After all the factors in one row of the design have been examined, we jump to the next row and repeat the process until all the factors in the different rows are updated. We continue this process until the algorithm converges to a locally optimal design.

The coordinate exchange procedure is very simple, yet quite powerful because it can be easily adapted to many criteria of interest, with the computational difficulty commensurate to the computational complexity of the criterion. In our
implementation, we adapted Matlab’s cordexch to the THA criterion. Clearly, the truncated HA criterion is significantly more computationally intensive than the D-criterion. However, based on (6), we only need to update $|X'X|$ and the hat matrix. Furthermore, both the computation of $|X'X|$ and the hat matrix can be done by QR factorization. Thus we are able to combine the coordinate exchange algorithm and (6) to produce designs that are robust against missing observations. For all THA-optimal designs, we used $\alpha = 0.1$ and 100 algorithm tries. The value of $\alpha$ is application-specific and THA designs could be created for any reasonable value. For our purposes, we used the midpoint of the range of values that we believe are most likely.

4. Impact of missing observations on screening designs

Given that the primary objective of our work is to compare classical and optimal designs in terms of their robustness to missing observations, in this section we consider the impact of missing observations on designs commonly used for factor screening. We focus primarily on fractional factorial designs (FFDs), along with D-optimal designs that have become more commonly used in recent years. We also include in our comparisons the THA-optimal designs of Section 3.4. The model of interest is (2), which includes only main effects.

Two-level factorial designs are widely used in screening experiments to investigate main effects and interaction effects. When the number of factors is more than about 4 or 5, the number of runs required in a factorial design is often prohibitively large, so fractional factorial designs are frequently used instead. Resolution III fractions – those that cannot estimate any interactions free of the main effects – are often used when main effects are of primary interest (see Wu and Hamada, 2009). Alternatively, D-optimal designs can also effectively estimate the first-order model (see Goos and Jones, 2011). For each design, we calculate the D-efficiencies from (4), for each set of 1, 2 and 3 missing runs, as well as the leverage criterion of Section 3.3 (lower is better) and the THA criterion of Section 3.4.2 (higher is better).

Consider the situation in which $k = 5$ and the experimental goal is to determine which of the main effects are significant, using $n = 16$ or fewer runs. Two classical designs are the resolution III $2^{5-2}$ fraction with $I = ABCD = BCE = ADE$, and the resolution $V$ $2^{5-1}$ fraction with $I = ABCDE$ which have 8 and 16 runs, respectively. We assess these five-factor fractional factorial designs, along with D-optimal and THA-optimal designs with $n = \{8, 10, 12, 16\}$. Results are shown in Fig. 1 and Table 1.

From the results, we observe little difference among the designs with the same number of runs. For $n = 8$, the D-optimal and THA-optimal designs that we used are regular fractions, with defining relations $I = ABCE = ABD = CDE$ and $I = aBCE = bCD = ADE$, respectively. Thus, it is unsurprising that the three designs appear equivalent in terms of missing-robustness. For $n = 16$, the D-optimal and THA-optimal designs are orthogonal, but not regular fractions. All three 16-run designs are D-optimal, and from Table 1 it is clear that they are equivalent in terms of the THA and leverage criteria as well. This illustrates that there can be multiple D-optimal designs and in fact neither the THA nor leverage criteria will necessarily discriminate between them. Interestingly, there appears to be a slight difference between the regular fraction
and the other two designs in terms of D-efficiencies when two or three runs are missing (Fig. 1). That is, the fractional factorial has less variation in terms of D-efficiency than the other two. This may be because the FFD is a regular fraction while the others are not. There is little or no difference between the D-optimal and THA-optimal designs when \( n = 10 \) and \( n = 12 \).

In Fig. 1, we see that as the number of missing runs increases, the D-efficiency values tend to decrease while their variability increases. When \( n = 8 \), two missing runs often result in an inestimable model and three missing runs guarantee an inestimable model. Thus, the impact of a few missing runs is exaggerated when the original design is close to saturated.

We performed additional investigations for \( k = 10 \), but do not report them since similar conclusions were obtained.

Overall, compared to the Truncated HA missing-robust criterion, the classical and optimal designs are not only robust, but nearly equivalent. Furthermore, the impact of missing runs on screening designs appears small as long as the number of runs is sufficient.

### 5. Impact of missing observations on response surface designs

Now we move beyond screening designs to study the impact of missing values on designs used to fit the quadratic model (3). As before, we wish to compare classical and optimal designs, so we study the most common response surface design, the central composite design (CCD, see Myers et al., 2009), along with I-optimal designs (see Goos and Jones, 2011) and the THA-optimal designs of Section 3.4. We consider experiments with both \( k = 3 \) and \( k = 5 \), for a variety of sample sizes, and assess the impact of 1, 2, and 3 missing observations on both I- and D-efficiencies. We also compare the designs in terms of the leverage and THA criteria.

#### 5.1. Three-factor response surface experiments

For a CCD with three factors, there are typically \( n_f = 2^3 = 8 \) runs for the factorial portion of the design, 6 runs for the face-centered axial points, and \( n_c \) center points. Here, we take \( n_c = 2, 4, 6 \) and consequently study the three-factor, face-centered CCD with 16, 18, and 20 runs, respectively. For these three run sizes, as well as \( n = 12 \), we constructed I-optimal designs using JMP® (JMP, 2015) software and THA-optimal designs using the methods described in Section 3.4.

In Fig. 2, we see that the efficiency losses due to missing observations are especially acute when \( n = 12 \). There appears to be little difference between the efficiency loss for the central composite and I-optimal designs, though the optimal designs are generally higher in both I- and D-efficiencies. Because the THA criterion is closely related to the D-criterion, the THA-optimal designs have a higher baseline D-efficiency than the other designs and in fact appear to be D-optimal. However, the THA-optimal designs have a lower baseline I-efficiency, though they seem to absorb less efficiency loss than the other designs as the number of missing observations increases.

The first part of Table 2 reveals that when \( k = 3 \) the CCDs and I-optimal designs have similar THA criterion values, while the THA-optimal designs are predictably superior. The THA-optimal designs also tend to have more uniform leverage values, while the I-optimal designs are a bit better on this measure than the CCDs.

#### 5.2. Five-factor response surface experiments

For \( k = 5 \), the standard CCD has either \( n_f = 2^{5-1} = 16 \) or \( n_f = 2^5 = 32 \) factorial points, 10 axial points, along with \( n_c \) center runs. We considered four CCDs with \( n = \{28, 32, 36, 44\} \), \( n_f = \{16, 16, 16, 32\} \), and \( n_c = \{2, 6, 10, 2\} \), respectively. We also constructed I-optimal and THA-optimal designs with \( k = 5 \) and \( n = \{24, 28, 32, 36, 44\} \). (For \( n = 36 \), the number of center runs in the CCD is excessive; better classical-type designs could be chosen, but we retained this one to maintain consistency in the types of designs we compared.)

Compared to \( k = 3 \), Fig. 3 shows larger efficiency differences between the CCDs and the I-optimal designs, though this may be partially explained by the CCDs with 6 and 10 center runs. There is no obvious difference between the designs in
terms of efficiency loss. The THA-optimal designs are again essentially D-optimal, and are typically inferior in terms of the I-criterion and superior in terms of the D-criterion. Interestingly, when design resources are constrained (n = 24 and n = 28), the THA-optimal designs are competitive with the I-optimal and central composite designs, in terms of I-efficiency, as the number of missing observations increases.

Note that several of the THA-optimal designs have a D-efficiency greater than 100%. This is because the “D-optimal” designs generated to serve as D-efficiency baselines were algorithmically generated using JMP® software (JMP, 2015) and

Table 2
THA and leverage criterion values for response surface designs.

<table>
<thead>
<tr>
<th>k</th>
<th>n</th>
<th>Design</th>
<th>THA value</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12</td>
<td>I-optimal</td>
<td>3.50</td>
<td>8.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>3.90</td>
<td>8.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>5.46</td>
<td>6.86</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>I-optimal</td>
<td>5.46</td>
<td>6.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>5.88</td>
<td>6.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>5.53</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>I-optimal</td>
<td>5.96</td>
<td>6.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>6.56</td>
<td>5.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>6.31</td>
<td>5.66</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>I-optimal</td>
<td>6.31</td>
<td>5.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>7.15</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>I-optimal</td>
<td>6.16</td>
<td>18.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>7.09</td>
<td>18.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>6.54</td>
<td>17.76</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>I-optimal</td>
<td>7.14</td>
<td>17.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>8.16</td>
<td>16.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>5.87</td>
<td>17.50</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>I-optimal</td>
<td>5.08</td>
<td>15.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>8.38</td>
<td>13.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>5.11</td>
<td>18.73</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>I-optimal</td>
<td>7.02</td>
<td>14.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>8.25</td>
<td>12.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCD</td>
<td>6.24</td>
<td>10.29</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>I-optimal</td>
<td>5.95</td>
<td>11.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>THA-optimal</td>
<td>7.05</td>
<td>10.10</td>
</tr>
</tbody>
</table>
Fig. 3. I- and D-efficiencies for possible designs for the second-order model, for \( k = 5 \) and \( n = 24, 28, 32, 36, \) and \( 44 \), according to the number of missing runs.

thus do not guarantee a globally D-optimal design. This suggests that the D-criterion alone may be insufficient to evaluate a design; here we have THA-optimal designs which are clearly superior in terms of robustness (second part of Table 2) but also as good or better in terms of the D-criterion than the designs generated explicitly to produce D-optimal designs. These results also highlight the similarity between the THA-criterion and the D-criterion.

The second part of Table 2, with \( k = 5 \), shows that the I-optimal designs are typically more robust than the CCDs in terms of the THA and leverage criteria. However, this is slightly misleading because the experiments for which the disparity is most pronounced (\( n = 32 \) and \( n = 36 \)) are cases in which the CCD has an abnormally large number of center points (6 and 10, respectively). When \( n = 44 \) and the full \( 2^5 \) factorial is used, the CCD actually outperforms the I-optimal design. The THA-optimal designs are clearly preferred on the THA and leverage criteria.

Overall, the CCDs and I-optimal designs do not appear to differ substantially in terms of their efficiency loss as the number of missing observations increases. However, since the I-optimal designs are more I-efficient, this advantage is retained even when design runs are missing. The THA-optimal designs are more robust and D-efficient, but less I-efficient. Their lack of I-efficiency reduces their utility as response surface designs except when resources are severely constrained.

6. Discussion and conclusions

We have presented some empirical results for the effect of missing observations on various classical and optimal designs, as well as a new type of missing-robust design, in both screening and response surface settings. In general, we found missing observations have little impact on screening designs as long as the number of runs in the initial design is not too small. The D-optimal and fractional factorial designs were also robust when compared to the new THA-optimal missing-robust screening designs. A practitioner who wants to have a more robust design to protect against the possible loss of information from missing observations would do well to add a few additional runs to the appropriate design.

Secondly, we did not find evidence that optimal response surface designs are less robust to missing observations than classical designs. In terms of efficiency, they are superior though that is expected based on their construction. More interestingly, they performed well in terms of average efficiency when a few observations were lost. The classical CCDs had a similar amount of efficiency loss, but were less efficient overall. Since the new THA-optimal missing-robust designs are based on the D-criterion, they suffered significantly in terms of I-efficiencies. However, they appeared to have less efficiency loss in the face of missing observations. Notably, when the number of observations was small relative to the model being estimated, these missing-robust designs sometimes outperformed their I-optimal counterparts in terms of I-efficiency, when two or three runs were missing. These THA-optimal designs also tend to be D-optimal, or nearly so, which suggests that D-optimal designs have good missing-robustness properties.

These results should inspire some confidence that optimal designs are similar to their classical counterparts in terms of robustness to missing observations. We also suggest that for severely resource-constrained experiments in which missing observations are a nonnegligible possibility, either extra runs should be added or an explicitly missing-robust design should
be used. Finally, we note that there are no missing-robustness criterion in the literature that are based on I-optimality; such a criterion might produce designs that are attractive alternatives to an I-optimal design when the possibility of missing observations cannot be ignored.

Appendix A. Supplementary data

The designs used in this study are included as material supplementary to this article, which can be found online at [http://dx.doi.org/10.1016/j.csda.2016.12.001](http://dx.doi.org/10.1016/j.csda.2016.12.001).

References


Factorial and response surface designs robust to missing observations

Marcelo A. da Silva, Steven G. Gilmour, Luzia A. Trinca

Programa Interinstitucional de Pós-Graduação em Estatística, USP/UFSCar, São Carlos, Brazil
Department of Mathematics, King’s College London, UK
Departamento de Bioestatística, IB, Universidade Estadual Paulista, Botucatu, Brazil

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form 28 April 2016
Accepted 13 May 2016
Available online 7 June 2016

Keywords:
Compound criteria
Cook’s distance
Leverage
Optimum design
Pure error

ABSTRACT

Compound optimum design criteria which allow pure error degrees of freedom may produce designs that break down when even a single run is missing, if the number of experimental units is small. The inclusion, in the compound criteria, of a measure of leverage uniformity is proposed in order to produce designs that are more robust to missing observations. By appropriately choosing the weights of each part of the criterion, robust designs are obtained that are also highly efficient in terms of other properties. Applications to various experimental setups show the advantages of the new methods.

1. Introduction

Processes, products and methods in many areas are discovered and improved by performing controlled experiments in which the levels of several continuous inputs, experimental factors, are manipulated and at least one outcome is measured. Empirical models, such as low order polynomials, relating the response to the factor levels have been extremely useful for interpreting the data from such experiments. Such models and methods are part of the large area of Response Surface Methodology. Designs for experiments in this setup are known as Response Surface (RS) designs.

It has long been recognized that the experimental design should have several good properties. In the context of RS, Box and Draper (1975) started a list that was subsequently enlarged (Box and Draper, 1987, 2007) to 14 desired properties, some of them conflicting, indicating that in practice it is wise and necessary to compromise in order to choose a good design.

On the other hand, optimum design methodologies have concentrated on variance-based criteria such as $D$-, $A$- and $I$-optimality, the so called alphabetical optimality, see Atkinson et al. (2007) for an account of design criterion definitions. The use of a single optimality criterion may lead to designs that lack practical appeal. Gilmour and Trinca (2012) re-defined the alphabetical optimality criteria such that their properties are valid under inferences based on the randomization process only. They proposed adjustments to the traditional criteria allowing for pure error degrees of freedom in order to appropriately estimate random variation, the so called $DP$ and $AP$ criteria for instance. However, as recognized by the authors, these criteria may produce extreme designs with no spare degrees of freedom for inclusion of additional model terms. They further proposed compound criteria that aggregate into a single function the properties reflecting four experimental objectives, including a simple component, based on degree of freedom efficiency (Daniel, 1976) to drive the
design to allow some lack of fit degrees of freedom as well. The use of compound criteria as well as procedures for multiple objectives (Lu et al., 2011) has the power to produce designs that are very statistically efficient and useful for experimenters. Concerning the extreme designs produced by using a single property, e.g. $DP$, it was pointed out by Ridout (2012) that small designs would break down in case of even one missing observation from some treatment units. Robustness to missing observations is closely related to insensitivity to wild observations, a desired design property highlighted by Box and Draper. A design is said to be robust to missing observations if the model parameters are still estimable, without too much loss of precision, when observations from some experimental units are not available. Just as there are different design optimality criteria for estimation and for inference, there are different criteria for robustness. Surrogate measures related to the so called leverages, associated with a regression model, have been used to compare designs in this sense, as well as measures related to precision.

For example, Box and Draper (1975) studied the connections of the sum of squares of leverages and other design measures and found the best replication of center points and axial point values in central composite designs (CCD). Herzberg and Andrews (1976) and Andrews and Herzberg (1979) noted that such a measure does not discriminate well between designs and proposed extended measures incorporating some probability for the event of a missing observation. Akhter and Prescott (1986) developed an efficiency measure relating the $D$ criterion and the leverages and compared several CCDs, while Ahmad and Gilmour (2010) studied efficiency loss with respect to several optimality criteria due to missing data from different types of points in subset designs (Gilmour, 2006) and Ahmad et al. (2012) did the same for augmented pairs designs. Related investigations were also presented by Ghosh (1982a,b) who studied robustness of certain designs under sets of $s$ missing runs and found the maximum $s$ value for given designs. Adding to these works, Ghosh (1983, 1989) proposed measures to study influence on estimation and prediction of observations. To the best of our knowledge, a property related to robustness to missing data has not yet been incorporated in a criterion function in order to algorithmically construct an efficient RS design robust to missing data.

In this paper we incorporate a measure for the contribution of leverages, related to Cook’s distance, in a compound design criterion in order to prevent the optimal design from being too sensitive to some observations or to breakdown in case of missing data. We show through several examples that such a property is particularly important in the case of limited experimental resources. In Section 2, a criterion for assessing design robustness is developed and in Section 3 a brief description of the algorithm is presented. The proposed criterion is shown to work well in several illustrative experiments in Section 4. Some final comments are made in Section 5.

2. Efficient and robust designs

Consider a completely randomized design in which there are $t$ treatments, the distinct combinations of levels of $q$ quantitative factors, to be allocated to $n$ experimental units ($t < n$), treatment $r$ being replicated $n_r$ times ($n_r \in \mathbb{N}$, $\sum_{r=1}^{t} n_r = n$). The underlying model for the continuous random response variable $Y$ is

$$y_{ij} = \mu + \alpha_f r + \varepsilon_{ij} \quad r = 1, 2, \ldots, t; j = 1, 2, \ldots, n_r,$$

(1)

where, in matrix notation, $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2I$. Once data are collected the fitting of this model allows $d = n - t$ pure error degrees of freedom to estimate $\sigma^2$ unbiasedly. As argued in Box and Draper (2007), in RS experiments we want to simplify the model and add interpretability by approximating

$$\mu_r \approx f(x_r)' \beta \quad r = 1, 2, \ldots, t,$$

(2)

where $x_r$ is the vector of levels of the $q$ factors defining treatment $r$ (the design experimental points), $f$ is the function that expands the levels according to the desired approximating function, usually a low order polynomial, and $\beta$ is the $p$-dimensional vector of parameters with its first element being the intercept denoted by $\beta_0$. In matrix notation, let $X_p = [1 \mid X]$ be the $n \times p$ model matrix for Eq. (2) where each row of $X_p$ corresponding to treatment $r$ is $f(x_r)'$, $1$ is the $n$ dimensional column vector with all elements equal to 1 and $X$ is a $n \times (p - 1)$ matrix.

For the $DP$ design criterion (Gilmour and Trinca, 2012) we should minimize $(F_{p,d;1-\alpha})^p/[X_p'X_p]$, where $F_{p,d;1-\alpha}$ is the $1-\alpha$ quantile of the $F$ distribution with $p$ numerator and $d$ denominator degrees of freedom and $1 - \alpha$ is the confidence level of the confidence region for the $p$-parameter vector $\beta$. Other alphabetical optimality criteria can be defined similarly. Using the $(DP)_{\alpha}$ criterion, for the case of interest in a subset of $p_2$ ($p_2 < p$) parameters we should minimize $(F_{p_2,d;1-\alpha})^{p_2}/([M^{-1}]^{22})$, where $M = X_p'X_p$ and $(M^{-1})^{22}$ is the portion of $M^{-1}$ referring to the subset of $p_2$ parameters of interest. See Atkinson et al. (2007) for details on $D_3$ and other useful design criteria. If we use $p_2 = p - 1$ which drops only the intercept ($\beta_0$) from the set of parameters of interest, minimizing $([M^{-1}]^{22})$ is equivalent to maximizing $|X'QX|$, where $Q = I - 1/41'$ and $I$ is the $n \times n$ identity matrix. Focusing on four design objectives, each with a priority weight $k_1$ ($\sum_{k=1}^{4} k_1 = 1$), representing

1. global $F$ test for treatment effects in $\beta$, with significance level $\alpha_1$;
2. partial confidence intervals for each regression parameter each with confidence level of $1 - \alpha_2$;
3. point estimation of each regression parameter; and
4. lack of fit degrees of freedom,
Gilmour and Trinca (2012) formulated the compound criterion

$$\frac{|X'QX|_F^{1/2} (n - d)^{s_2}}{(F_{p-1,d:1-\alpha_1})^{s_1} (F_{1,d:1-\alpha_2})^{s_2} [\det (W(X'QX)^{-1})]^{1/2 + \gamma_1}}.$$

where $W$ is a diagonal matrix of weights for a weighted-$\chi^2$ criterion and $\alpha_1$ and $\alpha_2$ are the significance levels used in objectives 1 and 2, respectively.

As discussed in Ridout (2012), for small $n$, considering only the first objective may result in designs that break down in the case of a single missing observation. For design breakdown we use the same meaning as Ghosh (1982a,b), that is, the rank of the $M$ matrix relative to the reduced design is less than $p$, if the data from some experimental unit is lost, and thus it is not possible to estimate all the elements of $\beta$. One step in the direction of constructing efficient designs with respect to several properties and simultaneously safeguarding against the related missing observation problems is incorporating in the compound criterion some measure to guide the design search in this respect. Missing observations are a fairly common problem in response surface studies, since many of the combinations of factor levels will never have been studied before and might lead to no response being possible. We will see an example of this in practice in Section 4.

Ghosh (1982a) defined a design to be robust against the unavailability of $s$ observations if the design does not break down after omitting any set of $s$ runs. A complete investigation of robustness would require evaluation of the design under each possible set of $s$ missing data points, which can become very computationally intensive. In practice, if the experimenter faces a situation of high risk of having more than one or two missing observations, he or she should be prepared to start with a reasonably large $n$, in which case, the sensitivity of the design to unavailable data should be low. Our development, based on well known results in linear models, leads to the use of a surrogate measure of sensitivity that will prevent the design from breaking down when a single observation goes missing. From least squares theory, e.g. Hoaglin and Welsh (1978), if observation $i$ is removed from the data, the covariance matrix of $\hat{\beta}_{(-i)}$, the new estimator of $\beta$, except for the constant $\sigma^2$, is

$$M_{(-i)}^{-1} = (X_p'X_p)^{-1} = M^{-1} + \frac{M^{-1}f(x_i)f(x_i)'M^{-1}}{1 - h_i},$$

where $h_i$, called leverage, is the $i$th element of the diagonal of the projection or hat matrix given by $H = X_pM^{-1}X_p'$.

It is well known that $\text{rank}(H) = \text{trace}(H) = p$ and $\frac{1}{n} \leq h_i \leq 1$. The ideal $h_i$ value is $p/n$ ($i = 1, 2, \ldots, n$) in which case the contribution from each observation to estimate its response is the same for all points, none of them being influential due to the design. For replicated $X$, the reciprocal of the number of replications is the upper bound for $h_i$, so only for unreplicated treatments can $h_i$ reach the value 1. From Eq. (4) it is easily seen that if $h_i = 1$ and observation $i$ is removed from the data the design breaks down in the sense that the reduced data does not support the fitting of the intended model. Cook and Weisberg (1982) show that

$$h_i = \frac{1}{n} + \hat{x}_i'X_i^{-1}X_i,$$

where $X_i$ is the $i$th row of $X$, and

$$h_i = \frac{1}{n} + \sum_{i=1}^{p} \frac{(\psi_i'x_i)^2}{\lambda_i} = \frac{1}{n} + \hat{x}_i'X_i \sum_{i=1}^{p} \frac{\cos^2(\theta_i)}{\lambda_i},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of $X'X$, $\psi_1, \psi_2, \ldots, \psi_p$ are the corresponding eigenvectors and $\theta_i$ is the angle between $\psi_i$ and $x_i$. Thus $h_i$ is large if $X_i'x_i$ is large, that is, $x_i$ is far away from the bulk of the design points. $h_i$ can also be large when $X_i$ is in the direction of an eigenvector corresponding to a small eigenvalue. But note that if $X_i'x_i$ is small, $h_i$ is small no matter its direction. Thus, for level balanced factors, design points closer to the center of the experimental region have small $h_i$ values as we show in the illustrations in Section 4.

Each $h_i$ can also be expressed as

$$h_i = f(x_i)'M^{-1}f(x_i)$$

such that we can write $\text{Var}(\hat{y}(x_i)) = h_i\sigma^2$ where $\hat{y}(x_i)$ is the estimate of the mean response (estimated from the fitted polynomial) under treatment $x_i$. This expression establishes the relations between leverage measures and the $G$ and $I$ criteria (Box and Draper, 1975; Ahmad and Gilmour, 2010) and, together with Eq. (4), shows the dangers of having high leverage design points. The determinant of the information matrix of the design with design points removed can also be written as a function of the full design information matrix and the elements of the $H$ matrix, as shown in Andrews and Herzberg (1979), thus giving a link to the $D$ criterion.

For illustration, Table 1 shows a $(DP)_S$ design constructed by Gilmour and Trinca (2012) together with the $h_i$ value for each design point. The design allows 6 degrees of freedom for estimating $\sigma^2$ but, if at least one observation from the set of design points 1, 6, 9 and 12 goes missing, the second order model cannot be fitted. Note that when $h_i = 1$ the estimated response from the full data matches exactly its observed value. Even if the situation is not so drastic that the design breaks
down, a design point with high leverage may cause inflation in the variances of parameter estimators if its observation goes missing. Thus, when designing an experiment, especially if the number of experimental units is limited, we should caution against the use of design points with high leverage. For p and n fixed Box and Draper (1975) considered minimizing the variance of leverages given by

\[ n^{-1} \sum_{i=1}^{n} \left( h_i - \frac{p}{n} \right)^2, \]  

(7)

or its square root, for obtaining designs robust to wild observations. They showed how this measure of robustness varied with the number of center points and the values for the axial points used in CCDs.

Other quantities based on the \( h_i \) values are appealing. For example, recall that the contribution of each leverage to Cook’s distance is \( h_i \). Thus minimizing

\[ \frac{1}{n} \sum_{i=1}^{n} \frac{h_i}{(1-h_i)^2}, \]  

(8)

when choosing a design is also a good idea. However, as discussed by Andrews and Herzberg (1979), the use of (7) or (8) or any other measure based only on the leverages as a design criterion may result in very inefficient designs. One explanation for this comes from Eq. (5) which indicates, as pointed out by Cook and Weisberg (1982), that a design with low \( h_5 \) benefits from points that are closer to the center of the region. Since such points have less information for estimating many terms of the polynomial model the design will perform poorly in terms of estimation precision.

Following the line of compromising among objectives or design properties, in this paper we consider the construction of compound optimum designs that are efficient in terms of several properties but also control for high leverage. For this last property we choose to use Eq. (8) because of its relation to common diagnostic techniques and also because we can easily define a design efficiency measure with respect to leverage. Thus Eq. (8) will be referred to as the \( H \) criterion function and the \( H \)-efficiency of design \( X \) will be calculated by

\[ H_{\text{eff}} = \frac{1}{100} \left( \frac{\sum_{i=1}^{n} h_i}{\sum_{i=1}^{n} (1-h_i)^2} \right), \]  

(9)

where the subscript \( H \) refers to the \( H \) optimum design. Note that a design with at least one point with \( h_i = 1 \) is 0% efficient under the \( H \) criterion. For the theoretical or ideal \( H \)-optimum design the numerator in (9) reduces to \( \frac{p}{n} \), but such an ideal design rarely exists so we prefer to use Eq. (9) for measuring the efficiency with respect to the leverages.

Thus, the compound criteria we propose in this paper maximize the compound function

\[ \left( \frac{\left| X'QX \right|^{1/2}}{(n-d)^{k_1}} \right)^{k_1} \left( F_{n-p,1,\alpha_1} \right)^{k_1} \left( F_{n-1-d,1-\alpha_2} \right)^{k_2} \left( \text{tr}[W(X'QX)^{-1}] \right)^{k_2 + k_3} \left( \frac{\sum_{i=1}^{n} h_i}{(1-h_i)^2} \right)^{k_5}, \]  

(10)

which can be written as

\[ \left( \frac{\left| X'QX \right|^{1/2}}{(n-d)^{k_1}} \right)^{k_1} \left( F_{n-p,1,\alpha_1} \right)^{k_1} \left( F_{n-1-d,1-\alpha_2} \right)^{k_2} \left( \text{tr}[W(X'QX)^{-1}] \right)^{k_2 + k_3} \left( \frac{\sum_{i=1}^{n} h_i}{(1-h_i)^2} \right)^{k_5}, \]  

where

\[ X_{n-p,1,\alpha_1} \]

is the upper 1\% Point of the \( F \)-distribution with \( n-p \) and 1 degrees of freedom.

### Table 1

<table>
<thead>
<tr>
<th>( (DP)S )</th>
<th>( H )</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( h_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (2) )</td>
<td>( (4) )</td>
<td>( (5) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1.000</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0.500</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0.500</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>0.500</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0.500</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.000</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0.500</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0.500</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.500</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.500</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>0.500</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>0.500</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.500</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.500</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 1

Designs for Example 1: three 3-level factors in \( n = 16 \) and \( p = 10 \).
for finding optimal designs, where $\kappa_5$ is the weight for $H$-efficiency and $\sum_{i=1}^{5} \kappa_i = 1$. Choosing different values of the $\kappa_i$ is subjective and problem dependent, like the choice of utility function in most applications. However, small changes in their values make little difference in practice and small positive weights for particular criteria do not differ much from ignoring these criteria. As suggested by Gilmour and Trinca (2012), it is reasonable to simplify the choice by using relative weights of 3, 1 and 0 for objectives that are considered major, minor or unimportant. When time allows, we would also recommend users to try various weights and to consider all the properties of the designs produced. As with all methods of optimal design, that proposed here should be considered to be mainly a way to produce interesting designs for consideration by the experimental team.

3. An algorithm

To find the optimum designs we used exchange algorithms. Both point exchange (Cook and Nachtsheim, 1980) and coordinate exchange (Meyer and Nachtsheim, 1995) versions were explored. The nature of these algorithms is sequentially performing changes in design points or factor coordinates that improve a given initial design, usually a random initial design. The search is performed many times from different initial designs in order to increase the probability of finding the best solution. In this paper we consider candidate design points belonging to the complete factorial design with the minimum number of levels necessary for the response surface model aimed at. A basic description of the algorithm for the coordinate exchange version is given in Algorithm 1 and R (R Core Team, 2016) code is provided in the Supplementary Material.

**Algorithm 1: Coordinate Exchange Algorithm**

**Input:** number of factors $k$; levels[$k$][ ]; model terms indicator; design size $N$; number of tries $N_{tries}$; weight vector for compound criteria $\kappa = (\kappa_1, \ldots, \kappa_5)$; weight vector for $W$; significance level for pure error adjustments $\alpha_1$ and $\alpha_2$; number of iterations for recalibration $N_{cali}$

1. for $m \leftarrow 1$ to $N_{tries}$ do
2.   $X$ ← model matrix from a randomly generated initial design
3.   crit ← evaluate equation (10)
4.   crit$_0$ ← crit
5.   improve ← 1
6.   while improve == 1 do
7.     improve ← 0
8.     for $j \leftarrow 1$ to $k$ do
9.       for $i \leftarrow 1$ to $N$ do
10.      exchange($X[i][j]$)
11.     complete all elements of $X[i][.]$ according to the model
12.     crit$_0$ ← evaluate equation (10)
13.     if crit$_0$ > crit then
14.       improve ← 1
15.       crit ← crit$_0$
16.     else
17.       revert exchange($X[i][j]$)
18.   if $m == 1$ then
19.     $X$ _best ← $X$
20.     crit$_{best}$ ← crit
21.   else
22.     if crit > crit$_{best}$ then
23.       $X$ _best ← $X$
24.       crit$_{best}$ ← crit
25.   Return $X$ _best

For criteria that require the calculation of the determinant and/or the inverse of the appropriate information matrix we should use updating formulas after each exchange in step 12. Here we used the methods of Cook and Nachtsheim (1989). If desired, determinants and inverses can be recalculated after each change or after a prespecified number of changes, e.g. to recalibrate calculations in order to avoid building up numerical rounding errors. However, our experience shows that recalibration is unnecessary. The number of degrees of freedom for pure error is $n$ minus the number of distinct treatments. For the point exchange algorithm, a treatment label is attached to each point in the candidate set which is carried forward during the search. For coordinate exchange, treatment labels are attached to the rows of $X$ after each change in the design.
Table 2
Properties of designs for the second order model, Example 1 (q = 3, p = 10 and n = 16).

<table>
<thead>
<tr>
<th>Design</th>
<th>Criterion</th>
<th>DF</th>
<th>Efficiencies</th>
<th>h_{max}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D_5; A_5</td>
<td>(0; 6)</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>(DP)_S</td>
<td>(6; 0)</td>
<td>83.00</td>
<td>100.00</td>
</tr>
<tr>
<td>3</td>
<td>(AP)_S</td>
<td>(5; 1)</td>
<td>93.03</td>
<td>96.17</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>(0; 6)</td>
<td>66.22</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>\kappa_1 = \kappa_5 = 0.5</td>
<td>(4; 2)</td>
<td>92.58</td>
<td>76.13</td>
</tr>
<tr>
<td>6</td>
<td>\kappa_1 = 0.3; \kappa_5 = 0.7</td>
<td>(3; 3)</td>
<td>96.62</td>
<td>54.09</td>
</tr>
<tr>
<td>7</td>
<td>\kappa_1 = 0.5; \kappa_4 = 0.5</td>
<td>(5; 1)</td>
<td>93.03</td>
<td>96.17</td>
</tr>
<tr>
<td>8</td>
<td>\kappa_1 = 0.2; \kappa_4 = 0.8</td>
<td>(4; 2)</td>
<td>95.10</td>
<td>78.21</td>
</tr>
<tr>
<td>9</td>
<td>CCD</td>
<td>(1; 5)</td>
<td>93.15</td>
<td>1.91</td>
</tr>
</tbody>
</table>

4. Applications

In this section we show the performances of the proposed criteria in four experimental layouts. We used \( \alpha_1 = \alpha_2 = 0.05 \) for the \((DP)_S\) and \((AP)_S\) criteria throughout and at least 1000 tries. The number of tries was increased when the best design appeared just once.

4.1. Example 1

We considered Example 1 from Gilmour and Trinca (2012) and constructed several new designs for the experiment which involved three three-level factors in 16 runs with the second order model as the primary model. In Table 1 we show the \((DP)_S\) - and \(H\)-optimum designs and a compromise design obtained by composing the \((DP)_S\) and \(H\) properties with equal weights. The \(h_i\) values for the design points are also shown. As already discussed in Section 2 the \((DP)_S\)-optim design has four points with the upper bound leverage value. For the \(H\)-optimum design the best that can be done is choosing all 16 points with leverages ranging from 0.571 to 0.644 (note that the ideal value would be \( \frac{10}{16} = 0.625 \)). The efficiencies in Table 2 show that this design lacks efficiency in terms of other properties. For the compromise design the \(h_i\) values range from 0.482 to 0.789.

These three designs are contrasted with several others in Table 2 (designs 1, 3 and 8 are shown in Gilmour and Trinca (2012), the CCD is the central composite design in a cubic region with two center points). We see that \(D_5\) and \(H\) single criteria produce designs that do not allow degrees of freedom (DF) for estimating pure error (see designs 1 and 4) while the pure error adjusted criteria produce designs that are poor for lack of fit checking and leverage efficiency (designs 2 and 3). The CCD has just one DF for error and thus has low efficiencies for the adjusted criteria. Designs from 5 to 8 show that we may drive the design by changing the weight pattern in the compound criterion, recalling that \(\kappa_1\) represents the weight given to checking, \(\kappa_4\) the weight given to checking for lack of fit and \(\kappa_5\) the weight given to robustness to missing values. In this case for example we obtained an attractive compromise design (design 5) when considering \((DP)_S\) and \(H\) with equal weight, which performs reasonably well in several respects and will not break down if a point goes missing. Designs 7 and 8 were obtained composing the \((DP)_S\) and DF efficiencies (as in Gilmour and Trinca (2012)) and show that although DF efficiency attenuates leverage problems, the resulting designs are different from those using the \(H\) property. Note that design 7 is very similar to design 3 but not equivalent. This highlights the value of constructing several designs and comparing them in terms of a wide range of properties, although in this case the difference is very small.

4.2. Example 2

In this illustration we use as motivation the experiment from Mountzouris et al. (1999) that studied the effects of substrate concentration \((X_1)\), enzyme concentration \((X_2)\) and transmembrane pressure \((X_3)\) on several quantitative characteristics of the product formed (types of oligodextrins). The investigation aimed to fit empirical models, in particular second order polynomials and thus the design used was a three-level CCD with four center runs \((n = 18)\). For one of the treatments, \(X_1\) and \(X_2\) at the highest levels and \(X_2\) at the lowest level (run 15 in Table 3), the reaction did not work and thus the respective run was removed from the data analysis. Fortunately the CCD design with run 15 removed still allowed the fitting of the assumed model. Note however that all points from the two-level factorial have the maximum \(h_i\) value that is close to 0.8.

We constructed several alternative designs for this experiment, some of which are shown in Tables 3 and 4. The properties of all designs are shown in Table 5. The ideal value of \(h_i\) is \( \frac{10}{16} = 0.556 \) and the \(H\)-optimum design gives \(h_i\) ranging from 0.514 to 0.589 but no pure error DF. Despite the high number of distinct treatments in this design, its efficiencies in terms of point estimation are not very high. The other designs constructed by single criteria have very low efficiencies in at least one of the properties studied. Among them, the \(D_5\) optimum design has the best performance in terms the \(H\) property but only 2 pure error DF. Not that designs 3, 4 and 11 are all \((DP)_S\) optimal but the three designs behave slightly differently with respect to the \(A_5\) and \((AP)_S\) criteria, highlighting the importance of evaluating several properties for design choice. Designs
from 6 to 13 were obtained by the compound criteria varying the associated weights. The results show that many interesting designs can be constructed and the use of many properties with some weight greater than zero on each is promising as in design 10.

4.3. Example 3

Subset designs under minimax loss due to missing design points were studied by Ahmad and Gilmour (2010). The loss for design point $i$ was defined as $h_i$. In their Example 1 they studied several possible subset designs for fitting the four factor second order model in 36 runs. Here we study the properties of their nine designs (Table 5 of Ahmad and Gilmour (2010)) in a cuboidal region and several other alternatives such as $D_5$, $A_5$, $(DP)_5$ and $(AP)_3$-optimum designs and some compromise designs obtained by using the compound criterion. Some designs are shown in Tables 6 and 7 and Table 8 shows the properties of the designs.

From the subset designs (designs 1–9) the most efficient in terms of leverage is design 4, composed of the $S_5$ subset plus some center points. It is followed by design 5 and by the modified CCDs which are more efficient in terms of pure error DF. Because the experiment is reasonably large for the model, single criterion optimum designs perform quite well generally except the $D_5$ and $A_5$ optimum designs which result in low pure error DF. Note that the ideal $h_i$ value is 0.417. For the sake of curiosity we also searched for the design that minimizes the maximum $h_i$ value (maximum loss) and found design 16,
with no pure error DF. In this example we found design 13, the best in terms of criterion $H$, only when using a compound criterion. Even trying the single $H$ criterion on 50,000 initial designs, the algorithm returned an inferior design to this one. Again, by using compound criteria we obtained interesting designs for the experiment.
### Table 8
Properties of designs for the second order model, Example 3 ($q = 4$, $p = 15$ and $n = 36$).

<table>
<thead>
<tr>
<th>Design</th>
<th>Criterion</th>
<th>DF</th>
<th>Efficiencies</th>
<th>$h_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$S_4 + 2S_1 + 4S_0$</td>
<td>(11; 10)</td>
<td>$D_S = 78.82$; $(DP)_S = 70.36$; $A_5 = 75.03$; $(AP)_S = 75.46$; $H = 71.61$</td>
<td>0.636</td>
</tr>
<tr>
<td>2</td>
<td>$S_4 + S_1 + 12S_0$</td>
<td>(11; 10)</td>
<td>$D_S = 67.07$; $(DP)_S = 59.88$; $A_5 = 59.15$; $(AP)_S = 59.50$; $H = 65.09$</td>
<td>0.658</td>
</tr>
<tr>
<td>3</td>
<td>$S_2 + S_1 + 4S_0$</td>
<td>(3; 18)</td>
<td>$D_S = 42.29$; $(DP)_S = 11.86$; $A_5 = 30.21$; $(AP)_S = 34.53$; $H = 84.53$</td>
<td>0.532</td>
</tr>
<tr>
<td>4</td>
<td>$S_1 + 4S_0$</td>
<td>(3; 18)</td>
<td>$D_S = 89.17$; $(DP)_S = 25.02$; $A_5 = 87.00$; $(AP)_S = 41.86$; $H = 97.93$</td>
<td>0.438</td>
</tr>
<tr>
<td>5</td>
<td>$S_4 + \frac{1}{2}S_1 + 4S_0$</td>
<td>(3; 18)</td>
<td>$D_S = 96.78$; $(DP)_S = 27.15$; $A_5 = 92.92$; $(AP)_S = 44.70$; $H = 92.72$</td>
<td>0.668</td>
</tr>
<tr>
<td>6</td>
<td>$S_4 + \frac{1}{2}S_1 + S_1 + 4S_0$</td>
<td>(11; 10)</td>
<td>$D_S = 84.95$; $(DP)_S = 75.84$; $A_5 = 70.97$; $(AP)_S = 71.39$; $H = 90.02$</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>$S_4 + \frac{1}{2}\kappa_{1} S_1 + 4S_0$</td>
<td>(11; 10)</td>
<td>$D_S = 84.78$; $(DP)_S = 75.68$; $A_5 = 70.65$; $(AP)_S = 71.06$; $H = 89.12$</td>
<td>0.570</td>
</tr>
<tr>
<td>8</td>
<td>$\frac{1}{2}S_4 + S_2 + 4S_0$</td>
<td>(3; 18)</td>
<td>$D_S = 71.88$; $(DP)_S = 20.17$; $A_5 = 60.37$; $(AP)_S = 29.04$; $H = 77.53$</td>
<td>0.669</td>
</tr>
<tr>
<td>9</td>
<td>$\frac{1}{2}S_4 + S_2 + 4S_0$</td>
<td>(3; 18)</td>
<td>$D_S = 67.73$; $(DP)_S = 19.00$; $A_5 = 47.72$; $(AP)_S = 22.96$; $H = 87.82$</td>
<td>0.596</td>
</tr>
<tr>
<td>10</td>
<td>$D_5$; $A_5$</td>
<td>(4; 17)</td>
<td>$D_S = 100.00$; $(DP)_S = 41.63$; $A_5 = 100.00$; $(AP)_S = 63.21$; $H = 99.48$</td>
<td>0.517</td>
</tr>
<tr>
<td>11</td>
<td>$(DP)_S$</td>
<td>(18; 3)</td>
<td>$D_S = 93.67$; $(DP)_S = 100.00$; $A_5 = 86.73$; $(AP)_S = 95.74$; $H = 98.87$</td>
<td>0.484</td>
</tr>
<tr>
<td>12</td>
<td>$(AP)_S$</td>
<td>(16; 5)</td>
<td>$D_S = 95.63$; $(DP)_S = 98.51$; $A_5 = 92.23$; $(AP)_S = 100.00$; $H = 98.08$</td>
<td>0.520</td>
</tr>
<tr>
<td>13</td>
<td>$H$</td>
<td>(12; 9)</td>
<td>$D_S = 93.39$; $(DP)_S = 86.58$; $A_5 = 87.30$; $(AP)_S = 89.60$; $H = 100.00$</td>
<td>0.432</td>
</tr>
<tr>
<td>14</td>
<td>$\kappa_1 = \kappa_2 = 0.2$; $\kappa_5 = 0.6$</td>
<td>(16; 5)</td>
<td>$D_S = 95.25$; $(DP)_S = 98.12$; $A_5 = 90.06$; $(AP)_S = 97.65$; $H = 98.65$</td>
<td>0.482</td>
</tr>
<tr>
<td>15</td>
<td>$\kappa_1 = 0.05$; $\kappa_3 = 0.10$; $\kappa_5 = 0.85$</td>
<td>(10; 11)</td>
<td>$D_S = 95.95$; $(DP)_S = 81.89$; $A_5 = 92.95$; $(AP)_S = 91.23$; $H = 99.43$</td>
<td>0.464</td>
</tr>
<tr>
<td>16</td>
<td>$H_{\text{max}}$</td>
<td>(0; 21)</td>
<td>$D_S = 87.69$; $(DP)_S = 80.00$; $A_5 = 83.56$; $(AP)_S = 80.00$; $H = 99.56$</td>
<td>0.432</td>
</tr>
</tbody>
</table>
at random, and improving the minterms of leverages may lead to even better designs. Criterional ones somehow drive the design to local solutions and perhaps, starting with generally better designs, instead of discriminating well between designs. We should highlight that in some examples we were able to find the best showing that it is good practice to consider several properties in the design criterion.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$D_1$: $A_1$: $H$</th>
<th>$(DP)_1$: $(AP)_1$</th>
<th>$\kappa_1 = 0.15; \kappa_5 = 0.85$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
</tr>
<tr>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>2</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>3</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>4</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>5</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>6</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>7</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>8</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>9</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>10</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>11</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>12</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>13</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>14</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>15</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>16</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
</tbody>
</table>

### Table 10
Properties of designs for main effects and two-factor interactions, Example 4 ($q = 4, p = 11, n = 16$ or $n = 24$).

<table>
<thead>
<tr>
<th>Size</th>
<th>Design</th>
<th>Criterion</th>
<th>$DF$</th>
<th>Efficiencies $h_{opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D_1$: $A_1$: $H$</td>
<td>$(PE; LoF)$</td>
<td>$D_1$</td>
<td>$(DP)_1$, $A_1$, $(AP)_1$, $H$</td>
</tr>
<tr>
<td>$n = 16$</td>
<td>1</td>
<td>$D_1$: $A_1$: $H$</td>
<td>(0; 5)</td>
<td>100.00</td>
</tr>
<tr>
<td>2</td>
<td>$(DP)_1$: $(AP)_1$</td>
<td>(5; 0)</td>
<td>76.68</td>
<td>100.00</td>
</tr>
<tr>
<td>3</td>
<td>$\kappa_1 = 0.2; \kappa_3 = 0.1; \kappa_5 = 0.7$</td>
<td>(4; 1)</td>
<td>81.23</td>
<td>84.09</td>
</tr>
<tr>
<td>4</td>
<td>$\kappa_1 = 0.15; \kappa_5 = 0.85$</td>
<td>(2; 3)</td>
<td>89.45</td>
<td>28.48</td>
</tr>
<tr>
<td>$n = 24$</td>
<td>5</td>
<td>$D_1$: $A_1$: $(AP)_1$: $S_4 + \frac{1}{2}S_4$</td>
<td>(8; 5)</td>
<td>100.00</td>
</tr>
<tr>
<td>6</td>
<td>$(DP)_1$</td>
<td>(11; 2)</td>
<td>90.79</td>
<td>100.00</td>
</tr>
<tr>
<td>7</td>
<td>$H$</td>
<td>(12; 1)</td>
<td>87.51</td>
<td>99.89</td>
</tr>
</tbody>
</table>

### 4.4. Example 4

In this example we consider designs with two-level factors, the model including main effects and two-factor interactions. For four factors in $n = 16$ runs, some designs are shown in Table 9. The $D_1$, $A_1$, and $H$-optimum designs are the same, the full factorial. Obviously no pure error estimation is possible from this design and thus we cannot estimate $\sigma^2$ unbiasedly. The $(DP)_1$, $(AP)_1$, $(AP)_5$, and $(DP)_5$-optimum designs are also equivalent allowing 5 degrees of freedom for pure error. All the six points from this design that are not replicated have $h_i = 1$ and thus the design will break down if at least one of these goes missing. Thus the efficiency in terms of Eq. (9) is 0%. A compound criterion involving estimation and leverages produced the design with 4 pure error DF that is 24.28% efficient in terms of (9). This could be increased, if desired, by giving more weight to $H$ as in design 4 (see top part of Table 10).

Increasing $n$ to 24 runs we obtained the designs shown in Table 11, whose properties are given in the lower part of Table 10. In this case, as $n$ is more than twice the number of parameters in the model, the usual criteria give reasonably efficient designs as does the use of the full factorial plus a half replicate. Note that the $H$-optimum design allows more degrees of freedom than the $(DP)_5$-optimum design at the cost of some loss of efficiency for estimating the regression parameters. All compound criteria we tried returned a design equivalent either to the $(DP)_5$ or to the $H$-optimum design, showing that it is good practice to consider several properties in the design criterion.

### 5. Discussion

Robustness of designs to missing observation has been of concern in the planning of experiments. In this paper we propose the inclusion of a property, the $H$ property, in the compound criteria of Gilmour and Trinca (2012), in order to construct optimum designs that will not break down if an observation goes missing. Similar properties have been used by other authors to evaluate the performances of CCDs or subset designs but have not been used to drive the search of an optimum design. We also confirm that the use of the $H$ property as single criterion is not interesting because it does not discriminate well between designs. We should highlight that in some examples we were able to find the best $H$ design only when using a compound criterion, even when trying many thousands of initial random designs. This may indicate this criterion alone somehow drives the design to local solutions and perhaps, starting with generally better designs, instead of at random, and improving them in terms of leverages may lead to even better designs.
Table 11
Possible designs for Example 4: four 2-level factors in $n = 24$ and $p = 11$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$D_5$; $A_5$; $(AP)_5$; $S_4 + 1.5S_4$</th>
<th>$(DP)_5$</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>14</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>16</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>17</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>18</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>19</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>20</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>21</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>22</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>23</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>24</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Although the surrogate measure used in this paper guarantees robustness of the design to just a single missing observation, for reasonably sized designs the approach produces designs robust to three or four missing points. This was shown by a small study carried out for each of the examples and reported in the Supplementary Material, Appendix. For small $n$, missing more than one run is likely to lead to very little information being obtained from the experiment. Of course, the failure of a fairly large proportion of the runs to produce a response might in itself tell the experimenters something important about the system under study.

The overall message of this paper is to reiterate the popular advice to consider many properties of factorial and response surface designs before committing to use one for a particular experiment. The compound criterion used here, including the $H$ criterion will allow experimenters to build designs which are robust to missing or outlying observations. We recommend it for use in practice.

Acknowledgments

The authors are grateful to grant # 2014/01818-0, São Paulo Research Foundation (FAPESP). The first author is grateful to CAPES/MEC/Brazil Government for his scholarship while he was an M.Sc. student in the Biometry Program, Unesp, Botucatu. The authors thank two referees, an associate editor and the editor for helpful comments which led to improvements in the article.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2016.05.023.

References


Ghosh, S., 1982b. Robustness of designs against the unavailability of data. Sankhyā 44 (B), 50–62.
Optimal designs for comparing regression models with correlated observations

Holger Dette *, Kirsten Schorning, Maria Konstantinou

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

A R T I C L E   I N F O

Article history:
Received 23 January 2016
Received in revised form 25 June 2016
Accepted 30 June 2016
Available online 20 July 2016

Keywords:
Linear regression
Correlated observations
Comparing regression curves
Confidence band
Optimal design

A B S T R A C T

The problem under investigation is that of efficient statistical inference for comparing two regression curves estimated from two samples of dependent measurements. Based on a representation of the best pair of linear unbiased estimators in continuous time models as a stochastic integral, a pair of linear unbiased estimators with corresponding optimal designs for finite sample size is constructed. This pair minimises the width of the confidence band for the difference between the estimated curves in a class of linear unbiased estimators approximating the stochastic integrals and is very close to the pair of weighted least squares estimators with corresponding optimal design. Thus results readily available in the literature are extended to the case of correlated observations and an easily implementable solution is provided which is practically non distinguishable from the weighted least squares estimators. The advantages of using the proposed pairs of estimators with corresponding optimal designs for the comparison of regression models are illustrated via several numerical examples.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The issue of comparing two regression models that relate a common response variable to the same covariates for two different groups, respectively, arises frequently in experimental work and particularly in drug development. The conclusions drawn from such comparisons are essential for assessing the non-superiority of one model to the other or for determining whether the difference between the two models is statistically significant or not. In the latter case, the two curves describe practically the same relationship between the response variable and the covariates and the pooled sample can be used to estimate parameters of interest thus improving subsequent inference.

Establishing the similarity of regression curves has also been of great interest in drug dissolution studies for which several methods have been discussed in the literature (see Yuksei et al., 2000; Costa and Sousa Lobo, 2001; Costa, 2001; Costa et al., 2003, among others). Many of the proposed methods in these references are based on measures of similarity such as the Rescigno index or similarity factors. On the more statistical side, various hypothesis tests have been proposed in the literature for assessing the equivalence of two regression curves. Linear and non-linear models with independent observations have been studied, for example, by Liu et al. (2009) and Gsteiger et al. (2011), respectively. Their approach is based on estimating the regression curves in the different samples and constructing confidence bands for the deviation distance between these estimates. More recently, Dette et al. (2015) propose to directly estimate the maximal deviation distance or an $L_2$-distance between the curves and to establish equivalence if the estimator is smaller than a given threshold.

* Correspondence to: Department of Mathematics, Institute of Statistics, Ruhr-Universität Bochum, Universitätsstrasse 150, Building NA 3/73, 44780 Bochum, Germany. Fax: +49 0 234 32 14559.
E-mail address: holger.dette@ruhr-uni-bochum.de (H. Dette).

http://dx.doi.org/10.1016/j.csda.2016.06.017
0167-9473/© 2016 Elsevier B.V. All rights reserved.
On the other hand, the efficient planning of experiments for comparing curves has not been dealt with in the literature although this would substantially improve the accuracy of the conclusions drawn regarding non-superiority or equivalence. To the best of the authors’ knowledge only Dette and Schorning (2016) investigate such a design problem. They consider regression models with independent observations and search for designs that minimise the width of the simultaneous confidence band proposed by Gsteiger et al. (2011), for the difference of the two estimated models. More precisely, an optimal pair of designs minimises an $L_p$-norm of the asymptotic variance of the difference between the two estimated curves. The $L_p$-norm is calculated in the common covariate region of interest and the regression curves are estimated via maximum likelihood. Dette and Schorning (2016) provide explicit solutions for some commonly used dose–response models and demonstrate that using an optimal pair of designs, such as the one they propose, instead of a “standard design” results in the width of the confidence band to be reduced by more than 50%. Although this improvement is impressive, the results of Dette and Schorning (2016) cannot be used, for example, to improve the statistical inference for the comparison of dissolution profiles since in such studies measurements are usually taken at the same patient and therefore cannot be considered as uncorrelated.

The goal of the present paper is to develop efficient statistical tools (estimation and design) for the comparison of two regression curves estimated from two samples of correlated measurements. The estimation of the parameters can easily be done by (weighted) least squares as in the uncorrelated case. However, it is well known that the construction of optimal designs for such estimators in the case of dependent data is a rather challenging problem since classical tools of convex optimisation theory are not applicable. Solutions of exact optimal design problem are only available for specific linear models (see Dette et al., 2008; KiselaK and Stehlík, 2008; Harman and Štulajter, 2010) and most of the literature is focused on asymptotic optimal designs without avoiding however, the issue of non-convex optimisation (see, for example, Sacks and Ylvisaker, 1968; Bickel and Herzberg, 1979; Pázman and Müller, 2001; Müller and Pázman, 2003; Näther, 1985; Zhigljavsky et al., 2010, among others). As a consequence, optimal designs have mainly been determined for a few one-parameter linear models.

Only recently, Dette et al. (2016, 2015) have made substantial progress towards overcoming the issue of non-convex optimisation. In contrast to the prevailing design of experiments practice, according to which an optimal design is constructed for a particular estimator, these authors propose to optimise the estimator and the design simultaneously. Their strategy yields new estimators with corresponding optimal designs which are very close to the best linear unbiased estimator for finite sample size, that is, the weighted least squares estimator (WLSE) with corresponding optimal design. Note that the latter designs require non-convex optimisation to be determined. Dette et al. (2015) improved the accuracy of the proposed class of estimators and optimal designs using a new representation of the best linear unbiased estimator in the continuous time model as a stochastic integral. They thus provide an improved solution practically non-distinguishable from the WLSE (with corresponding optimal design), which is also easier to implement and applicable to a broad class of linear regression models with various covariance kernels.

The aim is to fill in the gap in the literature regarding efficient planning of experiments for comparing regression models with dependent observations. In Section 2 the practical problem under investigation is introduced and the initial basis for the comparison of curves in the case of dependent observations is provided. The framework for continuous-time models is set-up in Section 3 where the best pair of linear unbiased estimators for comparing such models with dependent error processes is obtained using stochastic integrals. Finally, in Section 4 the corresponding best pair of estimators with optimal designs in a class of linear unbiased estimators approximating the stochastic integrals is derived for finite sample sizes. Thus the question of how to plan experiments for comparing curves with dependent observations is answered. Several numerical examples are discussed in Section 5 via which the benefits of the proposed methodology are demonstrated.

2. Comparing parametric curves

The practical scenario of two groups is considered where the dependence between the response and the covariates in these groups is described by two linear regression models with dependent observations given by

$$Y_i(t_{ij}) = f^1_i(t_{ij})\theta_1 + \epsilon_i(t_{ij}), \quad j = 1, \ldots, n_i; \quad i = 1, 2. \tag{2.1}$$

In each group a total of $n_i$ observations are taken at the time-points $t_{i1}, \ldots, t_{in_i}$ varying in a compact interval $[a, b]$ and $\{\epsilon_i(t) : t \in [a, b]\}$ are two independent Gaussian processes with $\mathbb{E}[\epsilon_i(t_{ij})] = 0$ and covariance kernels $K_i(t_{ik}, t_{il}) = \mathbb{E}[\epsilon_i(t_{ik})\epsilon_i(t_{il})]$ denoting the covariance between observations at the points $t_{ik}$ and $t_{il}$ ($i = 1, 2, k, l = 1, \ldots, n_i$). The vectors of unknown parameters $\theta_1$ and $\theta_2$ are assumed to be $m_1$- and $m_2$-dimensional respectively and the corresponding

$$f_i(t) = (f_{i1}(t), \ldots, f_{im_i}(t))^T, \quad i = 1, 2,$$

are vectors of continuously differentiable linearly independent regression functions.

Let $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ be a pair of linear unbiased estimators for each of the two models. Then each estimator $\hat{\theta}_i$ is normally distributed with $\mathbb{E}[\hat{\theta}_i] = \theta_i$ and covariance matrix $\text{Var}(\hat{\theta}_i) = \Sigma_i$ ($i = 1, 2$). Moreover, the prediction for the difference at the time-point $t$ satisfies

$$f_1^T(t)\hat{\theta}_1 - f_2^T(t)\hat{\theta}_2 - (f_1^T(t)\theta_1 - f_2^T(t)\theta_2) \sim \mathcal{N}(0, g(t, \hat{\theta})).$$
where the function \( g \) is defined by
\[
g(t, \hat{\theta}) = \text{Var}(f_1^T(t)\hat{\theta}_1 - f_2^T(t)\hat{\theta}_2) = f_1^T(t)\text{Var}(\hat{\theta}_1)f_1(t) + f_2^T(t)\text{Var}(\hat{\theta}_2)f_2(t) = f_1^T(t)\Sigma_1f_1(t) + f_2^T(t)\Sigma_2f_2(t). \tag{2.2}
\]
This result along with the results of Gsteiger et al. (2011) are used to obtain a simultaneous confidence band for the difference of the two curves. More precisely, if \([a, b]\) is the range where the two curves should be compared on, the confidence band is defined by
\[
\hat{t} = \sup_{t \in [a, b]} \frac{|f_1^T(t)\hat{\theta}_1 - f_2^T(t)\hat{\theta}_2 - (f_1^T(t)\theta_1 - f_2^T(t)\theta_2)|}{|f_1^T(t)\Sigma_1f_1(t) + f_2^T(t)\Sigma_2f_2(t)|^{1/2}} \leq D, \tag{2.3}
\]
where the constant \( D \) is chosen, such that \( P(\hat{t} \leq D) \approx 1 - \alpha \). Note that Gsteiger et al. (2011) propose the parametric bootstrap for choosing the critical value \( D \). Consequently, "good" estimators with corresponding "good" time-points \( t_1, \ldots, t_n \) should make the width of this band as small as possible at each \( t \in [a, b] \). This corresponds to a simultaneous minimisation of the variance in (2.2) with respect to the choice of the linear unbiased estimators and the time-points. As pointed out by Dette and Schorning (2016) this is only possible in rare circumstances and thus they propose to minimise an \( L_p \)-norm of the function \( g(t, \hat{\theta}) \) as a design criterion, that is
\[
\|g(\cdot, \hat{\theta})\|_p := \left( \int_a^b |g(t, \hat{\theta})|^p \, dt \right)^{1/p}, \quad 1 \leq p \leq \infty, \tag{2.4}
\]
where the case \( p = \infty \) corresponds to the maximal deviation \( \|g(\cdot, \hat{\theta})\|_{\infty} = \sup_{t \in [a, b]} |g(t, \hat{\theta})| \).

3. Comparison in continuous time models

Dette et al. (2016) showed that the optimal design problem in regression models with triangular covariance structures can be reduced to a design problem in regression models with a Brownian motion as correlation structure (see also Remark 3.1). Therefore, linear regression models in their continuous time version, that is,
\[
Y_i(t) = f_1^T(t)\theta_1 + \varepsilon_i(t), \quad t \in [a, b]; \quad i = 1, 2, \tag{3.1}
\]
are examined first, where \( 0 < a < b \) and the full trajectory of both processes \( Y_i = \{Y_i(t) | t \in [a, b]\} \) can be observed. The error processes \( \varepsilon_i = \{\varepsilon_i(t) | t \in [a, b]\} \) are independent Brownian motions with continuous covariance kernels given by \( K_i(t, s) = t \wedge s \). Furthermore, the processes \( Y_1 \) and \( Y_2 \) are assumed to be independent.

Following the discussion in Section 2, a "good pair" of estimators should make the \( L_p \)-norm of the variance given in (2.2) as small as possible. Therefore, the best pair of estimators is found by minimising
\[
\mu_p(\hat{\theta}) = \|\text{Var}(f_1^T(t)\hat{\theta}_1 - f_2^T(t)\hat{\theta}_2)\|_p, \quad p \in [1, \infty], \tag{3.2}
\]
with respect to all pairs of linear unbiased estimators. Recall that for \( p = \infty \), the criterion is \( \mu_{\infty}(\hat{\theta}) = \sup_{t \in [a, b]} \text{Var}(f_1^T(t)\hat{\theta}_1 - f_2^T(t)\hat{\theta}_2) \).

Dette et al. (2015) showed that the best linear unbiased estimators in the individual continuous time models are given by
\[
\hat{\theta}_{\text{BLUE}} = C_i^{-1} \left( \int_a^b \hat{f}_i(t) dY_i(t) + \frac{\hat{f}_i(a)}{a} Y_i(a) \right), \quad i = 1, 2, \tag{3.3}
\]
where
\[
C_i^{-1} = \text{Var}(\hat{\theta}_{\text{BLUE}}) = \left( \int_a^b \hat{f}_i(t)\hat{f}_i^T(t) \, dt + \frac{f_i(a)\hat{f}_i^T(a)}{a} \right)^{-1}, \quad i = 1, 2. \tag{3.4}
\]
These estimators have minimum variance with respect to the Loewner ordering, that is,
\[
C_i^{-1} = \text{Var}(\hat{\theta}_{\text{BLUE}}) \leq \text{Var}(\hat{\theta}_i), \quad i = 1, 2,
\]
in the Loewner ordering, for any linear unbiased estimator \( \hat{\theta}_i \) in model (3.1) \((i = 1, 2)\). Theorem 3.1 shows that the best pair of linear unbiased estimators is the pair \((\hat{\theta}_{\text{BLUE}}, \hat{\theta}_{\text{BLUE}})\) of the best linear unbiased estimators in the individual models. For its proof the following lemma is established first which is given as exercise 14 of Chapter 4.3 in Borwein and Lewis (2000).

**Lemma 3.1.** Let \( g : [a, b] \times L \to \mathbb{R} \) be a function on the non-empty set \( L \subset \mathbb{R}^{m_1 + m_2} \). If the point \((\bar{t}, \bar{\theta}) \in [a, b] \times L \) is a saddlepoint, that is
\[
g(t, \hat{\theta}) \leq g(\bar{t}, \bar{\theta}) \leq g(t, \hat{\theta}) \quad \text{for all} \quad t \in [a, b], \hat{\theta} \in L, \tag{3.5}
\]
the following equalities hold
\[
\inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta) = g(\tilde{t}, \tilde{\theta}) = \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta). \tag{3.6}
\]

**Proof of Lemma 3.1.** Note that \((\tilde{t}, \tilde{\theta}) \in [a, b] \times \mathcal{L}\) is a saddlepoint if and only if
\[
g(\tilde{t}, \tilde{\theta}) = \inf_{\theta \in \mathcal{L}} g(\tilde{t}, \theta) \quad \text{and} \quad g(\tilde{t}, \tilde{\theta}) = \sup_{t \in [a, b]} g(t, \tilde{\theta}).
\]

Using this formulation \(g(\tilde{t}, \tilde{\theta}) = \sup_{t \in [a, b]} g(t, \tilde{\theta}) \geq \inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta)\) and also \(g(\tilde{t}, \tilde{\theta}) = \inf_{\theta \in \mathcal{L}} g(\tilde{t}, \theta) \leq \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta)\). Hence
\[
\inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta) \leq \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta). \tag{3.7}
\]
On the other hand, \(g(t, \theta) \leq \sup_{t \in [a, b]} g(t, \tilde{\theta})\) for all \(t \in [a, b]\) and \(\theta \in \mathcal{L}\) and thus
\[
\inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta) \leq \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta) \leq \inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta).
\]
Hence
\[
\inf_{\theta \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \theta) \geq \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta). \tag{3.8}
\]
and by combining (3.7) and (3.8), the equality (3.6) follows. \(\square\)

**Theorem 3.1.** Let \(\hat{\theta}_{\text{BLUE}}\) be the best linear unbiased estimator, defined in (3.3) and (3.4), in the corresponding continuous time model (3.1) for \(i = 1, 2\). Then for any \(p \in [1, \infty]\), \(\hat{\theta}_{\text{BLUE}} = (\hat{\theta}_{\text{BLUE}}^1, \hat{\theta}_{\text{BLUE}}^2)^T\) is the best pair of linear unbiased estimators minimising the \(L_p\)-norm (3.2) of the variance of the estimate of the difference between the parametric curves in the class
\[
\mathcal{L} = \{\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)^T : \hat{\theta}^T t \text{ linear unbiased estimator for model (3.1) for } i = 1, 2\}.
\]

**Proof of Theorem 3.1.** Since both \(\hat{\theta}_{\text{BLUE}}^1\) and \(\hat{\theta}_{\text{BLUE}}^2\) have minimal variance with respect to the Loewner ordering, it follows that \(\hat{\theta}_{\text{BLUE}} = (\hat{\theta}_{\text{BLUE}}^1, \hat{\theta}_{\text{BLUE}}^2)^T\) minimises the variance of the difference between the estimated curves. That is, for any \(t \in [a, b]\)
\[
g(t, \hat{\theta}_{\text{BLUE}}) \leq g(t, \hat{\theta}_{\text{BLUE}}), \quad \text{for all } \hat{\theta} \in \mathcal{L}.
\]
When \(p \in [1, \infty]\), it is straightforward to check that \(\hat{\theta}_{\text{BLUE}}\) minimises \(\mu_p(\hat{\theta}) = \|g(t, \hat{\theta})\|_p\) using the fact that the \(L_p\)-norm is an increasing function. Now let \(p = \infty\) and also let \(\tilde{t} \in \arg \sup_{t \in [a, b]} g(t, \hat{\theta}_{\text{BLUE}})\). It follows from the definition of \(\tilde{t}\) that
\[
g(t, \hat{\theta}_{\text{BLUE}}) \leq g(\tilde{t}, \hat{\theta}_{\text{BLUE}}), \quad \text{for all } t \in [a, b].\]
Therefore,
\[
g(t, \hat{\theta}_{\text{BLUE}}) \leq g(\tilde{t}, \hat{\theta}_{\text{BLUE}}) \leq g(\tilde{t}, \hat{\theta}), \quad \text{for all } t \in [a, b] \text{ and } \hat{\theta} \in \mathcal{L}.
\]
This means that \((\tilde{t}, \hat{\theta}_{\text{BLUE}}) \in [a, b] \times \mathcal{L}\) is a saddlepoint and using Lemma 3.1 it is obtained that
\[
\inf_{\hat{\theta} \in \mathcal{L}} \sup_{t \in [a, b]} g(t, \hat{\theta}) = g(\tilde{t}, \hat{\theta}_{\text{BLUE}}) = \sup_{t \in [a, b]} \inf_{\theta \in \mathcal{L}} g(t, \theta).
\]
Thus \(\hat{\theta}_{\text{BLUE}}\) minimises \(\mu_{\infty}(\hat{\theta})\) in the class \(\mathcal{L}\) of pairs of linear unbiased estimators. \(\square\)

**Remark 3.1.** Brownian motion is a special case of the general class of triangular kernels which are of the form
\[
K_i(t, t') = u_i(t) v_i(t'), \quad \text{for } t \leq t': i = 1, 2,
\]
for each group and the simple kernel \(K_0(t, t') = t \wedge t' = \min\{t, t'\}\) corresponding to the Brownian motion is obtained by choosing \(u_i(t) = t\) and \(v_i(t) = 1\). Using the results of Dette et al. (2016, 2015) showed that a representation of the BLUE as a stochastic integral can be obtained for any continuous time model of the form (3.1) with a general triangular kernel. This is achieved by means of a simple non-linear transformation that reduces the model with triangular covariance kernel to a different continuous time model with Brownian motion as an error process. In particular, any model of the form (3.1) with \(\varepsilon_i(t)\) having a general triangular covariance kernel is equivalent to
\[
\tilde{Y}_i(t) = \tilde{f}_i^T (\tilde{t}) \theta + \tilde{\varepsilon}_i(\tilde{t}), \quad \tilde{t} \in [\tilde{a}, \tilde{b}]; \quad i = 1, 2, \tag{3.9}
\]
where \(\tilde{\varepsilon}_i(\tilde{t}) = \varepsilon_i(t)/v_i(t), i = 1, 2\), are Brownian motions on the interval \([\tilde{a}, \tilde{b}]\) and
\[
\tilde{t} = q_i(t) := \frac{u_i(t)}{v_i(t)}, \quad \tilde{f}_i(\tilde{t}) = \frac{f_i(t)}{v_i(t)} \quad \text{and} \quad \tilde{Y}_i(\tilde{t}) = \frac{Y_i(t)}{v_i(t)}, \quad i = 1, 2.
\]
Hence the \( \hat{\theta}_i, \text{BLUE} \) for any continuous time model (3.1) with a general triangular covariance kernel can be obtained from the \( \hat{\theta}_i, \text{BLUE} \) in the corresponding model (3.9) by the transformation \( t = q(t) \). Therefore, although throughout the theoretical part the focus is on the covariance kernel of the Brownian motion, the presented methodology is valid for all triangular kernels. Some examples of kernels other than that of the Brownian motion are given in Section 5 where the optimal designs are found for the transformed models (3.9) with Brownian motion as error processes and then the design points are transformed back to the original design space \([a, b]\) via \( t = q^{-1}(t) \).

### 4. Optimal inference for comparing curves

Using the results of Section 3 for continuous time models, the initial problem of comparing two regression curves estimated from two samples of dependent measurements, defined in (2.1), is considered. Following the discussion in Remark 3.1, the correlation structure in each group is assumed to be \( \text{Cov}(Y_i(t_{ij}), Y_i(t_{ik})) = 0 \) for all \( i \), which corresponds to the case of a Brownian motion. Let \( n = n_1 + n_2 \) denote the total sample size and define \( Y_i = (Y_i(t_{i,1}), \ldots, Y_i(t_{i,n})^T \) as the vector of observations in group \( i \). The corresponding weighted least squares estimator of \( \theta_i \) is defined as

\[
\hat{\theta}_{i, \text{WLS}} = (X_i^T \Sigma_i^{-1} X_i)^{-1} X_i^T \Sigma_i^{-1} Y_i, \quad i = 1, 2,
\]

where \( X_i = (f_i(t_{ij}))_{j=1}^{n_i} \) is the \( n_i \times m_i \) design matrix and the \( n_i \times n_i \) matrix \( \Sigma_i = (K_i(t_{ij}, t_{ik}))_{j,k} = K_i(t_{ij}, t_{ik}) \) is the variance-covariance matrix (\( i = 1, 2 \)). It is well known that \( \hat{\theta}_{i, \text{WLS}} \) is the BLUE in model (2.1) for \( i = 1, 2 \) and the corresponding minimal variance is given by

\[
\text{Var}(\hat{\theta}_{i, \text{WLS}}) = (X_i^T \Sigma_i^{-1} X_i)^{-1}, \quad i = 1, 2.
\]

As pointed out in the introduction the minimisation of a real-valued functional of this matrix is a demanding non-convex discrete optimisation problem and thus analytical and numerical results are rather scarce in the literature.

An alternative to the weighted least squares estimator is proposed by Dette et al. (2015) who use an approximation of the stochastic integral involved in (3.3) and require the resulting estimator to be unbiased. Following this approach, a numerical approximation of the stochastic integral in (3.3) is constructed and the estimators

\[
\hat{\theta}_{i,n} = C_i^{-1} \left\{ \sum_{j=1}^{n_i} \omega_{ij} f_i(t_{ij}) \left( \hat{\theta}_{i,n} - f_i(T_{ij}) \right) \right\}, \quad i = 1, 2,
\]

(4.1)

are considered for each of the two regression models (2.1), where \( C_i^{-1} \) is defined in (3.4) and \( \omega_{ij} \) are weight-matrices. Here \( a = t_{i,1} < t_{i,2} < \cdots < t_{i,n_i-1} < t_{i,n_i} = b, \quad i = 1, 2, \) are design points on the time interval \([a, b]\) and \( \omega_{ij} = \Omega_i(t_{ij}, t_{ij-1}) \) are the corresponding \( m_i \)-dimensional weight-vectors of both of which have to be determined in an optimal way. Both estimators \( \hat{\theta}_{i,n} \) are further conditioned to being unbiased. It is shown in Dette et al. (2015) that unbiasedness is equivalent to the condition

\[
M_i := \int_a^b \hat{f}_i(t)(t) \, dt = \sum_{j=1}^{n_i} \omega_{ij} \left( f_i(t_{ij}) - f_i(T_{ij}) \right), \quad i = 1, 2,
\]

(4.2)

under which \( \text{Var}(\hat{\theta}_{i,n}) = E[(\hat{\theta}_{i,n} - \hat{\theta}_{i, \text{BLUE}})(\hat{\theta}_{i,n} - \hat{\theta}_{i, \text{BLUE}})^T] + C_i^{-1} \) (see Theorem 3.1 in Dette et al., 2015).

For a pair \( \hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})^T \) of linear unbiased estimators of the form (4.1) the variance of the estimator \( f_1(t)\hat{\theta}_{1,n} - f_2(t)\hat{\theta}_{2,n} \) for the difference of the curves \( f_1(t)\hat{\theta}_1 - f_2(t)\hat{\theta}_2 \) at a specified time-point \( t \in [a, b] \) is given by

\[
g_n(t, \hat{\theta}_n) = \text{Var}(f_1(t)\hat{\theta}_{1,n} - f_2(t)\hat{\theta}_{2,n}) = \sum_{i=1}^{2} f_i(t) \text{Var}(\hat{\theta}_{i,n} f_i(t)) = \sum_{i=1}^{2} f_i(t) \left\{ \left( C_i^{-1} E \left[ \int_{t_{ij-1}}^{t_{ij}} \omega_{ij}(s) \, ds \right] \right) C_i^{-1} + C_i^{-1} \right\} f_i(t)
\]

\[
= \sum_{i=1}^{2} f_i(t) \left\{ C_i^{-1} E \left[ \int_{t_{ij-1}}^{t_{ij}} \omega_{ij}(s) \, ds \right] \right\} C_i^{-1} + C_i^{-1} \right\} f_i(t)
\]

\[
= \sum_{i=1}^{2} f_i(t) \left\{ -C_i^{-1} M_i + \sum_{j=1}^{n_i} (t_{ij} - t_{ij-1}) C_i^{-1} \omega_{ij} \omega_{ij} C_i^{-1} + C_i^{-1} \right\} f_i(t),
\]

(4.3)
where Itô’s formula is used for the fourth equality. Thus, it is required to choose the \( n - 4 \) design points \( t_{1,2}, \ldots, t_{1,n_1-1}, t_{2,2}, \ldots, t_{2,n_2-1} \) and the \( n - 2 \) weight-vectors \( \omega_1,2, \ldots, \omega_{1,n_1}, \omega_{2,2}, \ldots, \omega_{2,n_2} \) such that the \( L_p \)-norm of (4.3) is minimised in the class

\[
\mathcal{L}_n = \{ \theta_n = (\hat{\theta}_{1,n_1}^T, \hat{\theta}_{2,n_2}^T)^T \mid \theta_{i,n_i} \text{ is of the form (4.1)} \text{ and satisfies (4.2)} \ i = 1, 2 \},
\]

of all pairs of linear unbiased estimators of the form (4.1). Now let

\[
\gamma_{i,j} = \frac{\omega_{i,j}}{\sqrt{t_{i,j} - t_{i,j-1}}}, \quad j = 1, \ldots, n_i; i = 1, 2,
\]

and \( \Gamma = (\Gamma_{1}^T, \Gamma_{2}^T)^T = (\gamma_{1,1}^T, \ldots, \gamma_{1,n_1}^T, \gamma_{2,1}^T, \ldots, \gamma_{2,n_2}^T) \in \mathbb{R}^{m_1(n_1-1) + m_2(n_2-1)}. \) Then the variance \( g_n(t, \hat{\theta}_n) \) can be rewritten as

\[
\varphi_n(t, \Gamma) = \sum_{i=1}^{2} \int_{t_{i,j-1}}^{t_{i,j}} \left( \int_{t_{i,j-1}}^{t_{i,j}} \varphi_n(t, \Gamma) \right)^{\frac{1}{p}} dt^{-1}, \quad p \in [1, \infty), \]

is minimised. Note that \( \mu_{p,n} \) is convex in \( \Gamma \) for all \( p \in [1, \infty). \)

Similar to the case of continuous time models, the optimal pair \( \Gamma^* \) is the pair of optimal \( \Gamma_i^* \)'s which minimise the variance of the corresponding approximate estimator \( \hat{\theta}_{i,n_i} \), given in Theorem 3.2 in Dette et al. (2015) \( (i = 1, 2). \)

**Theorem 4.1.** Assume that each of the \( m_1 \times m_i \) matrices

\[
B_i = \sum_{j=2}^{n_i} \left[ \frac{f_i(t_{i,j}) - f_i(t_{i,j-1})} {t_{i,j} - t_{i,j-1}} \right] I_{i,j}, \quad i = 1, 2,
\]

is non-singular. Then the optimal \( \Gamma^* = (\Gamma_{1}^*), (\Gamma_{2}^*) \) minimising \( \mu_{p,n}(\Gamma) \), is given by the components

\[
\gamma_{i,j}^* = M_i B_i^{-1} \frac{f_i(t_{i,j}) - f_i(t_{i,j-1})} {\sqrt{t_{i,j} - t_{i,j-1}}}, \quad j = 1, \ldots, n_i; i = 1, 2.
\]

Moreover, the pair \( \hat{\theta}_{i,n_i} = (\hat{\theta}_{1,n_1}^*, \hat{\theta}_{2,n_2}^*) \) defined in (4.1) with weight-vectors given by

\[
\omega_{i,j}^* = M_i B_i^{-1} \frac{f_i(t_{i,j}) - f_i(t_{i,j-1})} {t_{i,j} - t_{i,j-1}}, \quad j = 1, \ldots, n_i; i = 1, 2,
\]

minimises the \( L_p \)-norm of the function \( g_n(t, \hat{\theta}_n) \) defined in (4.3) in the class \( \mathcal{L}_n \) defined in (4.4), with respect to \( \omega_{1,2}, \ldots, \omega_{1,n_1}, \omega_{2,1}, \ldots, \omega_{2,n_2}. \)

**Proof of Theorem 4.1.** Using similar arguments as in the proof of Theorem 3.2 in Dette et al. (2015) it can be shown that for any \( v \in \mathbb{R}^{m_i} \setminus \{0\}, i = 1, 2, \) each of the \( \Gamma_i^* \)'s, with components defined in (4.7), minimises the function

\[
v^T \vartheta(\hat{\theta}_{i,n_i}) v = v^T \left[ -C_i^{-1} M_i C_i^{-1} + C_i^{-1} \sum_{j=2}^{n_i} \gamma_{i,j} Y_{i,j}^T C_i^{-1} \right] v,
\]

subject to the constraint of unbiasedness given in (4.2). Therefore, the pair \( \Gamma^* = (\Gamma_{1}^*), (\Gamma_{2}^*) \) is Loewner optimal and it remains to show that \( \Gamma^* \) minimises the \( L_p \)-norm \( \mu_{p,n}(\Gamma) \) for any \( p \in [1, \infty]. \) The proof of this statement follows along the same lines of the proof of Theorem 3.1 and it is therefore omitted. \( \square \)

Inserting the optimal weights \( \gamma_{i,j}^* \) in the function given in (4.5) and using one of the functionals in (4.6) gives an optimality criterion which is finally minimised with respect to the choice of the design points \( a = t_{i,1} < t_{i,2} < \cdots < t_{i,n_i} = b. \) For example, if \( p = \infty \) the resulting criterion is given by

\[
\Phi_{\infty}(\{t_{i,j} \mid i = 1, \ldots, n_i; i = 1, 2\}) = \sup_{t \in [a,b]} \left\{ \sum_{i=1}^{2} \int_{t_{i,j-1}}^{t_{i,j}} \left( \int_{t_{i,j-1}}^{t_{i,j}} \varphi_n(t, \Gamma) \right)^{\frac{1}{p}} dt^{-1} \right\} C_i^{-1} f_i(t).
\]

Finally, the optimal points \( a = t_{i,1}^* < t_{i,2}^* < \cdots < t_{i,n_i}^* = b \) (minimising (4.9)) and the corresponding weights \( \omega_{i,j}^* \) defined by (4.8) provide the optimal linear unbiased estimator of the form (4.1) with corresponding optimal design. In the following section, the optimal points are determined using the heuristic algorithm Particle Swarm Optimisation algorithm.
5. Numerical examples

The advantages of the proposed pair of estimators (with corresponding optimal designs) are illustrated via several model, triangular covariance kernel and optimality criteria examples. A case study is also considered thus portraying the effectiveness of the suggested methodology when implemented to a real-life application.

At the design stage of an experiment the sample sizes $n_1$ and $n_2$ are known and the choice of these numbers of observation points in each group depends on the concrete application. For the considered problem of comparing two regression curves, one requires at least $n_i = m_i + 1$ observation points in each group, where $m_i$ ($i = 1, 2$) is the number of unknown parameters in each of the two models. In fact, according to the particular circumstances of each experiment (for example, costs, duration of the experiment), $n_1$ and $n_2$ will usually be close to these minimal numbers of required observations. Therefore, in what is to follow the values of $n_1$ and $n_2$ are chosen to be small, in accordance with the dimensions of the parameter vectors. Consequently the resulting design points are optimal for the case of small values of $n_1$ and $n_2$ which is of greater practical importance. Note that as $n_1$ and $n_2$ increase and thus move away from the minimal numbers of observation points, the corresponding optimal design approaches the equidistant design with equally spaced design points which eventually becomes optimal. Thus, the latter design (with optimal weights) can be used without detriment in the case of large values of $n_1$ and $n_2$ (with respect to the number of parameters in each model).

Regarding the calculation of the optimal design points, the heuristic algorithm Particle Swarm Optimisation (PSO) algorithm is used throughout (see for example Clerc, 2006). More specifically, after inserting the optimal weights given in Eq. (4.8) in the desired optimality criterion function specified by (4.6) and (4.5), the optimal design points are obtained by minimising the resulting optimality criterion using the PSO algorithm. For the case of $p = \infty$, the resulting optimality criterion is given in Eq. (4.9). The benefits of using the proposed estimators with corresponding optimal weights and design points are illustrated by means of the confidence bands proposed by Gsteiger et al. (2011) for the difference between the two estimated regression curves. These confidence bands are calculated as the averages of uniform confidence bands calculated by 100 simulations.

5.1. Equal sample sizes

The two regression curves

\[ Y_1 = Y_1(t) = f_1^T(t)\theta_1 + \varepsilon_1(t) = (\sin t, \cos t)^T\theta_1 + \varepsilon_1(t), \]

(5.1)

and

\[ Y_2 = Y_2(t) = f_2^T(t)\theta_2 + \varepsilon_2(t) = (\sin t, \cos t, \sin 2t, \cos 2t)^T\theta_2 + \varepsilon_2(t), \]

(5.2)

are to be used for the comparisons on the design space $[a, b] = [1, 10]$, and the cases of a Brownian motion and an exponential covariance kernel of the form $K(t, t') = \exp(-\lambda|t - t'|)$ for both error processes $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are studied separately. A common (and small) sample size is considered for both groups and in particular, the values $n_1 = n_2 = 5$ are used due to the fact that the dimension of the parameter vector for model (5.2) is $m_2 = 4$. Furthermore, the $\mu_\infty$-optimality criterion defined in (4.6) is examined. As pointed out by Dette and Schorning (2016), this criterion is probably of most practical interest as it directly refers to the maximum width of the confidence band. Note also that unlike the $\mu_p$-criteria for $p < \infty$, the $\mu_\infty$-optimality criterion is not necessarily differentiable.

Denote by $\hat{\theta}^*_n = (\hat{\theta}^*_1, \hat{\theta}^*_2, \hat{\theta}^*_3, \hat{\theta}^*_4)$ the best pair of linear unbiased estimators in the class (4.4) as determined in Theorem 4.1. Also define the equidistant design on $[1, 10]$ as the design with equally spaced points, that is, $[1, 3.25, 5.50, 7.75, 10]$, and optimal weights as in (4.8). Table 1 presents the $\mu_\infty$-optimal design points where observations should be taken in each group along with the $\mu_\infty$-optimality criterion values when either the $\mu_\infty$-optimal or equidistant design (in brackets) is used. All three combinations of comparisons using models (5.1) and (5.2) are studied separately and the error processes for both models are assumed to be Brownian motions. The corresponding design points and criterion values for the cases of both error processes having the same exponential kernel with $\lambda = 0.5$ and $\lambda = 1$ are given in Tables 2 and 3 respectively.

For all the covariance kernels considered, it is evident that the $\mu_\infty$-optimality criterion value decreases substantially when the $\mu_\infty$-optimal instead of the equally spaced design points are used. When model (5.1) is considered for both groups (first rows of Tables 1–3), this reduction is not as significant. This is probably due to the small dimension of the parameter vector in model (5.1) ($m_1 = 2$). Note that Dette et al. (2015) showed that for one-parameter models, equally spaced design points provide already an efficient allocation for each of the $\hat{\theta}^*_i, i = 1, 2$, given that the weights are chosen in an optimal way and that the derivative of the regression is not too large.

To further investigate the gains of using the proposed pair of estimators with corresponding optimal weights and design points, the uniform confidence bands proposed by Gsteiger et al. (2011) are calculated. Fig. 1 presents these confidence bands for the case of both error processes being Brownian motions. In the left panel the results for the proposed pair of estimators are shown, while the right panel gives the confidence bands obtained from the pair of weighted least squares estimators with corresponding optimal design (also determined by the PSO algorithm). Note again that the latter designs are difficult to determine because of the non-convex structure of the optimal design problem. Assuming Brownian motion for both error processes, the first and last row of graphs in Fig. 1 correspond to the cases where models (5.1) and (5.2) respectively are
As before, the maximal width of the confidence band decreases considerably when the models having the same exponential covariance kernel with \( \lambda = 0.5 \) are compared (first row of graphs), whereas the vectors of parameter values \((1, 1, 1)^T\) and \((1, 2, 1, 2)^T\) were used for the last row of graphs. The middle row of graphs corresponds to the direct comparison of the two models under consideration with \( \theta_1 = (1, 1)^T \) and \( \theta_2 = (1, 1, 1, 1)^T \) and assuming again Brownian motion for both error processes. In each graph, the confidence bands from the \( \mu_{\infty} \)-optimal and the equidistant design are plotted separately using the solid and dashed lines respectively, along with the plot for the true difference \( f_1(t)\theta_1 - f_2(t)\theta_2 \) (dotted lines).

The conclusions drawn from the graphs of Fig. 1 are in agreement with the observations from Table 1. In particular, when two models of the form (5.1) are compared (first row of graphs), the \( \mu_{\infty} \)-optimal and equidistant designs perform similarly and thus the latter design (with optimal weights) can be used without detriment. However, from the second and third row of graphs it can be observed that regardless of the estimator, if the \( \mu_{\infty} \)-optimal design is used instead of the equidistant design the maximal width of the confidence band is reduced substantially. Therefore, the use of the proposed \( \mu_{\infty} \)-optimal design, at least up to the optimal weights, does improve inference by significantly reducing the maximum variance of the difference of the two estimated regression curves.

Furthermore, a comparison of the left and right panels of Fig. 1 reveals that the proposed pair of estimators (with corresponding optimal design) produces similar results to those for the pair of weighted least squares estimators (with corresponding optimal design). Hence the recommended methodology provides an easily implemented solution which is practically non distinguishable from the solution for the weighted least squares estimators which requires non-convex optimisation. The advantages of the methodology are also illustrated in Fig. 2 for the cases of the error processes of both models having the same exponential covariance kernel \( K(t, t') = \exp(-\lambda|t - t'|) \) with \( \lambda = 0.5 \) (left panel) and \( \lambda = 1 \) (right panel). As before, the maximal width of the confidence band decreases considerably when the \( \mu_{\infty} \)-optimal design is used.

### Table 1

\( \mu_{\infty} \)-optimal design points for \( \hat{\mu}_0 \) of five observations in each group in the interval \([1, 10]\) and \( \mu_{\infty} \)-optimality criterion values for the \( \mu_{\infty} \)-optimal and equidistant (in brackets) designs. Both error processes are Brownian motions.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \mu_{\infty} )-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design points</td>
<td>Criterion value</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 3.10, 5.51, 8.40, 10]</td>
</tr>
<tr>
<td>Second group: ( f_1 )</td>
<td>[1, 3.04, 5.49, 8.29, 10]</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 3.30, 5.67, 7.34, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 1.44, 5.79, 9.58, 10]</td>
</tr>
<tr>
<td>First group: ( f_2 )</td>
<td>[1, 1.41, 5.86, 9.55, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 1.52, 5.76, 9.59, 10]</td>
</tr>
</tbody>
</table>

### Table 2

\( \mu_{\infty} \)-optimal design points for \( \hat{\mu}_0 \) of five observations in each group in the interval \([1, 10]\) and \( \mu_{\infty} \)-optimality criterion values for the \( \mu_{\infty} \)-optimal and equidistant (in brackets) designs. The error processes have the same exponential covariance kernel with \( \lambda = 0.5 \).

<table>
<thead>
<tr>
<th>Models</th>
<th>( \mu_{\infty} )-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design points</td>
<td>Criterion value</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 2.87, 5.41, 8.14, 10]</td>
</tr>
<tr>
<td>Second group: ( f_1 )</td>
<td>[1, 3.15, 5.47, 8.22, 10]</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 3.20, 5.09, 8.07, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 2.43, 5.60, 5.91, 10]</td>
</tr>
<tr>
<td>First group: ( f_2 )</td>
<td>[1, 1.54, 5.27, 9.70, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 5.15, 5.87, 9.34, 10]</td>
</tr>
</tbody>
</table>

### Table 3

\( \mu_{\infty} \)-optimal design points for \( \hat{\mu}_0 \) of five observations in each group in the interval \([1, 10]\) and \( \mu_{\infty} \)-optimality criterion values for the \( \mu_{\infty} \)-optimal and equidistant (in brackets) designs. The error processes have the same exponential covariance kernel with \( \lambda = 1 \).

<table>
<thead>
<tr>
<th>Models</th>
<th>( \mu_{\infty} )-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design points</td>
<td>Criterion value</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 2.98, 5.43, 8.03, 10]</td>
</tr>
<tr>
<td>Second group: ( f_1 )</td>
<td>[1, 2.72, 5.48, 7.91, 10]</td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td>[1, 2.11, 4.90, 7.77, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 2.44, 5.29, 5.90, 10]</td>
</tr>
<tr>
<td>First group: ( f_2 )</td>
<td>[1, 5.31, 6.13, 9.00, 10]</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td>[1, 2.90, 6.63, 7.48, 10]</td>
</tr>
</tbody>
</table>
Fig. 1. Confidence bands from the $\mu_\infty$-optimal (solid lines) and equidistant designs (dashed lines) with five time-points in each group and the true difference of the curves (dotted line). Left panel: the proposed estimator $\hat{\theta}_n^*$. Right panel: the weighted least squares estimator $\hat{\theta}_{WLS}$. First row: model (5.1) for both groups. Second row: model (5.1) for first group and model (5.2) for second group. Third row: model (5.2) for both groups. The covariance structure is Brownian motion in all cases.

at least up to the optimal weights. Based on these observations, in what is to follow the focus is on the alternative estimator $\hat{\theta}_n^*$ proposed and on the case of Brownian motion for both error processes.

5.2. Different sample sizes

As proposed by a referee, the case of unequal sample sizes is also considered and for purposes of comparison with the results of the previous section, the number of observation points are $n_1 = 5$ and $n_2 = 6$ for the first and second group respectively. As in the previous section, the $\mu_\infty$-optimal designs are calculated first for all the combinations of comparisons using models (5.1) and (5.2). These observation points along with their corresponding $\mu_\infty$-optimality criterion values are presented in Table 4 for the case of Brownian motion for both error processes. The values given in brackets correspond to the $\mu_\infty$-optimality criterion values when the equidistant design on $[1, 10]$ is used. For the case of $n_1 = 5$ and $n_2 = 6$, the observation points of the equidistant design are $[1, 3.25, 5.50, 7.75, 10]$ and $[1, 2.80, 4.60, 6.40, 8.20, 10]$ for the first and second group respectively and the weights are as in Eq. (4.8).

A comparison of the values in Tables 1 and 4 reveals that for either design, the criterion value decreases when $n_2 = 6$ is used instead $n_2 = 5$. This is more evident for the equidistant design thus indicating that its performance is improved. However, for the cases depicted in the second and third row of Table 4, the $\mu_\infty$-optimal design reduces the criterion value substantially.

A more direct comparison between the cases of equal and unequal sample sizes is achieved via the use of confidence bands. Fig. 3 presents the confidence bands for the pair $\hat{\theta}_n^*$ using the $\mu_\infty$-optimal design (solid lines) and the equidistant
Fig. 2. Confidence bands obtained from the estimator \( \hat{\theta}_n^* \) with the \( \mu_\infty \)-optimal (solid lines) and equidistant designs (dashed lines) with five time-points in each group and the true difference of the curves (dotted line). Left panel: covariance kernel \( \exp(-0.5|t-t'|) \). Right panel: covariance kernel \( \exp(-|t-t'|) \). First row: model (5.1) for both groups. Second row: model (5.1) for first group and model (5.2) for second group. Third row: model (5.2) for both groups.

Table 4

\( \mu_\infty \)-optimal design points for \( \hat{\theta}_n^* \) of five and six observations in the first and second group respectively in the interval \([1, 10]\) and \( \mu_\infty \)-optimality criterion values for the \( \mu_\infty \)-optimal and equidistant (in brackets) designs. Both error processes are Brownian motions.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \mu_\infty )-optimality</th>
<th>Design points</th>
<th>Criterion value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First group: ( f_1 )</td>
<td>( \mu_\infty )-optimality</td>
<td>[1, 2.82, 5.64, 8.30, 10]</td>
<td>0.59 (0.71)</td>
</tr>
<tr>
<td>Second group: ( f_1 )</td>
<td></td>
<td>[1, 2.69, 4.72, 6.05, 8.51, 10]</td>
<td></td>
</tr>
<tr>
<td>First group: ( f_1 )</td>
<td></td>
<td>[1, 2.79, 5.70, 8.18, 10]</td>
<td>1.08 (3.16)</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td></td>
<td>[1, 3.04, 5.14, 5.71, 8.24, 10]</td>
<td></td>
</tr>
<tr>
<td>First group: ( f_2 )</td>
<td></td>
<td>[1, 5.27, 5.82, 9.13, 10]</td>
<td>3.27 (27.97)</td>
</tr>
<tr>
<td>Second group: ( f_2 )</td>
<td></td>
<td>[1, 1.56, 3.27, 5.13, 5.95, 10]</td>
<td></td>
</tr>
</tbody>
</table>

design (dashed lines), whereas as before the dotted lines give the plot for the true difference \( f_1^T(t)\theta_1 - f_2^T(t)\theta_2 \). The left panel of Fig. 3 gives the results for \( n_1 = n_2 = 5 \) and the corresponding results for \( n_1 = 5, n_2 = 6 \) are depicted in the right panel. Furthermore, the first row of graphs corresponds to the case where models (5.1) and (5.2) are considered for the first and second group respectively whereas for the second row of graphs model (5.2) is used for both groups. The case of comparing two models of the form (5.1) is omitted since as it is shown in Fig. 1, the equidistant design produces a similar confidence band to that for the \( \mu_\infty \)-optimal design even for the smaller sample sizes \( n_1 = n_2 = 5 \).
As expected, an increase in the total sample size results in a smaller width of the confidence bands corresponding to the equidistant design. In particular, when models (5.1) and (5.2) are compared (first row), the equidistant design performs better when \( n_1 = 5 \) and \( n_2 = 6 \) since the corresponding confidence band is closer to that based on the \( \mu_\infty \)-optimal design. An explanation for this phenomenon is that assuming a common sample size, the equidistant design is already close to the \( \mu_\infty \)-optimal design for model (5.1) and thus increasing the number of observations for the second group for which model (5.2) is considered, will improve the accuracy of inference even if the equally spaced design points are used. Nevertheless, in both cases depicted in Fig. 3, the \( \mu_\infty \)-optimal design performs best both in terms of minimising the maximal width of the confidence band as well as in terms of producing a confidence band with a similar shape as that for the true difference \( f_1(t)\theta_1 - f_2(t)\theta_2 \).

5.3. \( \mu_1 \)-optimality criterion

Following a referee’s suggestion, an alternative \( L_p \)-norm criterion that can be used for the comparison of two parametric curves is the \( \mu_1 \)-optimality criterion given by

\[
\mu_{1,n}(\Gamma) = \int_a^b \varphi_n(t, \Gamma) \, dt = \int_a^b \sum_{i=1}^{2} f_i^T(t) \text{Var}(\hat{\theta}_{i,n}) f_i(t) \, dt = \sum_{i=1}^{2} \int_a^b f_i^T(t) \text{Var}(\hat{\theta}_{i,n}) f_i(t) \, dt = \sum_{i=1}^{2} \text{tr} \left\{ \text{Var}(\hat{\theta}_{i,n}) \int_a^b f_i(t) f_i^T(t) \, dt \right\}.
\]

Therefore, the \( \mu_1 \)-optimality criterion is the sum of two trace criteria of the form considered in Remark 3.3 of Dette et al. (2015). This simplifies the optimisation problem since the design points are optimised separately for each group using the differentiable trace criteria as in Dette et al. (2015). Furthermore, Theorem 4.1 proves that the weights determined are optimal in the Loewner-ordering and thus, the optimal weights minimising (5.3) are given in Eq. (4.8).

Table 5 gives the \( \mu_1 \)-optimal design points and the \( \mu_1 \)-optimality criterion values for this design and the equidistant design (in brackets), for all combinations of comparisons using models (5.1) and (5.2). The case of \( n_1 = n_2 = 5 \) and a Brownian motion for both error processes is considered.

Although the \( \mu_1 \)-optimal and \( \mu_\infty \)-optimal designs are different, the latter designs given in Table 1 for the case of two Brownian motions, the conclusions drawn from either of these criteria are the same. That is, the criterion value decreases significantly by the \( \mu_1 \)-optimal design except when model (5.1) is used for both groups in which case the reduction is less significant. Fig. 4 provides a better picture by presenting the confidence bands based on the \( \mu_1 \)- and \( \mu_\infty \)-optimal designs.
Based on the subsequent parameter estimates, the confidence bands for the proposed estimator where

Confidence bands obtained from the estimator is examined where the error processes corresponding to Test 1 and Test 3. The dissolution results for each of the two tests are given in Table 6.

which are plotted separately using solid and dashed lines respectively. The results for the comparison of two models of the form (5.1) are not presented since for this case the equidistant design (with optimal weights) can be used instead.

It is evident that both the $\mu_1$-optimal and the $\mu_\infty$-optimal designs result in confidence bands with small width and in fact, the confidence bands based on either design are very similar. Therefore, in this example although for the $\mu_\infty$-optimality criterion the corresponding $\mu_\infty$-optimal design is indeed optimal, the $\mu_1$-optimal design (with optimal weights) can be used without detriment since it will produce similar results and it is easier to evaluate as the sum of two differentiable trace-criteria.

5.4. Case study

As pointed out by a referee, it is of interest to apply the proposed methodology on a real situation. For this reason, a case study discussed in Yuksel et al. (2000) regarding the comparison of in vitro dissolution profiles is considered. In vitro dissolution is key in drug development for the assessment of bioequivalence and thus nowadays the pharmaceutical industry and registration authorities focus on drug dissolution studies.

Yuksel et al. (2000) conduct dissolution tests on six commercial tablets of naproxen sodium and the dissolution results are the percent dissolved versus time. Several kinetic models, commonly used to describe drug dissolution, were fitted to the individual data for each tablet and ANOVA-based, model-dependent and model-independent methods were applied for the comparison of the dissolution profiles.

For the application of the proposed methodology, a comparison between the dissolution results of Tests 1 and 3, as referred to in Yuksel et al. (2000), is considered and the Higuchi model is used for both the individual data. In particular, the models considered are given by

$$Y_i(t_j) = \sqrt{t_j} \theta_1 + \varepsilon(t_j), \quad j = 1, \ldots, 6; i = 1, 2,$$

where $t_j, j = 1, \ldots, 6,$ are the six common time-points used for both groups in Yuksel et al. (2000) and $i = 1, 2$ denotes the two groups corresponding to Test 1 and Test 3. The dissolution results for each of the two tests are given in Table 6.

Unlike Yuksel et al. (2000) who did not consider possible correlation structures, the scenario of dependent observations is examined where the error processes $\varepsilon_1(t_j)$ and $\varepsilon_2(t_j)$ are assumed to have an exponential covariance kernel of the form $\mathbb{E}[\varepsilon_1(t_k)\varepsilon_2(t_i)] = \exp(-\lambda_i|t_k - t_i|) \quad (i = 1, 2).$ The parameters $\lambda_i, i = 1, 2,$ are estimated via maximum likelihood and based on these estimates the $\mu_\infty$-optimal design points for a comparison between Test 1 and Test 3 turn out to be the time-points $[5, 13, 21, 29, 37, 45]$ for both groups.

The proposed estimator $\hat{\theta}_n^\star = (\hat{\theta}_{1,n_1}^\star, \hat{\theta}_{2,n_2}^\star)$, with optimal weights as in (4.8) and the $\mu_\infty$-optimal time-points $[5, 13, 21, 29, 37, 45]$ for each group, is evaluated for $n_1 = n_2 = 6$ also using the estimates $\hat{\lambda}_1 = 0.203$ and $\hat{\lambda}_2 = 0.223$. Based on the subsequent parameter estimates, the confidence bands for the proposed estimator $\hat{\theta}_n^\star$ are constructed when

<table>
<thead>
<tr>
<th>Models</th>
<th>$\mu_1$-optimality</th>
<th>$\mu_\infty$-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>First group: $f_1$</td>
<td>$[1, 3.03, 5.56, 8.03, 10]$</td>
<td>5.56 (5.78)</td>
</tr>
<tr>
<td>Second group: $f_1$</td>
<td>$[1, 3.03, 5.56, 8.03, 10]$</td>
<td>5.56 (5.78)</td>
</tr>
<tr>
<td>First group: $f_1$</td>
<td>$[1, 3.03, 5.56, 8.03, 10]$</td>
<td>15.47 (70.05)</td>
</tr>
<tr>
<td>Second group: $f_2$</td>
<td>$[1, 1.58, 5.60, 9.42, 10]$</td>
<td>25.38 (134.32)</td>
</tr>
</tbody>
</table>

![Fig. 4](image-url) Confidence bands obtained from the estimator $\hat{\theta}_n^\star$ with the five-point $\mu_1$-optimal (solid lines) and the $\mu_\infty$-optimal designs (dashed lines) and the true difference of the curves (dotted line). Left panel: model (5.1) for first group and model (5.2) for second group. Right panel: model (5.2) for both groups.
Table 6
Dissolution data of naproxen sodium tablets for Test 1 and Test 3.

<table>
<thead>
<tr>
<th>Time</th>
<th>Percent dissolved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test 1</td>
</tr>
<tr>
<td>5 min</td>
<td>29.8</td>
</tr>
<tr>
<td>10 min</td>
<td>61.4</td>
</tr>
<tr>
<td>15 min</td>
<td>85.5</td>
</tr>
<tr>
<td>20 min</td>
<td>98.5</td>
</tr>
<tr>
<td>30 min</td>
<td>104.5</td>
</tr>
<tr>
<td>45 min</td>
<td>104.8</td>
</tr>
</tbody>
</table>

Fig. 5. Confidence bands from the $\mu_\infty$-optimal (dashed lines) design points and the time points used in Yuksel et al. (2000) (solid lines) with six observations in the interval [5, 45] for each group and the true difference of the curves (dotted line). Left panel: the proposed estimator $\hat{\theta}_n^*$. Right panel: the weighted least squares estimator $\hat{\theta}_{WLSE}$. The covariance Kernel is $\exp\{-\lambda_i |t - t'|\}$ where $\lambda_1 = 0.203$ and $\lambda_2 = 0.223$ for each of the two groups.

either the $\mu_\infty$-optimal design points [5, 13, 21, 29, 37, 45] or the time-points [5, 10, 15, 20, 30, 45] considered in Yuksel et al. (2000) are used for both groups. These confidence bands are depicted in the left panel of Fig. 5 along with the corresponding confidence bands for the pair of weighted least squares estimators (right panel). In each graph, the dashed and solid lines are used for the confidence bands from the $\mu_\infty$-optimal design points and the time points used in Yuksel et al. (2000) respectively. As before, the dotted lines correspond to the true difference $\sqrt{t_{\theta_1}} - \sqrt{t_{\theta_2}}$.

It is observed that for either estimator, the decrease in the width of the confidence band when the $\mu_\infty$-optimal design is used is small. This is not surprising given that the Higuchi model considered is a one-parameter model (see also discussion in Section 5.1). However, for either choice of design points the optimal weights given in Eq. (4.8) were used for the proposed estimator $\hat{\theta}_n^*$ and the resulting confidence bands are similar to those for the weighted least squares estimator. Therefore, the proposed methodology is useful in real applications since it is an easily implementable alternative to the weighted least squares estimator that can be used without detriment.

Acknowledgements

This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Project C2) of the German Research Foundation (DFG), the European Union Seventh Framework Programme [FP7 2007–2013] under grant agreement no 602552 (IDEAL — Integrated Design and Analysis of small population group trials) and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References


Computation of $c$-optimal designs for models with correlated observations

Juan M. Rodríguez-Díaz
University of Salamanca, Faculty of Science, Pl. de los Caidos s/n, 37.008 Salamanca, Spain

ABSTRACT

In the optimal design of experiments setup, different optimality criteria can be considered depending on the objectives of the practitioner. One of the most used is $c$-optimality, which for a given model looks for the design that minimizes the variance of the linear combination of the parameters' estimators given by vector $c$. $c$-optimal designs are needed when dealing with standardized criteria, and are specially useful when $c$ is taken to be each one of the Euclidean vectors since in that case they provide the best designs for estimating the individual parameters. The well known procedure proposed by Elfving for independent observations is the origin of the procedure that can be used in the correlation framework. Some analytical results are shown for the model with constant covariance, but even in this case the computational task can become quite hard. For this reason, an algorithmic procedure is proposed; it can be used when dealing with a general model and some covariance structures.

1. Introduction

Let us assume the linear model

$$Y = X\beta + \epsilon,$$

where $Y = (y_1, \ldots, y_n)^T$ denotes the observations vector, $X = (f(x_1), \ldots, f(x_n))^T$ is the design matrix, with $f(x) = (f_1(x), \ldots, f_m(x))^T$ and the $f_i(x)$ are linearly independent on the experimental domain $\mathcal{X}$, $\beta$ is the $m$-vector parameters and $\epsilon$ the error terms vector, that will be assumed normally-distributed with covariance matrix $\Sigma$.

For non-independent observations, the information matrix for a design $\xi$ is given by

$$M(\xi) = X^T \Sigma^{-1} X.$$  

The inverse of the information matrix $M(\xi)$ is proportional to the covariance matrix of the estimators of the parameters of the model, thus an optimality criterion typically minimizes a function of $M^{-1}(\xi)$.

There is a known set of techniques for obtaining optimal designs when the model is linear in the parameters (see for instance Fedorov and Hackl, 1997 or Atkinson et al., 2007), but in most cases assuming independent observations. For a non-linear model, the usual approach is to linearize it and use the standard toolbox for the linearized model. In this case initial values are needed for the non-linear parameters, and thus the obtained designs will be locally optimal.

For a nonzero $m$-dimensional vector $c$, the $c$-optimality criterion tries to find the design that minimizes the variance of the best linear unbiased estimator of $c^T \beta$. When $c$ is taken to be each one of the Euclidean vectors $(1, 0, 0, \ldots), (0, 1, 0, \ldots), \ldots,$
c-optimality will provide the best designs for the estimation of each parameter, which in particular are needed for the standardized criteria (Dette, 1997), that take into account the scale of the parameters. They can be as well used for checking how good a specific design is for the estimation of each one of the model parameters. Specifically, a design $\xi$ is c-optimal if minimizes $\text{Var}(c^T \hat{\beta}) = c^T M(\xi)^{-1} c$, where $M$ is a generalized inverse (more information on generalized inverses can be found for instance in Yanai et al., 2011 or Pukelsheim, 2006, where a study of generalized inverses related with c-optimality is performed). c-optimal designs are very often singular (that is, the information matrix of the corresponding design is singular), which increases the difficulty of obtaining them. Elfving (1952) provided a graphical method for finding c-optimal designs when the observations are independent. This identifies a key point corresponding to the intersection of the line defined by $c$ with the boundary of the 'Elfving Set' (the convex hull of $f(\chi) \cup -f(\chi)$). This optimal point will be a convex combination of, at most, $m$ points of the set $\pm f(\chi)$ (Fallman, 1974). The result can be stated as follows:

**Theorem 1.1.** Assuming model (1) and independent observations, the approximate design

$$
\hat{\xi}^* = \left\{ x_1, \ldots, x_r \right\},
$$

where the $x_i$ are the support points and $p_i$ is the weight that point $x_i$ has in the design. ($p_1 + \cdots + p_r = 1$) is c-optimal if there exists a point $c^*$ belonging to the line defined by vector $c$ and to the boundary of the convex hull of $f(\chi) \cup -f(\chi)$ that can be expressed as $\pm p_1 f(x_1) \pm \cdots \pm p_r f(x_r)$. That is, it has a maximal norm within the points of the convex hull that can be expressed as $y \cdot c$ with $y$ a scalar number.

The proof to the original theorem can be found in Elfving (1952).

This procedure is specially suitable for two-dimensional models, for which the required convex hull is easy to obtain, and produces approximate designs. Pukelsheim and Torsney (1991) give a method for computing c-optimal weights given the support points, López-Fidalgo and Rodríguez-Díaz (2004) generalize the Elfving’s method to the multi-dimensional case and Harman and Jurik (2008) improve the computation task by using linear programming. Pukelsheim (2006) contains as well an updated approach to Elfving’s theorem.

In the next section, a procedure for obtaining c-optimal designs when the observations are correlated is presented. Roughly speaking, the idea is to use a change of variables such that it turns the information matrix (2) into the shape $\hat{X}^T \hat{X}$, and thus some ideas from Elfving’s method for independent observations can be applied, now to the new design matrix $\hat{X}$. Although the method is not suitable for every covariance structure it can be used to solve some cases, and several examples of these applications are shown in Section 3. Finally, Section 4 describes a summary of the results appearing in the paper, introducing as well some other cases where the technique could be applied. Recent papers related to the keys of the present work are Tommasi et al. (2014), which computes c-optimal designs for log-linear models, and Dette et al. (in press), where optimal designs for comparing models under correlation are computed.

2. A derivation of Elfving’s method for correlated observations

A strictly positive definite covariance matrix will be assumed, that is, every eigenvalue of $\Sigma$ will be assumed to be greater than zero. Through the paper, different examples using strictly positive definite stationary covariance kernels $\rho(\cdot)$ will be shown (see Dette et al., 2013, 2015), giving rise to the covariance structure $\text{cov}[y(x_1), y(x_j)] = $ $\rho(|x_1 - x_j|)$. For this type of correlation, different support points should be assumed in order to avoid singular covariance matrices, thus we should initially fix $n$ and restrict to exact designs, $\xi = \{x_1, \ldots, x_n\}$, with $x_i \neq x_j$ for all $i, j$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\Sigma$, the covariance matrix can be decomposed as

$$
\Sigma = ADA^T ,
$$

where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $A$ is an orthonormal matrix whose $i$th column is an eigenvector with eigenvalue $\lambda_i$. $A$ is unique (but the sign of the vectors) when the $n$ eigenvalues are different, otherwise a convenient basis can be chosen in each subspace of eigenvectors of $\Sigma$ with eigenvalue $\lambda_i$. More details about the construction of $A$ can be checked for instance in Magnus and Neudecker (1999) (Schur decomposition).

2.1. Characterization of c-optimal designs

For a given $c$, a c-optimal design is any $\hat{\xi}^*$ that minimizes $\text{Var}(c^T \hat{\beta}) = c^T M(\xi)^{-1} c$, with $M(\xi)$ given by (2). Using (3)

$$
\text{Var}(c^T \hat{\beta}) = c^T \{X^T \Sigma^{-1} X\}^{-1} c = c^T \{\tilde{X}^T \tilde{X}\}^{-1} c,
$$

with $\tilde{X} = D_1 A' X$ and $D_1$ the square root of $D^{-1}$, that is $D_1 = \text{diag}(\lambda_1^{-1/2}, \ldots, \lambda_n^{-1/2})$. $\tilde{X}$ is a function of the design $\xi$, and it will be expressed as $\tilde{X}(\xi)$ when it may be convenient to stress this fact. The change of variables makes the minimization problem equivalent to that of independent observations that has $\tilde{X}$ as design matrix. That is, the initial model (1) becomes

$$
\tilde{Y} = \tilde{X} \beta + \tilde{\epsilon},
$$

(4)
with $\tilde{Y} = D_cA^tY, \tilde{c} = D_cA^ce$. And now $\Sigma_\tau = DA^t \Sigma AD_c = I_n$, where $I_n$ denotes the identity matrix of order $n$. Thus, the new variables $\tilde{Y}$ are uncorrelated.

Let us consider unbiased estimators of $c^t\beta$ that are linear functions of the observations, $\varphi = \sum h_i y_i = h^t Y$, with $h = (h_1, \ldots, h_n)^t$. The unbiasedness condition,

$$E[\varphi] = h^t X \beta = c^t \beta,$$

produces $h^t X = c^t$ or equivalently

$$\tilde{h}^t \tilde{X} = c^t,$$

with $\tilde{h} = D^{1/2} A^t h$. For other hand

$$\text{Var}(\varphi) = \text{Var}(h^t Y) = \text{Var}(\tilde{h}^t \tilde{Y}) = \sum_{i=1}^n \tilde{h}_i^2 = \|\tilde{h}\|^2,$$

the square of the vector norm. For a specific design $\xi$, the vector $\tilde{h}$ defines a linear combination of the rows of the new design matrix $\tilde{X}(\xi)$ fulfilling (5). In the general case, there will be infinite vectors $\tilde{h}$ that verify this condition, but since the aim is minimizing $\text{Var}(\varphi)$ we are interested in the ‘shortest’ one, the one that has minimum norm. This is a similar result to that in Elfving (1952). There the search for linear combinations of rows of the design matrix was restricted to points of the so called ‘Elfving Set’, looking for the outermost one fulfilling the unbiasedness condition. Here the objective is to find the linear combination of points of $X \cup -X$ that verify (5) and has a coefficient vector $\tilde{h}$ with minimum norm. Having in mind the condition (5), the aim is to look for designs $\xi$ that produce design matrices $\tilde{X}(\xi)$ as ‘big as possible’, which implies that the corresponding $\tilde{h}$ will be ‘small’.

The optimal design will be $\xi^*$ such that for the corresponding $\tilde{X}(\xi^*)$ there exists $\tilde{h}$ verifying (5) and having minimum norm. That is

$$\xi^* = \arg \min_{\xi} \min_{\tilde{h}} \left\{ \|\tilde{h}\| : \tilde{h}^t \tilde{X}(\xi) = c^t \right\}.$$

In the next subsection, some guidelines will be proposed in order to formulate the minimization problems. Thus, the main difficulty is the computation of the eigenvalues of $\Sigma$, and in fact the diagonalization of the covariance matrix before the design points $x_1, \ldots, x_n$ are fixed is the main restriction of the procedure. For other hand, the $\tilde{X}$ matrix could be conveniently modified in order to facilitate the computations taking into account the following properties:

- The order of the design points does have influence in $\tilde{X}$ and thus in the shape of the corresponding Elfving Set. However, if a design $\xi^*$ is found to be $c$-optimal then so will be any permutation of its points, since a permutation just means a reordering of the rows of the original design matrix $X$ which will have no influence in the value of $\text{Var}(c^t \beta)$ (see Lemma A.1 in the Appendix).
- If for every design $\xi$ a proportional design matrix $k\tilde{X}(\xi)$ is used instead of $\tilde{X}(\xi)$, the best combination of rows fulfilling (5) will now be given by $\tilde{h}/k$, but the optimal design will remain unaltered.
- Changing the sign of one or more rows of $\tilde{X}$ has no influence in the Elfving Set neither in the $c$-optimal designs.

2.2. Procedure

When $n = m$, that is the number of observations is the number of parameters of the model, and a general expression for $\tilde{X}(\xi)$ can be obtained, a $c$-optimal design may be computed solving the equation system given by (5) and looking for the design $\xi$ that leads to a minimum norm of $\tilde{h}$ (see for instance Example 3.1.1).

For $n > m$, again assuming that a expression for $\tilde{X}(\xi)$ is at disposal, a $c$-optimal design can be obtained through the following steps:

i. For a design $\xi = \{x_1, \ldots, x_n\}$ we need to find a vector $\tilde{h}$ such that $\tilde{h}^t \tilde{X}(\xi) = c^t$. This can be done by choosing $m$ rows of $\tilde{X}$, setting $\tilde{h}_i = 0$ for every row different than these $m$, and getting the nonzero components of $\tilde{h}$ by solving (5). If the equation system has not full rank or any of the components $\tilde{h}_i$ goes to infinity, a different set of $m$ rows of $\tilde{X}$ may be chosen. This initial solution will be denoted by $\tilde{h}_0$.

ii. Consider the linear mapping $l_\xi$

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m \quad v \longrightarrow v^t \tilde{X}(\xi).$$

The design $\xi$ is said to be admissible for $c^t \beta$ if $c^t \in \text{Im}(l_\xi)$, that is if there exists $v_\xi \in \mathbb{R}^n$ such that $v_\xi^t \tilde{X}(\xi) = c^t$, or that $c^t$ could be expressed as a linear combination of the rows of $\tilde{X}(\xi)$. Let $\mathcal{K}_\xi$ be the kernel of $l_\xi$, $\mathcal{K}_\xi = \{v \in \mathbb{R}^n : v^t \tilde{X}(\xi) = 0\}$, with dimension $n - m$. Then, the set that contains every vector $\tilde{h}$ verifying (5) will be

$$\mathcal{H}_\xi = \{\tilde{h}_0 + v, v \in \mathcal{K}_\xi\}.$$
iii. Now we look for the ‘smallest’ element \( \tilde{h}^* \) of \( \mathcal{H} \), that is the one with minimum norm. This minimum norm is the (orthogonal) distance from the point of \( \mathcal{H}^0 \) defined by \( \tilde{h}_0 \) to the hyperplane \( \mathcal{K}_{\tilde{h}_0} \). The optimal coefficient vector can be expressed as \( \tilde{h}^* = \tilde{h}_0 - \tilde{h}_0 \), where \( \tilde{h}_0 \) is the orthogonal projection of \( \tilde{h}_0 \) onto \( \mathcal{K}_{\tilde{h}_0} \).

If \( v_1, \ldots, v_{n-m} \) are linearly independent vectors spanning \( \mathcal{K}_{\tilde{h}_0} \), let \( A = [v_1, \ldots, v_{n-m}] \) denote the \( n \times (n - m) \) matrix with the \( v_i \)'s as columns. Then, the orthogonal projection matrix onto \( \mathcal{K}_{\tilde{h}_0} \) is given by

\[
P = A(A' A)^{-1} A',
\]

(see Yanai et al., 2011, Section 2.2). Thus, \( h_0 = P\tilde{h}_0 \) and \( \tilde{h}^* = Q\tilde{h}_0 \), where \( Q = I - P \) is the orthogonal projection matrix onto \( \mathcal{K}_{\tilde{h}_0}^\perp \).

**Example 3.1.2** shows the application of this procedure when computing \( n \)-point \( c \)-optimal designs for two-parameter models assuming constant covariance.

### 3. Examples of application

In the following, some examples dealing with different stationary covariance kernels will be shown. In first place, the case of constant covariance will be studied.

#### 3.1. Constant covariance

Consider model (1) and assume constant covariance \( \sigma^2 \) between observations, where \( 0 < a < 1 \) and \( \sigma^2 \) is the variance of the observations. It can be assumed \( \sigma^2 = 1 \) since it has no influence in the design, thus it is equivalent to use the covariance kernel given by \( \rho(0) = 1 \) and \( \rho(t) = a^2 \) for \( t > 0 \). Then

\[
\Sigma = \begin{pmatrix}
1 & a & \cdots & a \\
a & 1 & \cdots & a \\
\vdots & \vdots & \ddots & \vdots \\
a & a & \cdots & 1
\end{pmatrix}.
\]

(8)

The eigenvalues of \( \Sigma \) are \((1 - a)\), with multiplicity \((n - 1)\), and \(1 + (n - 1)a\). An orthonormal basis can be found (for instance using Gram–Schmidt procedure) in the subspace of eigenvectors of \( \Sigma \) with eigenvalue \( 1 - a \). The normalized eigenvector with eigenvalue \( 1 + (n - 1)a \) is \((1, 1, \ldots, 1)/\sqrt{n} \). Then, the rows of the transformed design matrix \( \tilde{X} \) described in Section 2.1 can be expressed as

\[
\tilde{X}_i = -[i(i + 1)(1 - a)]^{-1/2} \left[ f(x_1) - i f(x_{n+1-i}) + \sum_{j=n+2-i}^{n} f(x_j) \right] \quad i = 1, \ldots, n - 1,
\]

\[
\tilde{X}_n = [n(1 + (n - 1)a)]^{-1/2} \sum_{j=1}^{n} f(x_j).
\]

(9)

The computation of \( c \)-optimal designs can now be done following the procedure described in Section 2.2. Analytical results for the two-parameter model \((1, g(x))^t\) when looking for two and three-point designs are described in Theorems 3.1 and 3.2, respectively.

**Theorem 3.1.** Consider the model \( f(x)^t = \{1, g(x)\} \) and assume constant covariance \( \sigma^2 \) between observations, where \( 0 < a < 1 \) and \( \sigma^2 \) is the variance of the observations. Then, the \( c \)-optimal two-point designs do not depend on the covariance \( a \) and are given by

\[
\begin{align*}
\{x_{\text{min}}, x_{\text{max}}\} & \quad \text{if } c = (0, 1)^t, \\
\{x_1^*, x_2^*\} & \quad \text{if } c = (1, \tilde{c})^t,
\end{align*}
\]

where \( \tilde{c} \) is any scalar number, \( x_{\text{min}}, x_{\text{max}} \) are, respectively, the points where the minimum and maximum of the function \( g(x) \) in the design interval \( \mathcal{X} \) are attained, and \( x_1^*, x_2^* \) are the points where the function \( |[2\tilde{c} - g(x_1) - g(x_2)]/[g(x_1) - g(x_2)]|^2 \) attains its minimum.

**Proof.** From (9), and since proportional design matrices produce equivalent \( c \)-optimal designs and changes of rows signs have no influence in the Elfving Set, we can assume \( \tilde{X} \) to be the matrix

\[
\tilde{X} = \begin{pmatrix}
0 \\
\sqrt{\frac{1 + a}{1 - a}} [g(x_1) - g(x_2)] \\
2 [g(x_1) + g(x_2)]
\end{pmatrix}.
\]
The condition $\tilde{h}_1, \tilde{h}_2)\tilde{X} = (c_1, c_2)$ produces
\[
\begin{align*}
\tilde{h}_1 &= \sqrt{\frac{1-a}{c_2 - c_1(g(x_1) + g(x_2)) / 2}} \\
\tilde{h}_2 &= c_1 / 2.
\end{align*}
\]
and now the proof is straightforward when taking into account that the c-optimal design minimizes $\tilde{h}_1^2 + \tilde{h}_2^2$.  

**Remark 3.1.** Since the design $[x_{\text{min}}, x_{\text{max}}]$ maximizes the denominator of the function $([2\tilde{c} - g(x_1) - g(x_2)]/[g(x_1) - g(x_2)])^2$ from **Theorem 3.1**, this will be the c-optimal design for many values of $c$, specially for $c = (1, \tilde{c})^t$ with $\tilde{c}$ big.

In the following, some examples of application of the procedures described previously will be shown.

**Example 3.1.1.** Table 1 shows two-point c-optimal designs for model $f(x)^t = \{1, g(x)\}$ taking different functions $g(x)$ and design interval $X$, computed as described in **Theorem 3.1** and assuming a minimum distance 0.1 between observations.

**Example 3.1.2.** Let us consider again model $f(x)^t = \{1, g(x)\}$ but now looking for n-point c-optimal designs. Assuming constant covariance and using the possibilities of changing the sign of some rows and taking a proportional matrix, the transformed design matrix can be written as
\[
\bar{X} = \begin{pmatrix}
0 & \sqrt{\frac{n(1 + (n - 1)a)}{2(1-a)} [g(x_1) - g(x_n)]} \\
0 & \sqrt{\frac{n(1 + (n - 1)a)}{6(1-a)} [g(x_1) - 2g(x_{n-1}) + g(x_n)]} \\
\vdots & \vdots \\
0 & \sqrt{\frac{n(1 + (n - 1)a)}{(n-1)(1-a)} [g(x_1) - (n - 1)g(x_2) + g(x_3) + \cdots + g(x_n)]} \\
n & [g(x_1) + \cdots + g(x_n)]
\end{pmatrix}.
\]  

Let us follow the procedure described in Section 2.2. Fixing for instance $\tilde{h}_i = 0$ for $i = 2, \ldots, n - 1$ (that is considering just the points of the Elving Set given by rows 1 and n), the condition (5) produces
\[
\tilde{h}_0 = \left( \frac{2(1-a)}{n(1 + (n - 1)a)} \frac{c_2 - c_1 \sum g(x_i)/n}{g(x_1) - g(x_n)} \cdot 0, \ldots, 0, \frac{c_1}{n} \right)^t.
\]
For other hand $v = (v_1, \ldots, v_n)^t \in \mathcal{K}_\bar{x}$ implies $v_n = 0$ and
\[
\sum_{i=1}^{n-1} \frac{v_i}{\sqrt{\tilde{h}(i+1)}} \left( g(x_1) - ig(x_{n+1-i}) + \sum_{j=n+2-i}^{n} g(x_j) \right) = 0,
\]  
thus $\mathcal{K}_\bar{x}$ is an hyperplane of dimension $n - 2$ that does not depend on the covariance $a$. Then
\[
\mathcal{H}_\bar{x} = \{ \tilde{h} = \tilde{h}_0 + v, v \in \mathcal{K}_\bar{x} \}
= \left\{ \tilde{h} = (\tilde{h}_0 + v_1, v_2, \ldots, v_{n-1}, c_1/n) : \{v_i\} \text{ verify condition (11)} \right\},
\]
and the c-optimal design will be
\[ \xi^* = \arg \min_{\xi} (\hat{h}_1^2 + \cdots + \hat{h}_n^2). \]

In the general case \( \xi^* \) will depend on \( a \), but not for the model with intercept (see Theorem 3.3). For instance, for \( n = 3 \)
\[ \mathcal{K}_\xi = \left\{ v = \left( \frac{g(x_1) - 2g(x_2) + g(x_3)}{\sqrt{3}[g(x_1) - g(x_3)]}, 1, 0 \right) \right\}, \]
and choosing rows 1 and 3, we get
\[ \hat{h}_0 = \left( \frac{2(1 - a)}{3(1 + 2a)} c_2 - c_1 [g(x_1) + g(x_2) + g(x_3)]/3, 0, c_1/3 \right)^t. \]

Then, any \( \tilde{h} \) verifying (5) has the shape \( \hat{h}_0 + \lambda v \) and a c-optimal design can be obtained as
\[ \xi^* = \arg \min_{\{\xi, \lambda\}} \|\hat{h}_0 + \lambda v\|. \]

If the actual c-optimal design verifies \( g(x_1) = g(x_3) \) it will not be detected, since for this case the norms of \( \hat{h}_0 \) and \( v \) go to infinity. The reason is that when \( g(x_1) = g(x_3) \) the first row of \( \tilde{X} \) is \((0, 0)\) and thus it should not be used. Under this assumption and taking row two instead we get \( v = (1, 0, 0) \) and
\[ \hat{h}_0 = \frac{1}{3} \left( 0, \frac{1 - a}{2(1 + 2a)} \frac{3c_2 - c_1 [2g(x_1) + g(x_2)]}{g(x_1) - g(x_3)}, c_1 \right)^t. \]

It would be sensible to try both minimization problems in order to cover all cases. Furthermore, from (9), it is clear that the first \( n - 1 \) rows of \( \tilde{X} \) vanish when \( f(x_1) = \cdots = f(x_n) \), but even that this design would be admissible when \( f(x_i) \) is proportional to \( c \), because in that case the last row of \( \tilde{X} \) would be proportional to \( c \) as well. This procedure will not detect those cases, that are not easy to occur because both conditions \( f(x_1) = \cdots = f(x_n) \) and \( f(x_i) \propto c \) should be true at the same time, but this possibility should be checked.

In this case (three-point designs), it is possible to obtain explicit expressions of the final function that needs to be minimized in order to find the c-optimal design, as it is shown in Theorem 3.2:

**Theorem 3.2.** Consider the model \( f(x)^i = \{1, g(x)\} \) and assume constant covariance \( a\sigma^2 \) between observations, where \(- (n - 1)^{-1} < a < 1\) and \( \sigma^2 \) is the variance of the observations. Then, the c-optimal three-point design for \( c = (c_1, c_2)^t \) does not depend on \( a \) and is given by \( \xi^* = \{x_1, x_2, x_3\} \) such that the \( x_i \) minimize
\[ \frac{3c_2 - c_1 \sum_{i=1}^3 g(x_i)}{\sum_{i,j=1}^3 (-1)^{1-\delta_{ij}} g(x_i) g(x_j)} \]
where \( \delta_{ij} \) is the Kronecker’s delta, that is, \( \delta_{ij} = 1 \) when \( i = j \) and 0 otherwise.

**Proof.** For three observations and two model parameters \( \mathcal{K}_\xi = \langle v \rangle \) has dimension 1, thus the projection matrix (7) is just
\[ P = \frac{v v^t}{\|v\|^2}. \]

Then, using the results above, it can be obtained by straightforward computation
\[ \|\hat{h}^*\|^2 = \frac{\hat{h}_{01}^2}{1 + v_1^2} + \frac{c_1^2}{9} = \frac{1 - a}{18(1 + 2a)} \left[ \frac{3c_2 - c_1 \sum_{i=1}^3 g(x_i)}{\sum_{i,j=1}^3 (-1)^{1-\delta_{ij}} g(x_i) g(x_j)} + \frac{c_1^2}{9} \right], \]
and the variance of the estimator of \( c^t \beta \) (that has been corrected taking into account the proportional \( \tilde{X} \) (10) employed in the computations) is \( \text{Var}(\varphi) = 3(1 + 2a) \|\hat{h}^*\|^2 \). \( \Box \)
Example 3.1.4. Finally, the model without intercept \( f(x)^T = \{\cos(x), e^x\} \) in the design interval \( \mathcal{X} = \left[ -\pi/2, \pi/2 \right] \) will be considered, again assuming constant covariance \( a \) and a minimum distance 0.1 between observations. Table 4 shows \( c \)-optimal two-point designs for this model and different values of \( a \).
Table 3
$c$-optimal three and four-point designs for model $f(x)^T = \{1, x, x^2\}$.

<table>
<thead>
<tr>
<th>$c^t$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 0, 0)</td>
<td>(0, 0.1, 0.2)</td>
<td>(0, 0.1, 0.652, 1)</td>
</tr>
<tr>
<td>(0, 1, 0)</td>
<td>(0.0536, 1)</td>
<td>(0.499, 0.599, 1)</td>
</tr>
<tr>
<td>(0, 0, 1)</td>
<td>(0.5, 1)</td>
<td>(0.45, 0.55, 1)</td>
</tr>
<tr>
<td>(1, 2, 1)</td>
<td>(0.445, 0.545)</td>
<td>(0.387, 0.487, 0.587)</td>
</tr>
</tbody>
</table>

Table 4
$c$-optimal two-point designs for model $f(x)^T = \{\cos(x), e^x\}$ and different values of $a$.

<table>
<thead>
<tr>
<th>$c^t$</th>
<th>$a$</th>
<th>$0.1$</th>
<th>$0.5$</th>
<th>$0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1)</td>
<td>$[-\pi/2, \pi/2]$</td>
<td>$[\pi/2 - 0.1, \pi/2]$</td>
<td>$[\pi/2 - 0.1, \pi/2]$</td>
<td>$[\pi/2 - 0.1, \pi/2]$</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>$[-0.021, \pi/2]$</td>
<td>$[0.075, \pi/2]$</td>
<td>$[0.290, 1.542]$</td>
<td>$[0.290, 1.542]$</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>$[-0.052, 0.047]$</td>
<td>$[-0.052, 0.047]$</td>
<td>$[-\pi/2, 0.160]$</td>
<td>$[-\pi/2, 0.160]$</td>
</tr>
<tr>
<td>(1, 10)</td>
<td>$[0.730, \pi/2]$</td>
<td>$[0.641, \pi/2]$</td>
<td>$[0.484, \pi/2]$</td>
<td>$[0.484, \pi/2]$</td>
</tr>
<tr>
<td>(10, 1)</td>
<td>$[-0.017, \pi/2]$</td>
<td>$[0.080, \pi/2]$</td>
<td>$[0.294, 1.532]$</td>
<td>$[0.294, 1.532]$</td>
</tr>
</tbody>
</table>

Table 5
$c$-optimal two-point designs for model $f(x)^T = \{1, x\}$ and different covariance structures.

| $c^t$ | $\mathcal{W}(x_1, x_2), X$ | $\max(0, 1 - |x_1 - x_2|, [-1, 1])$ | $\{e^{-|x_1 - x_2|^2}, [0, 1]\}$ |
|-------|-----------------------------|---------------------------------|---------------------------------|
| (0, 1) | $[0, 1]$                   | $[-1, 1]$                       | $[0, 1]$                       |
| (1, 0) | $[0, 0.1]$                 | $[-x, x]$                       | $[0, 1]$                       |
| (1, 1) | $[0.9, 1]$                | $[0.9, 1]$                      | $[0.9, 1]$                     |
| (1, 10) | $[0, 1]$                | $[-1, 1]$                       | $[0.9, 1]$                     |
| (10, 1) | $[x_1, x_2] : x_1 + x_2 = 0.2, for every w(x_1, x_2)$ | $\mathcal{X}$ is the variance of the observations. The $c$-optimal two-point design is $\xi_0 = \{x_1^*, x_2^*\}$, where $x_1^*, x_2^*$ are the points where the minimum of the function $\min(1-w(x_1, x_2) = |2c_2 - c_1[g(x_1) + g(x_2)]|/|g(x_1) - g(x_2)|)$ is attained.

Proof. Again $\sigma^2 = 1$ will be assumed. For two-point designs, the covariance matrix $\Sigma$ has the same structure than that of constant covariance, thus the proof can be done using that of Theorem 3.1 taking into account that now $w(x_1, x_2)$ takes the role of $a$. \( \Box \)

Example 3.2.1. Table 5 shows two-point $c$-optimal designs for model $f(x)^T = \{1, x\}$ and design interval $\mathcal{X}$ assuming different covariance structures $\mathcal{W}(x_1, x_2)$ (proposed in Dette et al., 2015) and a minimum distance 0.1 between observations. The results have been obtained as described in Theorem 3.4.

4. Conclusions and discussion

A new procedure for the computation of $c$-optimal designs in the correlated setup is introduced, relating the problem to that of independent observations and thus being able to use some ideas from Elfving (1952). Analytical results have been obtained for two-parameter models with an intercept for two-point designs, both assuming constant covariance or a strictly positive definite covariance kernel. A general procedure for a number of observations greater than the number of parameters has been produced, being able to deal with some covariance structures. Examples of three-point $c$-optimal designs for a two-parameter model show its applications.

When applying the procedure to the heteroscedastic case of independent observations, the transformed matrix $\tilde{X}$ coincides with the transformed observations that (Elfving, 1952) proposes for this model. The difference is that there the application was on approximate designs while here the objective are exact designs with non-repeated points. This model
was already studied from the point of view of c-optimality from Dette and Holland-Letz (2009) in terms of generalized inverses.

Another covariance structure suitable for the proposed procedure is for instance the blocks model with constant covariance in blocks (which are independent among them). The covariance parameter may be different in the distinct blocks, thus the observations covariance matrix is a Block Diagonal matrix \( \Sigma = BD[\Omega_{n_1}, \ldots, \Omega_{n_p}] \), with submatrices

\[
\Omega_{n_i} = \begin{pmatrix}
1 & a_i & \cdots & a_i \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}
\]

of order \( n_i \) where \( p \) is the number of blocks and \( n_1, \ldots, n_p \) are the blocks sizes. The eigenvalues of \( \Omega_{n_i} \) are \((1-a_i),1+(n_i-1)a_i\). The eigenvalues of \( \Sigma \) are the set of all eigenvalues of the \( \Omega_{n_i} \), thus the transformed matrix \( \tilde{X} \) can be built. A related model is the random blocks model with different blocks variance (see Rodríguez-Díaz et al., 2016, Section 3). The development of a procedure for computing c-optimal designs for more complex models and covariance structures using as well the results of López-Fidalgo and Rodríguez-Díaz (2004) and Harman and Jurik (2008) is a work in progress.

Finally, when taking into account the correction in the estimation of the covariance matrix of the parameters of the model (Kenward and Roger, 1997, 2009), the optimality criteria should be corrected (see Rodríguez-Díaz et al., 2016). Should this correction be considered, the procedure should be adapted in order to accommodate to the new estimation of the variance–covariance matrix.

Acknowledgments

Research was supported by the Spanish Ministry of Economy and Competitiveness and Junta de Castilla y León (Grants ‘MTM 2013-47879-C2-2-P’ and ‘SA130U14’).

Appendix

Lemma A.1. Let \( \xi = \{x_1, \ldots, x_n\} \) be an exact design. Then, \( \Var(c^T \hat{\beta}) \) does not depend on the reordering of the points of \( \xi \)

Proof. Since every permutation of \( \{x_1, \ldots, x_n\} \) can be expressed as composition of basic (two-point) permutations, it will be enough to prove the property for these ones. Talking about matrices, the permutation \( P_{ij} \) = ‘Exchange the vectors in columns \( i \) and \( j \)’ is given by a \( n \times n \) matrix that has every component equal to zero but the ones in the positions \((i,j), (j,i)\) and \((k,k) \) for \( k \neq i,j \), that have the value 1. An exchange of the points \( x_i \) and \( x_j \) of \( \xi \) affects the matrices appearing in (2) in the following way (the exponent \( (i,j) \) will denote the design and matrices transformed by the permutation): \( X^{(i,j)} = P_{ij}X \) and \( \Sigma^{(i,j)} = P_{ij} \Sigma P_{ij} \). Then, \( M(\xi^{(i,j)}) = (X^{(i,j)} \Sigma^{(i,j)})^{-1}(X^{(i,j)} \Sigma^{(i,j)})^{-1}X^{(i,j)} \Sigma^{(i,j)} = M(\xi) \) since \( P_{ij} \) is symmetric and idempotent, and consequently \( \Var(c^T \hat{\beta}) \) does not change. \( \diamond \)

Proof of Theorem 3.3. Let \( f(x) = (f_1(x), \ldots, f_m(x))^T \) be the model, \( \xi = \{x_1, \ldots, x_n\} \) the design, and \( q_i = -\sqrt{n/(1+(n-1)a)/i(i+1)(1-a)} \), \( i = 1, \ldots, n-1 \), \( q_n = 1 \) the \( a \)-dependent factors of the rows of \( \tilde{X} \) as they appear in (10). For any function \( g(x) \), let us define

\[
\phi_i(g, \xi) = \left[ g(x_1) - bg(x_{n+1-i}) + \sum_{j=n+2-i}^n g(x_j) \right],
\]

\( i = 1, \ldots, n-1 \), and \( \phi_n(g, \xi) = \sum_{j=1}^n g(x_j) \). \( \phi_i(g, \xi) \) will be written just \( \phi_i(g) \) when there is no ambiguity about the design. Then, the transformed design matrix can be written as

\( \tilde{X} = (q_1 \phi_1(f), \ldots, q_{n-1} \phi_{n-1}(f), \phi_n(f))^T \),

and the condition (5) produces the system of equations

\[
\sum_{i=1}^n q_i \tilde{h}_i \phi_i(f_j) = c_j, \quad j = 1, \ldots, m.
\]

From the first one

\[
h_n = \left[ c_1 - \sum_{i=1}^{n-1} q_i \tilde{h}_i \phi_i(f_1) \right]/\phi_m(f_1) = \varphi_n(\xi, c) + \sum_{i=1}^{n-1} q_i \tilde{h}_i \psi_{1,n}(\xi, c),
\]
where the functions \( \varphi, \psi \) are not depending on the covariance parameter \( a \). Replacing the value of \( h_n \) in the rest of the equations, and operating recurrently we can obtain

\[
\tilde{h}_{n-j+1} q_{n-j+1} = \varphi_{n-j+1}(\xi, c) + \sum_{i=1}^{n-j} \tilde{q}_i \tilde{h}_i \psi_{j,n-j+1}(\xi, c), \quad j = 1, \ldots, m.
\]

Thus, when \( n = m \) the expressions \( \tilde{h}_i q_i \) are independent of \( a \), for \( i = 1, \ldots, n \), and \( \| \hat{h} \|^2 = \sum_{i=1}^{n} \tilde{h}_i^2 \) can be expressed as

\[
\| \hat{h} \|^2 = \tilde{q} \sum_{i=1}^{n} \tilde{\varphi}_i(\xi, c) + \tilde{\varphi}_n(\xi, c),
\]

where \( \tilde{q} = \tilde{q}(a) = (1 - a)/n(1 + (n - 1)a) \) and the functions \( \tilde{\varphi}_i \) are not depending on \( a, j = 1, \ldots, n - 1 \). Then, if \( n = m \) and \( \tilde{\varphi}_n \) is independent on the design \( \xi \), the \( n \)-point \( c \)-optimal designs minimizing (13) will not depend on the covariance parameter \( a \). This happens for instance for the model with intercept, because in this case \( f_1(x) = 1 \) it will be \( \phi_i(f_1) = 0 \) for all \( i = 1, \ldots, n - 1 \) and \( \phi_n(f_1) = 1 \). Thus \( \tilde{h}_n = c_1/n \) and \( \tilde{\varphi}_n(\xi, c) = (c_1/n)^2 \).

For \( n > m \) the vector \( \hat{h} \) can be expressed as \( \hat{h} = (\hat{h}_0 + v, v \in \mathcal{J}_\xi) \), where \( \hat{h}_0 \) has \( n - m \) coordinates equal to 0 and \( v = (v_1, \ldots, v_{n-1}, 0) \) is not depending on \( a \) (see Example 3.1.2, the last coordinate of \( v \) equal to 0 is due to the intercept term). Using arguments similar to those of the case \( n = m \) above, we can assume \( \hat{h}_0 \) to have the shape \( \hat{h}_0 = (\tilde{q} v_1, \ldots, \tilde{q} v_{m-1}, 0, \ldots, 0, c_1/n) \). For other hand, if \( v^{(1)}, \ldots, v^{(n-m)} \) are the generating vectors of \( \mathcal{J}_\xi \) then any \( \tilde{h} \) verifying (5) can be written as

\[
\tilde{h} = \tilde{h}_0 + \tilde{\lambda}_1 v^{(1)} + \cdots + \tilde{\lambda}_{n-m} v^{(n-m)} = \tilde{h}_0 + \tilde{q} \lambda_1 v^{(1)} + \cdots + \tilde{q} \lambda_{n-m} v^{(n-m)},
\]

where \( \tilde{\lambda}_i = \tilde{q} / \tilde{a} \) and

\[
\tilde{h}_i = \tilde{q} \left( \tilde{\varphi}_i + \lambda_1 v^{(1)} + \cdots + \lambda_{n-m} v^{(n-m)} \right), \quad i = 1, \ldots, n - 1; \quad \tilde{h}_m = c_1/n,
\]

with \( \tilde{\varphi}_j = 0 \) for \( j = m, \ldots, n - 1 \). We need to find the design \( \xi^* \) and coefficients \( \lambda_i^* \) such that minimize the norm of \( \tilde{h} \),

\[
\{\xi^*, \lambda_1^*, \ldots, \lambda_{n-m}^*\} = \arg \min_{\xi, \lambda_1, \ldots, \lambda_{n-m}} (\tilde{h}_1^2 + \cdots + \tilde{h}_m^2).
\]

For a different value of \( a \), the \( \lambda_i^* \) in general may change (namely \( \lambda_i^* (a') = \lambda_i^* (a) \tilde{q}(a)/\tilde{q}(a') \)), but not the design \( \xi^* \).

References


Optimal response and covariate-adaptive biased-coin designs for clinical trials with continuous multivariate or longitudinal responses

Anthony C. Atkinson\textsuperscript{a,\textdagger}, Atanu Biswas\textsuperscript{b}

\textsuperscript{a} Department of Statistics, London School of Economics, London WC2A 2AE, UK
\textsuperscript{b} Applied Statistics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700 108, India

\textbf{HIGHLIGHTS}

- First paper for clinical trials with normally distributed longitudinal responses that balances efficiency and randomness.
- Design is both covariate and response adaptive.
- Extensions introduced to optimum design theory for multivariate responses.

\textbf{ABSTRACT}

Adaptive randomization of the sequential construction of optimum experimental designs is used to derive biased-coin designs for longitudinal clinical trials with continuous responses. The designs, coming from a very general rule, target pre-specified allocation proportions for the ranked treatment effects. Many of the properties of the designs are similar to those of well-understood designs for univariate responses. A numerical study illustrates this similarity in a comparison of four designs for longitudinal trials. Designs for multivariate responses can likewise be found, requiring only the appropriate information matrix. Some new results in the theory of optimum experimental design for multivariate responses are presented.

© 2016 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

Response-adaptive designs are becoming increasingly popular in phase III clinical trials with sequential entrance of patients. The ethical objective is to use the accumulating data to skew the allocation in favour of the better treatments, so ensuring that as few patients as possible receive bad treatments. The advantages of response-adaptive designs are extolled by Zelen and Wei (1995), Hu and Rosenberger (2003) and Rosenberger and Hu (2004). Gallo et al. (2006) provide a perspective from the pharmaceutical industry.

Our procedure is based on the adaptive randomization of treatment allocations from the sequential construction of optimum experimental designs. As a consequence, we require optimum designs for multivariate continuous responses that
provide balance over the prognostic factors that may be included in the estimation of treatment effects. Unfortunately, the majority of the adaptive designs that have been developed are for a single binary response per patient in the absence of covariates. Examples include the play–the-winner (PW) design (Zelen, 1969), the randomized play–the-winner (RPW) design (Wei and Durham, 1978), the success driven design (Durham et al., 1998) and the drop–the-loser (DL) rule (Ivanova, 2003). Related designs for continuous responses, using non-parametric methods to discretize the problem, include Rosenberger (1993) and Bandyopadhyay and Biswas (2004).

These designs work well in skewing the allocation in favour of the better treatment, although they are not derived from any optimality criterion. One form of optimality, for binary responses, consists of minimizing an aspect of behaviour, such as the total expected number of failures, for a given variance of the estimated treatment difference. Such designs include those of Rosenberger et al. (2001) and Biswas and Mandal (2007) for binary responses. Zhang and Rosenberger (2006, 2007) and Biswas et al. (2007) find optimum designs for continuous responses.

Several of these procedures have been extended to design in the presence of covariates, giving rise to Covariate Adjusted Response Adaptive (CARA) designs. For the randomized play–the-winner rule, Bandyopadhyay and Biswas (1999) combined polytomous covariates with binary responses and Bandyopadhyay and Biswas (2001) incorporated covariates in their design for continuous responses. Zhang et al. (2007) studied asymptotic properties of CARA designs under widely satisfied conditions. Optimum biased-coin designs for covariate balance, without response adaptivity, were introduced by Atkinson (1982). This form of optimality was extended to response-adaptive designs for univariate responses by Atkinson and Biswas (2005a, b). Rosenberger and Sverdlov (2008) discuss the arguments that have been advanced in the clinical trials literature for and against treatment allocation rules that provide some balance over covariates, as do Shao et al. (2010).

There is an appreciable literature on the analysis of data from clinical trials when the responses are observed at a series of monitoring times, for example Everitt and Pickles (2004, Chapters 5–7). Moltenberghs et al. (2004) describe data from three clinical trials of anti-depressants in which the responses can be treated as continuous. Galbraith and Marschner (2002) provide guidelines for designing non-adaptive longitudinal clinical trials.

By comparison there is very limited literature on the design of adaptive longitudinal trials. Biswas and Dewanjii (2004b) describe a trial of pulsed electro-magnetic field therapy in which each patient was monitored for about 16 weeks. The original responses in this trial had a complicated multivariate structure, which was ignored in the design. Instead a binary variable ‘recurrence’ was used. Biswas and Dewanjii developed an urn design for longitudinal binary responses, which is a modification and simple extension of the RPW design where the covariates were ignored. See also Biswas and Dewanjii (2004a, c). Sutradhar et al. (2005) used a similar urn model based design and allowed for the possibility of time-dependent covariates. Subsequently, Sutradhar and Jowaheer (2006) extended this approach for longitudinal count data. Biswas et al. (2012) provided an optimum response-adaptive design for longitudinal binary responses. Atkinson and Biswas (2014, Chapter 5) provide an account of work on response-adaptive designs for longitudinal responses. Further, Huang et al. (2013) proposed a general framework for longitudinal covariate-adjusted response-adaptive randomization procedures, and studied the related asymptotic properties.

In contrast, we obtain optimum biased-coin designs for multivariate and longitudinal responses by the extension of methods for univariate responses. The optimum designs in both cases are functions of the information matrix for the observations. The model for multivariate data is introduced in Section 2.1. In the rest of Section 2 we explore the consequences of a general formulation for randomized response-adaptive designs for univariate or multivariate responses. These designs use optimum design theory to provide covariate balance in a general adaptive rule that skews allocation to the better treatments, whilst maintaining a controllable degree of randomness. We stress that these results are extremely general; to apply the rules we merely need to be able to provide the information matrix of the observations. Loss and bias, used to compare the designs, are presented in Section 3 with the information matrix for longitudinal designs explicitly presented in Section 4.

Four specific allocation rules are described in Section 5. These include the extension of the rule of Atkinson and Biswas (2005a), which achieves adaptivity through use of the link function of Bandyopadhyay and Biswas (2001), and our new rule. We take the particular form of this rule which targets specified proportional allocations to the ranked treatments: in our numerical example with two treatments, our target is that 80% of patients should be allocated to the unknown better treatment. This procedure overcomes the instability in early allocations with the link-function based rule that can lead to imbalance if the trial is stopped early. In our application we apply the results to the particular information pattern and covariance structure arising with longitudinal responses developed in Section 4. The numerical results are in Section 7.

Two main contributions of our paper are the provision of our general rule and its application to longitudinal trials. In the form we use here, the design ceases to be response-adaptive once the correct ordering of the treatments has been established. We can then extend standard results of the effect of randomization on inference (Burman, 1996; Atkinson, 2002) to multivariate designs. For longitudinal designs with correlated observations we define an effective number of observations that permits calculation of the loss from randomization. This important quantity indicates the average number of patients on whom information is lost due to a particular randomization rule. Simulations in Section 7 confirm the accuracy of this definition.

The methods of optimum experimental design are central to our construction of allocation rules. In Appendix A.1 we develop new results on multivariate $D_{A}$-optimality that allow us to estimate linear combinations of the treatment effects, such as differences, in the presence of the parameters associated with the prognostic factors over which we are
balancing. An equivalence theorem satisfied by the optimum designs is in Appendix A.2. These contributions are discussed in Section 8.

2. Multivariate data

2.1. Models

Patients arrive sequentially and are to be allocated one of \( t \) treatments. The particular treatment to be allocated to patient \( n + 1 \) depends on a vector of prognostic factors \( x_{n+1} \), on previous allocations and on the information available from the responses of previous patients. In a longitudinal study this information will increase as extra readings become available on earlier patients. But we start with the simpler multivariate case of \( n_h \) responses from each patient, all of which are available before the next patient arrives. In Section 4 we extend the model to the longitudinal case of incomplete time series of observations.

We assume that the results of the trial, perhaps after data transformation, will be analysed using a regression model which, for the \( n_h \) readings on patient \( i \), is written

\[
E(y_{i}) = F_i \beta = H_i \alpha + Z_i \theta + G \zeta,
\]

where \( y_{i} \) is \( n_h \times 1 \). Although we take all observations to have the same dimension, there is no difficulty, other than notational, in \( y_{i} \) being of dimension \( n_h(i) \).

Here \( \alpha \) is the \( t \times 1 \) vector of treatment effects that are the focus of inferential interest and \( H_i \) is the \( n_h \times t \) matrix of \( t \) indicator variables, the one non-zero column indicating which treatment the patient received. The \( n_h \times v \) matrix \( Z_i \) contains those covariates, including any powers or interactions of the elements of \( x_i \), which may be used to adjust the responses when estimating \( \alpha \). Because of the way we have parameterized the treatment effects, \( Z_i \) does not include a constant column. In the context of longitudinal data the \( n_h - 1 \) elements of \( \zeta \) are arbitrary period effects, the same for all patients; the \( n_h \times (n_h - 1) \) matrix \( G \) is the matrix of indicator variables for the period. These matrices are highly structured. All rows of \( H_i \) are the same for patient \( i \), as are those of \( Z_i \), whereas the rows of \( G \) are different.

Conditionally on the values of the \( x_i \), the additive errors of observation in (1) have an \( n_h \) dimensional normal distribution with covariance matrix

\[
\text{cov} y_{i} = \sigma^2 V.
\]

For longitudinal data \( V \) has a known structure; in our numerical example this comes from an AR(1) process of errors. Efficient estimation is by weighted least squares, with a block diagonal weight matrix with diagonal elements \( V^{-1} \). The conditional information matrix for all \( n \) patients is

\[
I(n) = \sum_{i=1}^{n} F_i^T V^{-1} F_i.
\]

The design criterion does not depend on the value of \( \sigma^2 \), which is suppressed in (3). However, \( \sigma^2 \) is necessary in power calculations. In response adaptive designs the value of \( x_i \) is determined by previous values of \( y_i \), so that, unconditionally, the observations are not independent. In Section 6 we give the argument that (3) is, however, the correct unconditional asymptotic information matrix.

2.2. Parameter estimation and randomization

To construct adaptive designs for the efficient estimation of \( \alpha \) in (1) we employ randomized versions of the sequential construction of optimum experimental designs (Fedorov, 1972; Atkinson et al., 2007; Fedorov and Leonov, 2014). The cost of randomization is that the design is likely to be unbalanced when it is stopped after a number of patients that is unknown at the planning stage; there is a consequent loss of efficiency in estimation.

In order to balance parameter estimation and randomization in general non-sequential designs, Ball et al. (1993) suggested finding designs to maximize the utility

\[
U = U_V - \gamma U_R,
\]

where the contribution of \( U_V \) is to provide estimates with low variance, whereas \( U_R \) provides randomness. The parameter \( \gamma \) provides a balance between the two. We extend their method to the sequential construction of longitudinal clinical trials.

With \( \pi_j \) the probability of allocating treatment \( j \), let

\[
U_V = \sum_{j=1}^{t} \pi_j \phi_j,
\]

where \( \phi_j \) is a measure of the information from applying treatment \( j \). In the next section this is defined in terms of D_{A}-optimality.
Ball et al. (1993) are interested in randomness with equal allocation, when

$$U_R = \sum_{j=1}^{t} \pi_j \log \pi_j.$$ 

To combine randomness with greater allocation to preferred treatments we introduce a set of gains $G_1, \ldots, G_t$ for the allocation of the individual treatments. These gains can be quite general, although we require $G_i \geq 0 \ \forall \ i$. Then

$$U_R = \sum_{j=1}^{t} \pi_j (-G_j + \log \pi_j).$$

(5)

In Section 2.5 we associate the allocation of the individual treatments. These gains can be quite general, although we require $G_i \geq 0 \ \forall \ i$. Then

$$U_R = \sum_{j=1}^{t} \pi_j (-G_j + \log \pi_j).$$

To maximize the utility (4) subject to the constraint $\sum_{j=1}^{t} \pi_j = 1$ we introduce the Lagrange multiplier $\lambda$ and maximize

$$U = \sum_{j=1}^{t} \pi_j \phi_j - \gamma \sum_{j=1}^{t} \pi_j (-G_j + \log \pi_j) + \lambda \left( \sum_{j=1}^{t} \pi_j - 1 \right).$$

(6)

Since the $G_i$ occur in $U$ with a positive coefficient, maximization of $U$ gives large values of $\pi_j$ for treatments with larger $G_j$. Differentiation of (6) with respect to $\pi_j$ leads to the $t$ relationships

$$\phi_j - \gamma (-G_j + 1 + \log \pi_j) + \lambda = 0,$$

so that all quantities

$$\phi_j / \gamma + G_j - \log \pi_j$$

must be constant. Since $\sum_{j=1}^{t} \pi_j = 1$, we obtain

$$\pi_j = \{\exp(\phi_j / \gamma + G_j)\} / S = \{\exp(\psi_j / \gamma)\} / S,$$

(7)

where

$$\psi_j = \phi_j + \gamma G_j$$

and

$$S = \sum_{j=1}^{t} \exp((\phi_j / \gamma) + G_j) = \sum_{j=1}^{t} \exp(\psi_j / \gamma).$$

As $\gamma \rightarrow \infty$, emphasis is solely on randomization and now all quantities $G_j - \log \pi_j$ must be constant so that $\pi_j \propto \exp(G_j)$. When the gains $G_i$ for the allocation of the individual treatments are equal, we obtain the equal randomization rule of Ball et al. (1993) with $\pi_j = 1/t$. This also follows directly from (5) since, with $G_j = G \ \forall \ j$. $\Sigma_j \pi_j G_j = G$, which does not affect the allocation.

2.3. Optimum and sequential designs

The probabilities of allocation $\pi_j$ in (7) depend on the information measure $\phi_j$ and on the utility $G_i$ from allocating treatment $j$. We first consider $\phi_j$.

In the methods of optimum experimental design, treatments are allocated to make large some function of $I(n)$. We allocate to minimize the variance of $s$ linear combinations of the treatment estimates which are adaptively chosen to give the desired probability of allocation of each treatment. For this measure we follow Atkinson (1982) and use $D_A$-optimality (Atkinson et al., 2007, §10.2).

The $s$ linear combinations of the treatment effects are $L^T \alpha$, where $L^T$ is $s \times t$, $s < t$. The $s$ combinations of all parameters can be written

$$A^T \beta = L^T \alpha + W_1^T \theta + W_2^T \zeta,$$

(8)

where $A^T$ is $s \times (t + v + n_0 - 1)$, $W_1^T$ is $s \times v$ and $W_2^T$ is $s \times (n_0 - 1)$. If the effects of the variables $Z_i$ and $G$ are not of interest in themselves, the parameters $\theta$ and $\zeta$ in (1) become nuisance parameters and the elements of $W_1$ and $W_2$ are zero. For any $A$ the variance of the estimated combination of coefficients is

$$\text{var}[A^T \hat{\beta}] = \sigma^2 A^T (L(n))^{-1} A,$$

(9)

where $\hat{\beta}$ is the least squares estimate of $\beta$.

The properties of the design depend on the number of treatments and on the dimension of the space of the nuisance parameters over which allocations are randomized. The exact relationship depends on the form of randomization.
Atkinson (2014) provides many examples. There is no randomization over the values of the time profile G. Since, in (8), $W_l^T$ is $s \times v$, the dimension of the nuisance parameters is $q = t + v - s$.

$D_A$-optimum designs minimize the logarithm of the determinant of the covariance matrix (9). Thus we seek designs to maximize the information measure

\[ \phi = - \log |A^T (I(n))^{-1} A| = - \log \Psi. \quad (10) \]

If treatment $j$ is allocated to the $(n + 1)$st patient we extend the notation of (10) and obtain

\[ \phi_j = - \log |A^T (I(n + 1, j))^{-1} A| = - \log \Psi_j. \]

Once the allocation has been made, we can suppress the subscript $j$ which is, however, required when we are comparing treatments for allocation. Substitution of this expression for $\phi_j$ in (7) yields

\[ \pi_j = \Psi_j^{-1/\gamma} \exp(G_j)/S. \quad (11) \]

In the sequential generation of optimum designs we would make the allocation for which $\phi_j$ was a maximum.

For clinical trials with univariate responses Atkinson (1982) exploited the results on sequential generation of $D_A$-optimum designs from Silvey (1980). Here we have multivariate responses. The requisite extension of $D_A$-optimality is derived in Appendix A.1, with the Equivalence Theorem for multivariate $D_A$-optimality in Appendix A.2. It is clear from (A.1) that $\Psi_j$ is the product of two terms, one of which is the same for all allocations. Substitution into (11) yields the allocation probability

\[ \pi(j|x_{n+1}) = \frac{\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma} \exp(G_j)}{\sum_{i=1}^{l} \{1 + d_A(i, n, x_{n+1})\}^{1/\gamma} \exp(G_i)}, \quad (12) \]

where $d_A(j, n, x_{n+1})$, the directional derivative of the $D_A$-optimality criterion, is given by (A.3) and $x_{n+1}$ is the vector of covariates for the new patient that are included in $I(n + 1, j)$.

2.4. Gain and allocation probabilities

We have derived our very general allocation rule in terms of undefined gains $G_i$ from allocation of the treatment $j$. We now find appropriate $G_i$ for an allocation rule which targets proportions of adaptively ranked treatments.

Let the target proportion of patients receiving treatment ranked $j$ be $p^*_j$. Then we require that

\[ p^*_1 \geq p^*_2 \geq \cdots \geq p^*_j \geq \cdots \geq p^*_t, \quad (13) \]

with, to avoid uniform allocation, at least one inequality. With $p^*_j = 1$ (and all other $p^*_j = 0$) we obtain a rule in which only the most highly ranked treatment is allocated. Plausible rules allow allocation to all treatments that have not been eliminated from the study and have the $p^*_j$ a decreasing function of $j$. The purpose of the rule is to ensure a specified ethical gain without going through possible extreme designs even if, for example for a two treatment design, $\alpha_1$ is very much greater than $\alpha_2$. By using ranks, we ensure both a prefixed allocation which is ethically skewed and sufficient allocation to each treatment to ensure that the design is not too inferentially inefficient.

At the optimum design it follows from the Equivalence Theorem of Appendix A.2 that all $d_A(j, n, x_{n+1})$ are equal and the treatments are correctly ordered. Let the correct, but unknown, rank of treatment $j$ be $R(j)$. Then, from (12)

\[ \pi(j|x_{n+1}) = \frac{\exp(G_{R(j)})}{\sum_{i=1}^{l} \exp(G_{R(i)})}. \quad (14) \]

The probabilities of allocation in (12) and (14) are unaltered if we replace $G_{R(j)}$ with

\[ G_{R(j)}^c = G_{R(j)} + c. \]

We choose $c$ so that $\sum_{i=1}^{l} \exp(G_{R(i)}^c) = 1$. Then (14) becomes

\[ G_{R(j)}^c = \log p^*_R(j) \]

and the allocation probabilities (14) have the simple form

\[ \pi_G(j|x_{n+1}) = \frac{\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma} p^*_R(j)}{\sum_{i=1}^{l} \{1 + d_A(i, n, x_{n+1})\}^{1/\gamma} p^*_R(i)}. \quad (15) \]

provided the ranking of the treatments is known. In designing the trial, the $p^*_j$ are the fundamental quantities which are to be specified, rather than the gains $G_j$.  

2.5. Skewed allocations

Replacement of the ranks \( R(j) \) in (15) with the ranks \( \hat{R}(j) \) based on the estimated ordering of the treatment effects \( \alpha \) gives an operational rule, but we have also to specify the coefficients \( l \) that give an efficient design. We develop the argument for \( s = 1 \), when the properties of a linear combination \( l^T \alpha \) are of interest. First consider univariate responses, variance \( \sigma^2 \), with a proportion \( p_j = n_j/n \) of the \( n \) patients receiving treatment \( j \). When there are two treatments and in (8) we take \( l^T = (1 \ -1) \), or equivalently \((0.5 \ -0.5)\), the inferential purpose is to estimate \( \alpha_1 - \alpha_2 \) with minimum variance. The design minimizing (9) provides balance over the covariates and equal allocation to the two treatments, so that \( r_1 = r_2 = 0.5 \).

To obtain skewed allocation for \( t \) treatments combined with efficient parameter estimation we find designs for estimation of the linear combination with

\[
l^T \alpha = \pm l_1 \alpha_1 \mp \cdots \pm l_t \alpha_t,
\]

where the coefficients \( l_j \), \( j = 1, \ldots, t \) are such that \( 0 < l_j < 1 \) and \( \sum l_j = 1 \). Then, in the absence of covariates,

\[
\text{var} \{l^T \hat{\alpha}\} = \left(\frac{\sigma^2}{n}\right)\sum_{j=1}^{t} \frac{l_j^2}{r_j} \quad \text{with} \quad \sum_{j=1}^{t} r_j = 1.
\]

Use of a Lagrange multiplier shows that this variance is minimized when the proportion of patients receiving treatment \( j \) is \( l_j \), as it is when the design is balanced across treatments, in the sense of the covariates having the same conditional distribution for each treatment. The signs in (16) are a generalization to \( t \) treatments of the weights 0.5 and \(-0.5\) that give efficient designs with \( r_j = 0.5 \) for the treatment difference.

To obtain an adaptive design targeting (13) we take weights

\[
l_j = p_{R(j)}^5.
\]

2.6. Adaptive design: Rule G

With estimated rankings the probability of allocation of treatment \( j \) for Rule G is

\[
\pi(j|x_{n+1}) = \frac{\{1 + d(s, n, x_{n+1})\}^{1/\gamma} p_{R(j)}^*}{\sum_{s=1}^{t} \{1 + d(s, n, x_{n+1})\}^{1/\gamma} p_{R(s)}^*}.
\]

The effect of different values of the parameter \( \gamma \) can be elucidated by simulation, as in Atkinson (2014) for univariate responses. In our example we take \( \gamma = 0.1 \). Distributional results for Rule G are in Section 6.

It is a characteristic of this scheme that the probability of allocating the treatments depends on the \( p_j^* \) and on the ordering of the \( \alpha_j \), but not on the differences between them. Suppose there are two treatments. Then, if \( \alpha_1 > \alpha_2 \), treatment 1 will eventually be allocated in a proportion \( p_1^* \) of the trials regardless of the value of \( \Delta = \alpha_1 - \alpha_2 \). Of course, if \( \Delta \) is small relative to the measurement error, in many of the initial trials, \( \hat{\alpha}_1 < \hat{\alpha}_2 \) and it will seem that treatment 2 is better. Then some individual allocations will be skewed in favour of treatment 2 with target \( p_2^* \), that is \( p_{R(2)}^* \). When \( \hat{\alpha}_1 > \hat{\alpha}_2 \), treatment 1 will be preferred. If the trial is terminated before a clear difference between the treatments has been established, each treatment may have been allocated to around half the patients.

3. Loss, bias and the assessment of designs

3.1. Loss

Our adaptive designs have some randomness in allocation. The effect of randomness is slightly to unbalance the designs while reducing the chance of guessing the next allocation. To compare designs we need measures of these two aspects.

We start with the effect of imbalance and extend the idea of loss (Burman, 1996) to multivariate responses. For univariate responses the variance of the estimated linear combination (9) has a minimum value of \( \sigma^2/n \) for the optimum design with proportions \( l_j \) and balance across the covariates. For multivariate data with \( n_h \) independent observations on each patient, the minimum variance is \( \sigma^2/(n n_h) \). However, in general the variance depends on the correlation of the multivariate observations.

For observations with structure given by (1), the contribution of \( y_i \) to the total sum of squares is \( y_i^T V^{-1} y_i \). For positively correlated data this is a smaller contribution than if the observations were independent. We define the effective number of observations as

\[
n_{\text{effect}} = l^T V^{-1} l,
\]

where \( l \) is a vector of ones.
The effective number of observations \( n_{\text{effec}} \) depends on the structure of \( V \). For independent observations \( n_{\text{effec}} = n \). For correlated observations it decreases with increasing correlation. With this definition the variance of the estimated linear combination for the optimum design has the minimum value

\[
\text{var} \{ T \hat{\alpha} \} = \sigma^2 / (n \times n_{\text{effec}}),
\]

(20)

where \( \hat{\alpha} \) is the least squares estimate of \( \alpha \) from the optimum design.

We can find from (9) the variance of the same linear combination for any other design. The efficiency of the design is then

\[
E_n = 1 / \left[ n \text{var} \{ T \hat{\alpha} \} \right].
\]

The loss \( L_n \) is defined on comparing the variance (9) with the minimum value given by (20) as

\[
\text{var} \{ T \hat{\alpha} \} = \frac{\sigma^2}{n \times n_{\text{effec}} - L_n}.
\]

Comparisons can use either the efficiency \( E_n \), or the loss, calculated by Atkinson and Biswas (2014) for numerous rules for skewed and unskewed allocations when the responses are univariate.

For all reasonable designs the efficiency tends to one as \( n \to \infty \). However, distinct limits of loss are known for several classes of design for univariate responses and comparisons of loss provide an incisive means of comparing designs for the quality of the estimates of the treatment effects. The number of such parameters does not depend on the dimension of the multivariate observations and, in the multivariate setting, the loss can be interpreted as \( n_{\text{effec}} \) times the number of patients on whom information is lost due to the lack of optimality of the design.

3.2. Selection bias

The purpose of including randomization in these rules is to prevent various kinds of bias. Selection bias occurs when the clinician is able correctly to guess the next treatment to be allocated. For two treatments it can be written as

\[
B_n = \left( \begin{array}{c}
\text{probability of correctly guessing the allocation to patient } n \\
- \text{probability of incorrectly guessing the same allocation}.
\end{array} \right)
\]

(21)

As do Heritier et al. (2005), we directly use the allocation probability \( \pi_n(j) \) for patient \( n \). The probability of correctly guessing the allocation of treatment \( j \) when \( \pi_n(j) \geq 0.5 \) is \( \pi_n(j) \) and of an incorrect guess is \( 1 - \pi_n(j) \). The selection bias can then be estimated by simulation as the average value of \( 2\pi_n(j) - 1 \). In simple cases the bias can be calculated explicitly. For example, in the non-randomized sequential construction of optimum designs the next treatment to be allocated is known exactly and the value of \( B_n \) is one.

4. Longitudinal designs

The main algebraic difference between multivariate designs and longitudinal designs comes from the reduced amount of information that is available from previous patients when allocation is made to patient \( n + 1 \). As an example, in our calculations we assume that patients arrive, or are grouped to arrive, in cohorts of size \( n_g \) (\( n_g \) can equal one). In the general case of \( n_g > 1 \) the various members of each cohort can be allocated different treatments.

Consider the first patient of cohort \( k + 1 \), so that \( n + 1 = kn_g + 1 \). We assume the responses are delayed, so that there is no information on the responses from cohort \( k \). Patients from cohort \( k - 1 \) contribute one response, those from cohort \( k - 2 \) two responses and so forth. Working backwards, the first cohort to contribute all \( n_g \) responses is number \( k - n_h \). Let \( S(i, n) \) denote this set of indexes for patient \( i \) as cohort \( k + 1 \) starts. Then the information matrix for the allocation of patient \( n + 1 \) is, in an extension of the notation of (3),

\[
I(n) = \sum_{i=1}^{n} F_{S(i, n)}^T V_{S(i, n)}^{-1} F_{S(i, n)}. \tag{22}
\]

The same process of counting applies to the sufficient statistics and so to the estimates of \( \alpha \) used in the adaptive allocation rules. For allocation of the remaining observations in cohort \( k + 1 \) we increment the information matrix by the complete value of \( F_{S(i, n)}^T V_{S(i, n)}^{-1} F_{S(i, n)} \) for each allocated observation. This is a temporary measure to aid balance, so that we remove these contributions when recalculating (22) for the next cohort.

The other difference between longitudinal data and the multivariate data of Section 2.1 is the structure of the covariance matrix \( V \). In our numerical example the errors form an AR(1) process. A stationary process can be simulated by generating \( u_1 = \varepsilon_1 \sim N(0, 1/(1 - \rho^2)) \) and, for \( i > 1 \), \( u_i = \rho u_{i-1} + \varepsilon_i \) where \( \varepsilon_i \sim N(0, 1) \). Then \( \text{var} u_i = 1/(1 - \rho^2) \) and \( \text{cov}(u_i, u_{i-j}) = \rho^j/(1 - \rho^2) \). The errors of observation will be \( \sigma u_i \), where \( \sigma \) is to be estimated.
For this error structure

\[ V^{-1} = \begin{pmatrix}
1 & -\rho & 0 & 0 & \cdots \\
-\rho & 1 + \rho^2 & -\rho & 0 & \cdots \\
0 & -\rho & 1 + \rho^2 & -\rho & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots \\
\end{pmatrix} \]

and, from (19)

\[ n_{\text{effec}} = J^T V^{-1} J = n_h - 2(n_h - 1)\rho + (n_h - 2)\rho^2. \]

For \( \rho = 0 \), \( n_{\text{effec}} = n_h \). The number however decreases appreciably as \( \rho \) increases; for \( n_h = 4 \) and \( \rho = 0.5 \), \( n_{\text{effec}} = 1.5. \)

5. Four allocation rules

We now present and compare four specific allocation rules, two of which depend directly on the estimated ranking of the treatments. As in Section 2.5, let the ranking of treatment \( j \) be \( R(j) \), estimated by \( \hat{R}(j) \) from the ranking of the estimated treatment effects \( \hat{\alpha} \).

1. **Rule D** (Deterministic). For the purposes of comparison with adaptive rules, we assume that the correct ordering of the treatments is known. There is no randomization and the rule is that for the sequential construction of the optimum design. We allocate that treatment for which (A.2) is a maximum:

\[ \pi_D(j|x_{n+1}) = \begin{cases} 1 & j = \text{arg} \max_{j=1,\ldots,t} d_A(j, n, x_{n+1}) \\ 0 & \text{otherwise.} \end{cases} \]

Simulations of designs for univariate responses mentioned above show that the loss \( L_q \) for this rule rapidly tends to zero and the bias \( B_n \) is one, since it is always known which treatment will be allocated next. These are extreme values; other rules have higher loss and lower bias.

2. **Rule RA** (Random and Response Adaptive). In this rule the treatments are allocated with probabilities \( p^{\ast}_{ij} \) introduced in (13) and based on the estimated ranking of treatment effects; there is no attempt at covariate balance. Then (18) reduces to

\[ \pi_{RA}(j) = p^{\ast}_{ij}. \]

The best guessing strategy is always to guess that the seemingly best treatment will be allocated. Asymptotically the probability of being correct is \( p^{\ast}_{ij} \) (and of being wrong \( 1 - p^{\ast}_{ij} \)), so that the limit of \( B_n \) in (21) is \( 2p^{\ast}_{ij} - 1 \). For univariate responses the asymptotic value of the loss is \( q \) (Cox, 1951; Burman, 1996).

3. **Rule G** (General with ranks). This rule (18) extends Rule RA to include covariate balance. Since \( d_A(j, n, x_{n+1}) \) in (18) is not standardized for \( n \), the rule tends asymptotically to Rule RA.

4. **Rule L** (Link). For two treatments the target probabilities depend on the estimated difference in treatment means \( \hat{\Delta} = \hat{\alpha}_1 - \hat{\alpha}_2. \) Atkinson and Biswas (2005a) use a link function to relate the \( I_1 \) to \( \hat{\Delta} \). Following Bandyopadhyay and Biswas (2001) they take \( I_1 = \Phi(\hat{\Delta}/T) \), where \( \Phi(\cdot) \) is the standard normal cumulative distribution function (c.d.f.). The value of \( I_1 \) may be greater or less than 0.5 and the parameter \( T \) controls the degree of skewing of the allocation.

For generality we present the \( t \) treatment version of this rule. After \( n \) patients have been treated the estimated treatment parameters are \( \hat{\alpha}_j \). To preserve the invariance of the procedure to the overall treatment mean let

\[ \tilde{\alpha} = \sum_{j=1}^{t} \hat{\alpha}_j/t \quad \text{and} \quad \tilde{\Delta}_j = \hat{\alpha}_j - \tilde{\alpha}. \]

The cumulative normal distribution provides coefficients \( l_j \) by setting

\[ l_j = I_j/S_j, \quad \text{where} \quad S_j = \sum_{k=1}^{t} I'_k. \]  

For \( t = 2 \) this reduces to the design procedure of Bandyopadhyay and Biswas (2001) except that the standard deviation \( T \) is replaced by \( 2T. \)

This design procedure does not depend on the values of the covariates. To provide a rule that is covariate adjusted, we use the values of \( I'_j \) in place of \( p^{\ast}_{R(j)} \) in (18) to give allocation probabilities

\[ \pi_L(j|x_{n+1}) = \frac{\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma} I'_j}{\sum_{s=1}^{t} \{1 + d_A(s, n, x_{n+1})\}^{1/\gamma} I'_s}, \]  

(24)
Use of $l_0^r$, or the standardized $l_0$ from (23), gives identical allocation probabilities $\pi_L(j|x_{n+1})$ since the summation $S_l$ cancels in (24). As for Rule G, the emphasis on covariate balance reduces as $n$ increases, the rule asymptotically reducing to that of Bandyopadhyay and Biswas (2001).

The result is a rule in which the targets vary more than in the three other rules of this section. When the estimated treatment difference is small the allocations are closer to 0.5 and when the treatment differences are over-estimated, the allocation probabilities are more extreme.

6. Distributional results

The conditional distribution of the responses in (1) is multivariate normal. But, for the adaptive design, the value of $F_i$ depends on the preceding values of $y_i$, so that, unconditionally, the observations are no longer independent. However, asymptotically, the least squares parameter estimates for this model have the same distribution as those from a non-adaptive design.

The crucial property is that the four rules considered in our paper are such that $\pi(j|x_{n+1}) > c$ for all $n$ and $j$, with $c$ some positive constant. Then, with $n_{j,n}$ the number of patients allocated to treatment $j$, $n_{j,n} \to \infty$ with $n$. Asymptotic results require that the information matrix (3) is such that $I_n/n$ tends to a limit. For known covariance matrix $V$ in (2) the form of the rules ensures that the information matrix for the treatment terms in (1) has a limit, as does the matrix for the non-stochastic period indicators $G$. The remaining requirement is that the distribution of the covariates $Z_i$ is well behaved. It then follows, as in Lai and Wei (1982), that the asymptotic information matrix is that for least squares from the conditional model (1) and that the parameters are asymptotically normally distributed. If the value of $\rho$ is not known, a consistent estimator of $\rho$ is required in (2), together with an estimate of $\sigma^2$. Then the asymptotic information matrix is $J(n)$ given by (22) for longitudinal designs.

As $n$ increases, the treatment parameters $\alpha$ become increasingly precisely estimated and eventually the treatments are always correctly ordered. Then Rules D, RA and G are no longer response adaptive, although they still adapt to the values of the covariates. The rules become the analogue, for multivariate responses, of the covariate adaptive designs with skewed allocations included in Atkinson and Biswas (2005a,b). The losses for these univariate designs are similar to those of the rules for unskewed rules compared by Atkinson (2002), although convergence to asymptotic values is slower for more skewed allocations. Simulations in Table 1 for longitudinal responses show that these univariate results are also a good guide to the properties of the designs of this paper.

The value of loss indicates the effective reduction in sample size due to randomization. This value can then be used in power calculations to assess the effect of the randomization on average power. A different approach to calculation of the effect of randomization on power is that of Hu and Rosenberger (2003) who derive a relationship between power and the power calculation to assess the effect of randomization on average power. Under such conditions, a procedure with higher loss will in general have a higher variance of $n_{j,n}$. For random allocation to treatment $j$ with probability $p_j^n$,

$$\sqrt{n}\{\text{var}(n_{j,n}/n) - p_j^n\} = p_j^n(1 - p_j^n),$$

which equals 1/4 for unskewed randomization. Zhang et al. (2007, Remark 3.3) give the term for the additional variance above that of (25) for the rule of Bandyopadhyay and Biswas (2001) which depends on $\alpha_j$ and on $T$ in (23). We can therefore expect that asymptotically Rule L will have a higher loss than the other rules of Section 5.

7. Numerical results

We now present simulation results for the four rules of Section 5 when there are two treatments. The ethical goal is to allocate 80% of the patients to the better treatment, taken as treatment 1. With $p_{j,n}^* = 0.8$, the value of the bias $B_n$ tends to $2p_{1,n}^* - 1 = 0.6$ for large $n$. In these simulations we take cohorts of size 3 ($n_L = 3$) with $n_L = 4$. In (1)

$$\sigma^2 = 1, \Delta = \alpha_1 - \alpha_2 = 0.5$$

and there are four prognostic factors with independent standard normal distributions, so that $q = 5$. In (18) three values of $\gamma$, 0.1, 0.01 and 0.001 are taken, both for Rules L and G, so that in all eight rules are compared. For designs with univariate responses which are not response adaptive, smaller values of $\gamma$ initially force appreciable balance like Rule D, taking longer to move towards random allocation as $n$ increases. The same pattern is shown here. For Rule L we take $T = 0.5941$, giving the required value of 0.8 for $\Phi(\Delta/T)$. Each rule was simulated 10,000 times. All designs were regularized to avoid extreme designs by ensuring that the minimum number of allocations of each treatment was never below $\sqrt{n}$ nor the maximum above $n - \sqrt{n}$.

Table 1 gives results for $n = 48$ and 192, both multiples of $n_L$, for three values of $\rho$. The behaviour of loss and bias as $n$ increases is shown in Figs. 1 and 2.

The left-hand third of the table contains results for the average losses $\bar{L}_{48}$ and $\bar{L}_{192}$ for the three values of $\rho$. When $n = 48$ the values for Rule D (non-adaptive sequential design construction) are little affected by $\rho$. However, when $n = 192$ the loss for $\rho = 0$ decreases to 0.27 in line with the known limit of 0 as $n \to \infty$. The decrease with $n$ is less dramatic for the other
Table 1
The four rules of Section 5: average loss, bias and proportion of total allocations to treatment 1 when (a) \( n = 48 \) and (b) \( n = 192 \). Three values of \( \gamma \) each for rules G and L.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( n = 48 )</th>
<th>Loss ( \bar{L}_{n} )</th>
<th>Bias ( \bar{B}_{n} )</th>
<th>Total proportion ( t_{1,n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D )</td>
<td>2.42</td>
<td>2.49</td>
<td>2.73</td>
<td>1.00</td>
</tr>
<tr>
<td>( RA )</td>
<td>6.63</td>
<td>7.53</td>
<td>8.95</td>
<td>0.59</td>
</tr>
<tr>
<td>( G(0.1) )</td>
<td>3.94</td>
<td>4.88</td>
<td>6.45</td>
<td>0.63</td>
</tr>
<tr>
<td>( G(0.01) )</td>
<td>2.54</td>
<td>3.51</td>
<td>5.14</td>
<td>0.76</td>
</tr>
<tr>
<td>( G(0.001) )</td>
<td>2.55</td>
<td>3.47</td>
<td>5.08</td>
<td>0.96</td>
</tr>
<tr>
<td>( L(0.1) )</td>
<td>4.58</td>
<td>5.45</td>
<td>6.79</td>
<td>0.64</td>
</tr>
<tr>
<td>( L(0.01) )</td>
<td>3.68</td>
<td>4.17</td>
<td>5.42</td>
<td>0.82</td>
</tr>
<tr>
<td>( L(0.001) )</td>
<td>3.22</td>
<td>4.08</td>
<td>5.24</td>
<td>0.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 192 )</th>
<th>Loss ( \bar{L}_{n} )</th>
<th>Bias ( \bar{B}_{n} )</th>
<th>Total proportion ( t_{1,n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D )</td>
<td>0.27</td>
<td>1.35</td>
<td>2.40</td>
</tr>
<tr>
<td>( RA )</td>
<td>5.40</td>
<td>6.77</td>
<td>8.90</td>
</tr>
<tr>
<td>( G(0.1) )</td>
<td>3.96</td>
<td>5.31</td>
<td>7.40</td>
</tr>
<tr>
<td>( G(0.01) )</td>
<td>1.74</td>
<td>2.96</td>
<td>4.73</td>
</tr>
<tr>
<td>( G(0.001) )</td>
<td>0.43</td>
<td>1.60</td>
<td>3.24</td>
</tr>
<tr>
<td>( L(0.1) )</td>
<td>6.87</td>
<td>9.90</td>
<td>13.92</td>
</tr>
<tr>
<td>( L(0.01) )</td>
<td>4.85</td>
<td>8.06</td>
<td>12.12</td>
</tr>
<tr>
<td>( L(0.001) )</td>
<td>3.20</td>
<td>6.46</td>
<td>10.86</td>
</tr>
</tbody>
</table>

Fig. 1. Correlation \( \rho = 0 \). Left-hand panel: Average loss \( \bar{L}_{n} \) as a function of patient number \( n \). Reading down at \( n = 200 \), Rules L(0.1), RA, L(0.01), G(0.1), L(0.001), G(0.01), G(0.001), D. Right-hand panel: Average smoothed bias \( \bar{B}_{n} \). Reading down at \( n = 70 \), Rules D, L(0.001), G(0.001), L(0.01), G(0.1), L(0.1), R.

values of \( \rho \). For Rule RA (random allocation) there is no balancing over covariates. For \( n = 48 \) the loss is above 5 as it is for \( n = 192 \); for \( \rho = 0 \), the value of 5.40 is closest to that of \( q \), which would be the loss if the ordering of the treatments were known. The loss increases with \( \rho \), since the effective number of observations decreases, making the ordering of treatments more variable.

The losses for Rule G show the strong effect of the value of \( \gamma \), particularly when \( n = 192 \) and \( \gamma = 0.001 \). However, for \( \gamma = 0.1 \), there is a slight increase in loss with \( n \) as allocation becomes closer to that for Rule RA. For Rule L the loss decreases with \( \gamma \) although it increases going from \( n = 48 \) to 192 giving values significantly larger than those for Rule G with the same value of \( \gamma \), with a maximum value of 13.92 when \( \rho = 0.5 \) and \( \gamma = 0.1 \). As we described in Section 5, this design remains response adaptive and the larger values of loss reflect the effect of the estimated treatment difference \( \Delta \) on the values of \( l_{1} \) and \( l_{2} \).

The central third of the table gives the average biases for \( n = 48 \) and 192. The values agree with the theory of Section 3.2. For Rule D the bias is one for both values of \( n \); for Rule RA the values are 0.59, 0.60 or 0.61, close to \( 2p_{1}^{*} - 1 \). When \( n = 48 \) and \( \gamma = 0.1 \) the biases for Rule G and L are a little higher at 0.63 or 0.64. Both rules show that the effect of decreasing \( \gamma \) is to increase bias, strongly for \( n = 48 \), although much less so for \( n = 192 \), a feature clearly shown in the plots of Figs. 1 and 2.

The right-hand side of the table gives the average allocation to treatment 1, \( \bar{r}_{1,n} \). The regularization of the total number allocated to any treatment to be at most \( n \) precludes extreme allocations for \( n \) small so that all rules initially approach the overall target of 0.8 from below; Rule D is closest to the target when \( n = 48 \) and Rule RA the furthest from it. For Rule
The calculated values of $\hat{p}_1$ can be greater than 0.8. The entries in the table for $n = 192$ show this effect for small $\gamma$, when convergence to a proportion of 0.8 is even faster than for Rule D.

Figs. 1 and 2 provide further insight into the dependence of loss and bias on the value of $n$. Results for $\rho = 0$ are in Fig. 1. Apart from L all rules ultimately target 0.8. The upper limit of loss, as $n$ increases, is that for Rule RA which gradually decreases to $q = 5$. The Loss is bounded below by results for Rule D, which gradually decrease to zero. The losses for Rule G are smallest for $\gamma = 0.001$.

Rules with small loss generally have high bias, a phenomenon shown in the right-hand panel of the figure. The highest bias is one for deterministic design construction, Rule D, and the lowest close to 0.6 for Rule RA. The bias for Rule G is highest for the smallest value of $\gamma$, which is the value giving the lowest loss. Rule L behaves relatively poorly in these comparisons; loss is always higher than that for Rule G with the same value of $\gamma$, as is the bias, except for $\gamma = 0.1$, when there is little to choose between the rules.

These results for uncorrelated observations are close to those for individual patients exhibited in Atkinson (2014). The results in Fig. 2 for $\rho = 0.5$ have a similar structure to those of Fig. 1 but the numbers are slightly different. The values of loss for Rules RA, G and D do not decline so fast from the initial value and, at $n = 200$ are 2–3 higher than those in the uncorrelated case. The loss for Rule L increases more rapidly with $n$, again always being above that for Rule G with the same value of $\gamma$. The bias in the right-hand panel, apart from that for Rule D, again decreases to 0.6, but more gradually than in the uncorrelated case.

The advantage of Rule L, as is shown in Table 1, is that the proportion of patients receiving the better treatment converges more rapidly to 0.8 than for the other rules. The cost is the higher loss arising from the estimation of the target proportion, rather than its convergence to the given value of 0.8. In practice, an exact allocation of 0.8 is not likely to be important and the choice of rule should be based on loss and bias. Atkinson (2014) gives plots, for rules for individual patients, of loss against bias as $n$ increases. A rule with values of both quantities below those for all other rules at a particular $n$ is called “admissible”. The results of our comparisons show that Rule G is admissible when compared to Rule L. The choice of $\gamma$ depends on the relative importance of bias and loss.

Finally we consider the power of the $t$-test for equality of the treatment means. This is a maximum for equal allocation to the two treatments with balance over the covariates; skewing the design causes a reduction in power, as do increasing values of $\rho$, which effectively reduce the number of observations. We calculate the $t$-statistic using the elements of $I_n$ to allow for any correlation in the estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Since very small numbers of patients are not of interest in these simulations, power is assessed by counting the number of $t$-statistics that are greater than 1.96. Fig. 3 shows the logits of the powers for Rules D and L(0.1) as a function of $\rho$ for three values of $n$. The values for the six other rules of Table 1 lie between these two and are not shown. For lower values of $\rho$ Rule L has higher power than Rule D as it often gives a more nearly balanced allocation. As $\rho$ increases the performance of Rule L is sometimes degraded by poor estimates of the treatment difference leading to occasional designs which are very unbalanced over treatments.

If randomization is required, Rule D is not appropriate. Then Rules G and L are to be preferred to Rule RA. However, the choice between the two sets of rules does not only depend on statistical properties. In our comparison we chose $\Delta/T$ such that $l_1 = 0.8$. If this is the clinician’s ideal skewing towards the better treatment, Rule G should be used. If, on the other hand, greater skewing is required for larger values of $\Delta$, then Rule L is appropriate. In either case the value of $\gamma$ will have to be chosen to balance loss against bias.
Fig. 3. Power of Rules D and L(0.1) for testing treatment difference; logit of proportion significant for the \( t \) test of size 0.05. The three pairs of lines, reading down, are for \( n = 198, 99 \) and 48. The upper limit of power shown in the plot is approximately 0.95. The effect of the correlation \( \rho \) on the effective sample size is evident.

8. Discussion

For homoscedastic regression models with \( t \) treatments giving univariate responses, the optimum allocation for testing hypotheses about the equality of the means of the treatments is to allocate a proportion \( 1/t \) of the patients to each treatment. However equal allocation is not always required. Dumville et al. (2006) and Peckham et al. (2015) review the use of unequal allocation ratios in clinical trials.

In such work it is assumed that both the target weights, and the treatments to which they apply, are known. Unequal allocation targets arise naturally for models in which the variances of the responses to the treatments are not the same. For example, Wong and Zhu (2008) extend the \( D_A \)-optimum designs of Atkinson (1982) to heteroscedastic models in which the variances differ between treatments. Atkinson (2015) gives details for two-treatment designs in the presence of covariates. Baldi Antognini and Giovagnoli (2015) describe compound optimum designs balancing between inference and allocation of as many patients as possible to the better treatment; the allocation targets for the various treatments may depend adaptively on parameters estimated from the responses to earlier allocations; longitudinal responses are not considered.

The linear contrast (16) provides a mechanism for skewed allocation for which \( \text{var} \{A^T \beta\} \) is a scalar. Although the exposition here is in terms of \( D_A \)-optimality, criteria such as \( A \)- or \( E \)-optimality will yield the same optimum design minimizing \( \text{var} \{A^T \beta\} \).

Our general rule (12) provides a family of potential treatment allocation schemes which are covariate adaptive. The particular choice of \( G_j \) that led to (18) provides a rule for longitudinal responses that has many properties in common with better understood rules for univariate responses. Despite the relative computational complexity of the counting and consequent treatment allocation algorithm described in Section 4, the loss and bias of the rule are straightforward analogues of those for univariate responses. In particular, the analogy of the values of loss is striking, but depends on the correct definition of \( n_{\text{effec}} \).

For small \( n \) Rule G forces skewed allocation and the loss is close to that of Rule D. But, as \( n \) increases the rule becomes increasingly like skewed random allocation, with a higher loss but with bias tending to zero. Rules for smaller values of the tuning constant \( \gamma \) have a higher initial emphasis on targeting the target skewing proportions \( p^*_j \). In the selection of a biased-coin design for clinical trials the emphasis in the statistical literature is often on trials that provide allocations very close to these targets. Atkinson (2012) stresses the importance of considering both loss and selection bias. Rule G is such that the bias decreases as \( n \) increases, but in such a way that the efficiency of estimation goes to 100\%. A rule with constant loss of \( q/5 \) can be obtained by simplifying (18) and taking the allocation probabilities proportional to \( R(j)d_A(j, n, x_{n+1}) \) (Burman, 1996).

Two final points. First we note that we have assumed a known value of \( \rho \). Indeed, the evaluation of the inferential properties of the design, particularly \( n_{\text{effec}} \) and the loss will depend strongly on \( \rho \). However, the dependence of the design itself on \( \rho \) is slight. We would recommend designing for an arbitrary value, such as 0.3, whilst sequentially estimating \( \rho \). Of course, the value of the estimate \( \hat{\rho} \) should be used in any inferences drawn from the results of the trial. Secondly, the methods may be extended to regression models with distributions other than the normal through the use of elemental information matrices as described in Atkinson et al. (2014).

Acknowledgements

We are grateful to the referees for comments which helped clarify the exposition of our results.
Appendix. Multivariate Dₐ-optimum designs

A.1. Sequential design construction

For the sake of generality we extend the single linear combination of the parameters $a^T \beta$ in (8) to the set of s combinations $A^T \beta$, where A is a $(t + v + n_h - 1) \times s$ matrix of known constants. Dₐ-optimum experimental designs for the linear regression model (1) maximize $|A^T \{I(n)\}^{-1}A|^{-1}$ and so minimize the generalized variance of these linear combinations, providing a normal theory confidence region of minimum volume. Such optimum designs can be constructed sequentially.

We first consider univariate responses when $f_i$ in (1) becomes the vector $f_i^T$, which includes the vectors of allocation and prognostic factors for the ith patient. When allocation is made to patient $n + 1$, all other allocations are known. A useful matrix result for D-optimum designs maximizing $|I(n + 1)|$ is that

$$ |I(n + 1)| = [1 + f_{n+1}^T \{I(n)\}^{-1}f_{n+1}] |I(n)| = 1 + d(j, n, n_{n+1}) |I(n)|. \quad (A.1) $$

That treatment is allocated for which $d(j, n, n_{n+1})$ is a maximum.

In the iterative construction of Dₐ-optimum designs for a univariate response,

$$ d_A(j, n, n_{n+1}) = f_{n+1}^T \{I(n)\}^{-1} A^T \{I(n)\}^{-1} A \{I(n)\}^{-1} f_{n+1}, \quad (j = 1, \ldots, t). \quad (A.2) $$

In the absence of randomization, patient $n + 1$ would receive the treatment for which $d_A(j, n, n_{n+1})$ is a maximum.

We now turn to multivariate data and find the equivalent of (A.2). Let the ith row of $F_{n+1}$ be denoted $f_{i,n+1}^T$. We extend (A.2) to

$$ d_{A_{uv}}(j, n, n_{n+1}) = f_{i,n+1}^T \{I(n)\}^{-1} A^T \{I(n)\}^{-1} A \{I(n)\}^{-1} f_{i,n+1}. \quad (A.2) $$

With element $u$, $v$ of $V^{-1}$ written $V^{uv}$, the equivalent of (A.2) is

$$ d_A(j, n, n_{n+1}) = \sum_{u=1}^{n} \sum_{v=1}^{n} V^{uv} d_{A_{uv}}(j, n, n_{n+1}). \quad (A.3) $$

This is the function in our generic rule (18).

A.2. An equivalence theorem

As $n \to \infty$, $d_A(j, n, n_{n+1}) \to 0$. If we replace the number of patients allocated to each treatment by the continuous distribution of asymptotic proportions of allocation we obtain a design measure $\xi$ and an information matrix $I(\xi)$. Also $d_A(j, n, n_{n+1})$ tends to the directional derivative $d_A(j, \xi, \eta)$. In the case of non-sequential design we can consider choosing treatments from a space $X$ and covariates from a space $Z$. The design region is then $X = f \times Z$. We now extend the General Equivalence Theorem of optimum design theory (Kiefer and Wolfowitz, 1960; Whittle, 1973) to multivariate Dₐ-optimality.

If we let

$$ \tilde{d}_A(\xi) = \sup_{j,x} d_A(j, x, \xi), $$

the equivalence theorem states that the Dₐ-optimum design, denoted $\xi^*$, is such that

$$ \tilde{d}_A(\xi^*_D) = s. $$

Here $s$ is the number of independent linear combinations of the parameters specified by A.

In the sequential construction of optimum designs for clinical trials there will asymptotically be balance for each treatment over the prognostic factors. With such balance we ignore $x_{n+1}$ and write

$$ d_A(j, \xi^*) = s, \quad (j = 1, \ldots, t). $$

The balanced design used in (20) to derive the expression for loss is $n_{\text{eff}} n\xi^*$. The results in Table 1 for Rule D show how fast the sequential construction of the optimum design converges.

References


Automatic generation of generalised regular factorial designs

André Kobilinsky\textsuperscript{a}, Hervé Monod\textsuperscript{a,\ast}, R.A. Bailey\textsuperscript{b,c}

\textsuperscript{a} MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France
\textsuperscript{b} School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, UK
\textsuperscript{c} School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

\textbf{A B S T R A C T}

The R package \textit{plano\textsc{r}} enables the user to search for, and construct, factorial designs satisfying given conditions. The user specifies the factors and their numbers of levels, the factorial terms which are assumed to be non-zero, and the subset of those which are to be estimated. Both block and treatment factors can be allowed for, and they may have either fixed or random effects, as well as hierarchy relationships. The designs are generalised regular designs, which means that each one is constructed by using a design key and that the underlying theory is that of finite abelian groups. The main theoretical results and algorithms on which \textit{plano\textsc{r}} is based are developed and illustrated, with the emphasis on mathematical rather than programming details. Sections 3–5 are dedicated to the elementary case, when the numbers of levels of all factors are powers of the same prime. The ineligible factorial terms associated with users’ specifications are defined and it is shown how they can be used to search for a design key by a backtrack algorithm. Then the results are extended to the case when different primes are involved, by making use of the Sylow decomposition of finite abelian groups. The proposed approach provides a unified framework for a wide range of factorial designs.

© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Factorial designs that may include several block and treatment factors date back to the pioneering work of Fisher and Yates at Rothamsted Experimental Station (Yates, 1933, 1937; Fisher, 1942), followed by Finney (1945) and Bose (1947). Since then, the construction of fractional designs has been a constantly active field of research in the theory of design of experiments. It has also been widely applied in many different application areas, including food research, biology, industry, and – more recently – computer experiments.

The designs we are interested in are known today as regular factorial designs. Their construction, which is based on algebra and group theory, gives a large class of orthogonal factorial designs, including as special cases block and row–column designs (Yates, 1937) as well as fractional designs (Finney, 1945). Regular fractional designs became a standard method of construction very early. Their main principles in standard cases have been explained in numerous text books including classics (Kempthorne, 1952; Cochran and Cox, 1957) and more recent ones (Ryan, 2007; Bailey, 2008; Morris, 2011; Cheng, 2014).

\footnote{This paper also includes Supplementary material providing proofs (Appendix A)(and additional information on the algorithms (Appendices B, C).}

\footnote{Correspondence to: Hervé Monod, MaIAGE bât. 210, INRA, Domaine de Vilvert, 78350 Jouy-en-Josas, France. Fax: +33 1 34 65 22 17.
E-mail address: herve.monod@inra.fr (H. Monod).}

\url{http://dx.doi.org/10.1016/j.csda.2016.09.003}
\url{0167-9473/© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).}
Of course, there are situations where non-regular factorial designs are needed. Examples include incomplete-block designs where the block size does not divide the number of treatments, and orthogonal arrays, such as those obtained from Hadamard matrices (Hedayat et al., 1999) and by computer intensive algorithms (Kuhfeld and Tobias, 2005) or those given by Bose and Bush (1952) and Addelman and Kempthorne (1961), whose number of experimental units does not divide the number of treatments. See Grömping and Hu (2014) and Grömping and Bailey (2016) for further discussion of the definition of regularity. However, regular designs still have wide applicability, and these are what the software planor deals with.

An important notion in fractional designs is resolution (Box and Hunter, 1961a,b). If \( R \) is a positive integer, a fraction of resolution \( R \) allows the estimation of all factorial effects up to interactions of order strictly smaller than \( R/2 \), assuming that all interactions of order strictly larger than \( R/2 \) are zero. More discriminating criteria such as minimum aberration (Fries and Hunter, 1980) or maximum estimation capacity (Cheng and Mukerjee, 1998) have been developed. The construction of optimal designs with respect to these criteria is still an active field of research, which includes analytical as well as algorithmic issues. However, resolution and aberration imply that the model of interest be symmetric with respect to all factors.

Most work in this area thus deals with problems which are highly symmetric with respect to the factors and models of interest. However, there is also a need to find generic and user-friendly methods of construction adapted to much more flexible problem specifications, allowing for unconstrained numbers of levels and flexible model assumptions.

The construction of a regular fraction involves two steps: first, finding defining relationships or generators of the fraction ensuring that factorial effects of interest will be estimable; second, generating the actual design. Despite its apparent simplicity in standard cases, the first step still represents a major challenge in general situations. It has been applied and programmed mainly in the case of symmetric designs, which require all factors to have the same number of levels or, at least, numbers of levels that are powers of the same prime.

Algorithms were developed and studied in the 1970s and 1980s (Patterson, 1965, 1976; Bailey et al., 1977; Franklin and Bailey, 1977; Patterson and Bailey, 1978; Bailey, 1985; Franklin, 1985) and some of them implemented in statistical packages (SAS, 2010; Payne, 2012; Grömping, 2014). Lewis (1982) tabulated generators for asymmetric factorial designs with resolution 3. This paper extends these approaches to a generalised class of regular designs. We follow the theoretical framework of Kobilinsky (1985) and Kobilinsky and Monod (1991, 1995), although we occasionally modify a term or notation for simplicity. We also make intensive use of pseudofactors (Monod and Bailey, 1992). The generalised class includes symmetric as well as asymmetric designs, with estimation planned in one or more strata. For example, our results apply to the construction of generalised split-plot or criss-cross designs as well as the more usual fractional designs. The method allows the user to define the model and specify what should be estimated. The approach and results have been implemented in the R package planor (Kobilinsky et al., 2012; Monod et al., 2012) based on the initial APL version by Kobilinsky (2005).

2. Overview of the design search issue

2.1. Examples

We start with three examples to illustrate the diversity of situations we want to consider, all based on real applications from our consulting experience.

We follow standard practice in using the same notation \( A \) for a factor and for its main effect, and the notation \( AB \) for the interaction between factors \( A \) and \( B \). Hierarchy relationships are denoted as in Bailey (2008, Chapter 10). In particular:

- \( B \leq A \) means that factor \( B \) is or must be nested in factor \( A \), in other words \( B \) must be finer than or equivalent to \( A \) and, reciprocally, \( A \) must be coarser than or equivalent to \( B \), so that each level of \( B \) occurs with a single level of \( A \);

- \( A \land C \) denotes the product or infimum of \( A \) and \( C \), which is the factor whose levels are the combinations of levels of \( A \) and \( C \).

Note the following two properties:

- \( A \land C \leq A \land C \leq A \land C \leq C \);

- if \( A \leq B \) and \( A \leq C \), then \( A \leq B \land C \).

**Example 1.** There are four treatment factors \( F_1, F_2, F_3, F_4 \) with 6, 4, 3, 4 levels respectively. A complete factorial design would require 288 experimental units but we assume this is much larger than possible, so that a smaller fractional design is looked for.

The experimenter intends to analyse the data using a model that consists of the general mean, the four main effects of \( F_1, F_2, F_3, F_4 \) and the interaction \( F_1F_2 \), all with fixed effects, with good reasons to consider the other interactions as negligible. The factorial terms he or she wants to estimate are the four main effects only. The factorial terms in the model and those that must be estimated are listed in the following sets \( \mathcal{M} \) and \( \mathcal{E} \):

\[
\mathcal{M} = \{ \mu, F_1, F_2, F_3, F_4, F_1F_2 \} \quad \text{and} \quad \mathcal{E} = \{ F_1, F_2, F_3, F_4 \} = \mathcal{M} \setminus \{ \mu, F_1, F_2 \}.
\]

where \( \mu \) denotes the general mean.
Example 2. A row-and-column design has to be constructed with two columns (factor C), three rows (factor R) and two units in each of the six cells defined by a row and a column. There are three treatment factors: two 2-level factors D, E and one 3-level factor A. The experimenter wants to estimate the interactions D,A, E,A, considering a model that includes the general mean, row, column and cell effects as well as all main effects of treatment factors and all interactions between two treatment factors. The sets \( \mathcal{M} \) and \( \mathcal{E} \) now are:

\[
\mathcal{M} = \{ \mu, \ C, \ R, \ C.R, \ D, \ E, \ A, \ D.A, \ E.A, \ D.E \} \quad \text{and} \quad \mathcal{E} = [D.A, \ E.A].
\]

An additional constraint is that, for practical reasons, the factor A must remain constant on each row. We call this a hierarchy constraint, imposing that factor R be nested in factor A, and denote it by \( R \preceq A \).

Example 3. The experimental units consist of four blocks, each containing two subblocks of four units. This structure can be described by three factors \( P, \ Q, \ U \) with four, two and four levels respectively. The levels of \( P \) define the blocks, the levels of the infimum \( P \land Q \) define the subblocks, and the levels of the infimum \( P \land Q \land U \) determine the units. In addition, there are four treatment factors \( A, \ B, \ C, \ D \) with two levels. There is again a hierarchy constraint: we assume that the levels of \( A \) cannot be varied between the four units of a given subblock, which is denoted by \( P \land Q \ll A \).

The experimenter is interested in the main effects and two-factor interactions of \( A, \ B, \ C, \ D \). Thus the set \( \mathcal{E} \) is

\[
\mathcal{E} = \{ A, \ B, \ C, \ D, \ A.B, \ A.C, \ A.D, \ B.C, \ B.D, \ C.D \}.
\]

In addition to \( \mathcal{E} \), the model must contain the block and subblock effects and the general mean. However, the hierarchy constraint means that the main effect of \( A \) is necessarily confounded with subblock effects so that it cannot be estimated in a model with fixed subblock effects.

If a proper randomisation is performed, block and subblock effects can instead be considered as centred random effects and the analysis can be decomposed into three levels of variability, or strata (Bailey, 2008), associated with \( P, \ P \land Q \) and \( P \land Q \land U \) respectively. The blocks stratum consists of all contrasts between blocks. The between-subblocks stratum consists of all contrasts between subblocks which are orthogonal to blocks. The bottom stratum, which includes residual variability only, consists of all contrasts orthogonal to subblocks.

The objective now is to estimate all effects except the main effect \( A \) in the bottom stratum, and to estimate the main effect \( A \) in the between-subblocks stratum. These requirements can be described by using two pairs of model-estimate lists, one for each stratum where an effect will be estimated. The first pair deals with the bottom stratum by pretending that \( P \) and \( P \land Q \) have fixed effects and omitting \( A \) from the effects to be estimated. Thus

\[
\mathcal{M}_1 = \{ \mu, \ P, \ Q, \ P.Q \} \cup \mathcal{E} \quad \text{and} \quad \mathcal{E}_1 = \mathcal{E} \setminus \{ A \}.
\]

The second pair deals with the between-subblocks stratum by pretending that block effects, but not subblock effects, are fixed, and declaring that the only effect to be estimated in this stratum is the main effect \( A \). Thus

\[
\mathcal{M}_2 = \{ \mu, \ P \} \cup \mathcal{E} \quad \text{and} \quad \mathcal{E}_2 = \{ A \}.
\]

2.2. Factorial terms and model specifications

Let \( F_1, \ldots, F_h \) denote all the genuine block and treatment factors involved in the experiment, that is, those which have a direct meaning for the experimenter. Denote by \( T \) the set of all \( n \) combinations of levels of \( F_1, \ldots, F_h \), with \( n = n_1 \cdots n_h \) where \( n_i \) is the number of levels of \( F_i \).

The column vector \( \tau \) of block and treatment effects belongs to the vector space \( \mathbb{R}^n \). In the standard analysis of variance (ANOVA), this vector space is decomposed into mutually orthogonal subspaces \( \mathcal{W}_i \) associated with the \( 2^h \) subsets of factors \( \{ F_i : i \in I \} \), for \( I \subseteq \{ 1, \ldots, h \} \). These subspaces are given by the recurrence relation

\[
\mathcal{W}_I = \mathcal{V}_I \cap \left( \bigoplus_{J \subseteq I} \mathcal{W}_J \right)^\perp,
\]

where \( \mathcal{V}_I \) is the subspace spanned by the all-one vector and \( \mathcal{V}_I \) is the subspace spanned by the indicator vectors of the level-combinations of all factors in \( \{ F_i : i \in I \} \) (Bailey, 2008). According to this decomposition, the effects can be decomposed into factorial effects, as given by the equation

\[
\tau = \sum_{I \subseteq \{1, \ldots, h\}} S_I \tau,
\]

where \( S_I \tau \) is the orthogonal projection of \( \tau \) onto \( \mathcal{W}_I \). By convention, the subsets \( \{ F_i : i \in I \} \) are called factorial terms and denoted by \( \mathcal{F}(I) \). In examples, as in Section 2.1, we follow a more usual practice for writing factorial terms, so that \( F_1,F_2 = \mathcal{F}(\{1, 2\}) \) and \( \mu = \mathcal{F}(\emptyset) \). The order of a factorial term \( \mathcal{F}(I) \) is given by the cardinality of the subset \( I \). Factorial terms of order 1 are called main effects, and factorial terms of order 2 or more are called interactions.
When constructing a fractional design it is necessary to assume that some factorial terms \( F_i \) are negligible, that is, that \( S_{\tau} \) is zero. In the examples above, the model set \( \mathcal{M} \) contains the non-negligible effects and its subset \( \mathcal{E} \) contains the non-negligible effects that the experimenter wants to estimate. In multi-stratum experiments, such as Example 3, it is necessary to consider several such pairs of model and estimate sets, one pair for each stratum in which any fixed effect should be estimated.

### 2.3. Ingredients of the search

In this paper, generating a factorial design means specifying the combination of levels of the factors that must be allocated to each experimental unit. The situations that we consider generalise the examples of Section 2.1, using the generic factorial decomposition of Section 2.2. Their components consist of

(a) the list of genuine factors \( F_1, \ldots, F_i \), together with their numbers of levels and any hierarchy constraints;
(b) one or more joint model and estimate specifications \( (\mathcal{M}, \mathcal{E}) \), where \( \mathcal{M} \) contains the factorial terms in the model and \( \mathcal{E} \) contains the terms to estimate \( (\mathcal{E} \subseteq \mathcal{M}) \);
(c) the size of the experiment, i.e. the number \( N \) of experimental units.

Note that we allow for two ways to handle the case when a factor \( F_j \) at \( n_j \times n_j \) levels is nested in a factor \( F_i \) at \( n_i \) levels: either \( F_j \) is declared as a \( n_j \)-level factor and its actual levels are the levels of \( F_i \cap F_j \), or \( F_j \) is declared as a \( n_i n_j \)-level factor and the hierarchy constraint \( F_j \subset F_i \) must be specified. We advise using the first option when \( F_j \) is a natural refinement of \( F_i \) and the second option when \( F_j \) is due to experimental constraints rather than innate relationships between factors. Thus, in Example 3, we used the first option for the hierarchy between blocks and subblocks, with factors \( P \) and \( Q \) in the roles of \( F_i \) and \( F_j \) respectively. In contrast, we used the second option for the hierarchy between \( A \) and the block factors (with factors \( A \) and \( P \cap Q \) in the roles of \( F_i \) and \( F_j \) respectively), because these factors have no relationship except for experimental constraints.

In regular factorial designs, the number of units \( N \) is constrained by the other specifications. For a solution to exist, \( N \) must be a multiple of \( p_1^{q_1} \cdots p_i^{q_i} \), where \( p_1, \ldots, p_i \) are the prime numbers that divide \( n \) and \( \rho_1, \ldots, \rho_i \) are lower bounds on the exponents which depend on the model and estimate specifications. The algorithm described in this paper assumes that \( N \) is given by the user (or by a higher-level algorithm). So it is up to the user to propose for \( N \) a value of the form \( N = q \cdot p_1^{r_1} \cdots p_i^{r_i} \), where \( q \) is coprime to \( p_1, \ldots, p_i \) and \( \rho_1 \leq r_1, \ldots, \rho_i \leq r_i \). In practice, the user may proceed by trial and error by testing different values of \( N \), provided the computing time is not too long. In Examples 2 and 3, the number of units is imposed by the problem, with \( N = 2^5 \cdot 3 = 12 \) and \( N = 2^2 \cdot 3^2 = 32 \), respectively. In Example 1, the complete factorial design has size \( 2^5 \cdot 3^2 = 288 \) but we look for smaller design sizes. We know that \( \rho_1 \leq 5 \) and \( \rho_2 \leq 2 \) and so we can proceed by trial and error to find them.

### 3. Elementary regular factorial designs

In Sections 3–5, the number of units and the numbers of levels of all factors are powers of the same prime \( p \). Thus \( N = p^r \) and \( n = p^s \) for some scalars \( r \) and \( s \). This occurs in Example 3 with \( p = 2 \), but not in Examples 1 and 2, which both involve primes 2 and 3. The designs are constructed using elementary abelian groups of exponent \( p \), and so they are called elementary designs when they have to be distinguished from those considered in Section 6 and later.

The cyclic group of order \( p \), denoted \( C_p \), is identified with the integers modulo \( p \) under addition. The experimental units are identified with the elements of a product group \( G \cong (C_p)^s \), and the combinations of factor levels with the elements of a product group \( T \cong (C_p)^t \). A design \( d \) is a function from \( U \) to \( T \), allocating combination \( d(u) \) to unit \( u \). From now on, elements of both \( U \) and \( T \) are regarded as column vectors and \( v^T \) denotes the transpose of a vector \( v \).

#### 3.1. Pseudofactors

The canonical projections \( V_1, \ldots, V_i \) from \( U \) onto the cyclic group \( C_p \) are called the unit pseudofactors. If \( u = (u_1, \ldots, u_i)^T \) is a unit in \( U \) then \( V_i(u) = u_i \). The canonical projections \( A_1, \ldots, A_i \) from \( T \) onto the cyclic group \( C_p \) are called the treatment pseudofactors, even though some of the genuine factors involved may be block factors. If \( t = (t_1, \ldots, t_i)^T \) is in \( T \), then \( A_i(t) = t_i \).

Both kinds of pseudofactor must be viewed as technical intermediates in the design construction. Each genuine factor considered in Section 2 is the product of one or more treatment pseudofactors. If \( F_i \) is the product of a family \( \{ A_j \}_{j \in J} \) of pseudofactors, this family is said to be a decomposition of \( F_i \) into pseudofactors. For instance, if \( F_i = A_1 \wedge A_2 \), then \( F_i \) is said to be decomposed into two pseudofactors \( A_1, A_2 \), which means that \( F_i(t) = (A_1(t), A_2(t))^T \) for every \( t \in T \). The set of pseudofactors in the decomposition of \( F_i \) is denoted by \( \mathcal{P}(F_i) \). Alternatively, we use \( \mathcal{P}(i) \) to denote the indices of the pseudofactors in \( \mathcal{P}(F_i) \).

The decomposition of factors into pseudofactors induces a decomposition of factorial terms into pseudofactorial terms. Thus, the factorial term \( \prod_{j \in J} F_j \) is decomposed into the pseudofactorial terms \( \prod_{j \in J} A_j \) such that \( \bigcap_{j \in J} \mathcal{P}(i) \neq \emptyset \) if and only if \( i \in I \). We use the notation \( \mathcal{P}(\cdot) \) to denote the set of pseudofactorial terms that decompose a given factorial term.
Example 3 (Continued). There are $32 = 2^5$ units and thus five unit pseudofactors $V_1, \ldots, V_5$ at two levels. The nine treatment pseudofactors $(A_1, \ldots, A_9)$ will rather be denoted by $P_1, P_2, Q, U_1, U_2, A, B, C$ and $D$ to keep the correspondence with the genuine factors more explicit. With this notation, we have $\mathcal{P}(P) = \{P_1, P_2\}$, $\mathcal{P}(Q) = \{Q\}$, $\mathcal{P}(U) = \{U_1, U_2\}$, $\mathcal{P}(A) = \{A\}$, and $\mathcal{P}(D) = \{D\}$.

For the decomposition of factorial terms, we give the following examples, which include two main effects and three interactions:

- $\widetilde{\mathcal{P}}(A) = \{A\}$
- $\widetilde{\mathcal{P}}(P) = \{P_1, P_2, P_1P_2\}$
- $\widetilde{\mathcal{P}}(A) = \{A, B\}$
- $\widetilde{\mathcal{P}}(P, Q) = \{P_1Q, P_2Q, P_1P_2Q\}$, and
- $\widetilde{\mathcal{P}}(P, U) = \{P_1U_1, P_1U_2, P_1U_1U_2, P_2U_1, P_2U_2, P_2U_1U_2, P_1P_2U_1, P_1P_2U_2, P_1P_2U_1U_2\}$.

3.2. Characters

Our methods are based on the theory of duals of abelian groups, which can be found in Ledermann (1977). The group homomorphisms from $T$ into $C_p$ are all the linear combinations $A = a_1A_1 + \cdots + a_9A_9$ of the treatment pseudofactors, with $a_j \in C_p$. These homomorphisms are called the characters of $T$ and they make up a group $T^*$ called the dual of $T$. Each character can be represented by its vector of coefficients $a = (a_1, \ldots, a_9)^T$ in the product group $(C_p)^9$, and the group $T^*$ can be consequently identified with this product group. The characters and dual of $U$ are defined and represented similarly. The elements of $T^*$ will be called treatment characters and the elements of $U^*$ unit characters.

Each character $A$ of $T^*$ is associated with a pseudo-factorial effect $e_r(A)$. This means a precise linear combination of treatment effects in $\mathbb{R}^p$ or $\mathbb{C}^n$, either the general mean of $r$ if $A = 0$ or a contrast if $A \neq 0$ (see Kobilinsky, 1985 or Pistone and Rogantin, 2008 for more details). The important point is that each pseudo-factorial effect belongs to a unique pseudo-factorial term in the ANOVA decomposition of the treatment effects, and this term is easy to identify by the non-zero coefficients of $A$. For example, if there is only one non-zero coefficient $a_j$, then $A = a_jA_j$ and $e_r(A)$ belongs to the main effect of pseudofactor $A_j$. If $a_j \neq 0$ and $a_k \neq 0$ then the effect $e_r(a_jA_j + a_kA_k)$ belongs to the interaction between $A_j$ and $A_k$, and so on. When $p = 2$, each pseudofactorial term has one degree of freedom, so each includes a single character and a single pseudofactorial effect. In the general case, a pseudofactorial term of order $q$ includes $(p - 1)^q$ characters and the same number of pseudofactorial effects.

Throughout this paper, additive notation is used for characters and their associated pseudofactorial effects. Multiplicative notation, in which $A_1^{a_1} \cdots A_9^{a_9}$ is used instead of $a_1A_1 + \cdots + a_9A_9$, is more common. Then $A_j^{a_j}$ belongs to the main effect of pseudofactor $A_j$, while $A_j^{a_j}A_k^{a_k}$ belongs to the interaction $A_jA_k$, and so on. This has the disadvantage that $A_1A_2$ might be interpreted as an interaction or as one of the characters whose effect is part of that interaction.

3.3. Link between factorial terms and characters

Consider now the genuine factors $F_i$, for $i = 1, \ldots, h$. Let $E_i$ be the subset of $T^*$ consisting of all characters $\sum a_jA_j$ involving pseudofactors in $\mathcal{P}(F_i)$ only. Let $\widetilde{E}_i = E_i \setminus \{0\}$, so that $\widetilde{E}_i$ consists of all the non-zero elements of $E_i$. By extension, put $E_0 = \{0\}$ and, for each main effect or interaction $\mathcal{F}(I)$, put

$$E_i = \bigoplus_{i \in I} E_i \quad \text{and} \quad \widetilde{E}_i = \bigoplus_{i \in I} \widetilde{E}_i,$$

where $E \oplus E' = \{A + B : A \in E, B \in E'\}$. The first set $E_i$ includes all the characters associated with pseudofactors coming from the decomposition of the factors $F_i$, for $i \in I$. Those among them which have at least one non-zero coefficient for each factor make up the set $\widetilde{E_i}$, which is therefore the subset of $T^*$ associated with the factorial term $\mathcal{F}(I)$.

Example 3 (Continued). For brevity, we give only one example based on the interaction $P.Q$. The corresponding set of characters $E_i$ is

$$\{0, P_1, P_2, P_1 + P_2\} \oplus \{Q, 0\} = \{0, P_1, P_2, P_1 + P_2, Q, P_1 + Q, P_2 + Q, P_1 + P_2 + Q\},$$

while $\widetilde{E}_i$ is restricted to

$$\{P_1, P_2, P_1 + P_2\} \oplus \{Q\} = \{P_1 + Q, P_2 + Q, P_1 + P_2 + Q\}.$$

3.4. Elementary regular design and its key matrix

In a regular factorial design, the treatment pseudofactors are algebraically derived from the unit ones.
4.1. Ineligible characters due to model specifications

For the experimenter, the model and the effects to estimate consist of factorial terms defined on the genuine factors, as illustrated by the examples in Section 2. As shown in Section 3.3, each such factorial term $F(I)$ is associated with a set $E_I$ of characters in $T^*$. Thus, for each pair of model-estimate sets $(\mathcal{M}, \mathcal{E})$ of factorial terms, there is an associated pair $(\mathcal{M}, \mathcal{E})$ of character sets. The model set $\mathcal{M}$ is the union of the sets $\widehat{E}_I$, for all factorial terms $F(I)$ in $\mathcal{M}$, while the estimate set $\mathcal{E}$ is the union of the sets $\widehat{E}_I$ for all terms $F(I)$ in $\mathcal{E}$. If $E$ and $E'$ are any two subsets of $T^*$, we write $E - E' = \{A - B : A \in E, B \in E'\}$. 

The basic statistical properties of regular factorial designs are given by the following proposition (see e.g. Kobilinsky and Monod, 1995). Here Ker($\varphi$) denotes the kernel of $\varphi$, which is $\{A \in T^* : \varphi(A) = 0\}$.

3.5. Confounding

For a vector $c$ in $\mathbb{R}^n$ indexed by the genuine combinations, let $c^{(d)}_a$ denote the vector in $\mathbb{R}^N$ indexed by the units and defined by $(c^{(d)}_a)_u = (c)_d(u)$. Two treatment effects $c^{(d)}_1$ and $c^{(d)}_2$ are said to be confounded with each other in design $d$ if there is a constant $\gamma$ such that $c^{(d)}_1 = \gamma c^{(d)}_2$. In this case, it is impossible to estimate the effects $c^{(d)}_1$ and $c^{(d)}_2$ separately. Treatment effects $c^{(d)}_1$ and $c^{(d)}_2$ are said to be orthogonal to each other if $c^{(d)}_1c^{(d)}_2 = 0$.

The basic statistical properties of regular factorial designs are given by the following proposition (see e.g. Kobilinsky and Monod, 1995). Here Ker($\varphi$) denotes the kernel of $\varphi$, which is $\{A \in T^* : \varphi(A) = 0\}$.

Proposition 3.1. Let $A$ and $B$ denote two characters in $T^*$. A regular design with key matrix $\Phi$ and corresponding homomorphism $\varphi$ satisfies the following four properties:

(i) The pseudofactorial effect $e_\tau(A)$ is confounded with the general mean $e_\tau(0)$ if and only if $A \in \text{Ker}(\varphi)$;
(ii) The pseudofactorial effects $e_\tau(A)$ and $e_\tau(B)$ are confounded with each other if and only if $A - B \in \text{Ker}(\varphi)$;
(iii) The sets of mutually confounded pseudofactorial effects are given by the cosets of the subgroup Ker($\varphi$);
(iv) The pseudofactorial effects $e_\tau(A)$ and $e_\tau(B)$ are orthogonal to each other if $A$ and $B$ are in different cosets of Ker($\varphi$).

If $A \in T^*$, the effect $e_\tau(A)$ is estimable if and only if it is not confounded with any other non-zero effect. Proposition 3.1 shows that this occurs if and only if all the other characters $B$ in the same coset of Ker($\varphi$) as $A$ are assumed to have no effect on the expectation of the response to be measured ($e_\tau(B) = 0$).

4. Conditions on the design key matrix

4.1. Ineligible characters and factorial terms

4.1.1. Ineligible characters due to model specifications

For the experimenter, the model and the effects to estimate consist of factorial terms defined on the genuine factors, as illustrated by the examples in Section 2. As shown in Section 3.3, each such factorial term $F(I)$ is associated with a set $E_I$ of characters in $T^*$. Thus, for each pair of model-estimate sets $(\mathcal{M}, \mathcal{E})$ of factorial terms, there is an associated pair $(\mathcal{M}, \mathcal{E})$ of character sets. The model set $\mathcal{M}$ is the union of the sets $\widehat{E}_I$, for all factorial terms $F(I)$ in $\mathcal{M}$, while the estimate set $\mathcal{E}$ is the union of the sets $\widehat{E}_I$ for all terms $F(I)$ in $\mathcal{E}$. If $E$ and $E'$ are any two subsets of $T^*$, we write $E - E' = \{A - B : A \in E, B \in E'\}$.
Table 1
Construction of the ineligible set \( I \) by symmetric differences (Example 4).

<table>
<thead>
<tr>
<th>Estimate set ( \bar{E} )</th>
<th>Model set ( \mathcal{M} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( A )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( A, B )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( A, C )</td>
</tr>
<tr>
<td>( A, B )</td>
<td>( A, B, C )</td>
</tr>
<tr>
<td>( A, C )</td>
<td>( B )</td>
</tr>
<tr>
<td>( B, C )</td>
<td>( A )</td>
</tr>
<tr>
<td>( C )</td>
<td>( \mu )</td>
</tr>
<tr>
<td>( A, B )</td>
<td>( A, C )</td>
</tr>
<tr>
<td>( B, C )</td>
<td>( \mu )</td>
</tr>
<tr>
<td>( A, B, C )</td>
<td>( A, B, C )</td>
</tr>
</tbody>
</table>

Definition 4.1 (Ineligible Characters). Let \((\mathcal{M}, \bar{E})\) be model-estimate sets of factorial terms and \((\mathcal{M}, E)\) be the associated model-estimate sets of characters. Put

\[
I = \{ A - B : A \in \bar{E}, \ B \in M, \ A \neq B \} = (\bar{E} - M) \setminus \{0\}.
\]

Then \( I \) is called the set of ineligible characters with respect to \((\mathcal{M}, \bar{E})\) or, equivalently, to \((\mathcal{M}, E)\).

Proposition 4.1. Consider a regular design with key matrix \( \Phi \) and corresponding homomorphism \( \varphi \), and the model consisting of the factorial terms in \( \mathcal{M} \), considered as fixed effects. All factorial terms in \( \bar{E} \) are estimable if and only if \( \text{Ker}(\varphi) \cap I = \emptyset \), where \( I \) is the set of ineligible characters with respect to \((\mathcal{M}, \bar{E})\).

All proofs are in the Supplementary material (see Appendix A).

4.1.2. Ineligible factorial terms

Not every set of characters of \( T \) is a union of the subsets \( \bar{E}_I \) associated with factorial terms \( \mathcal{F} (I) \). However, Proposition 4.2 shows that, for any given model-estimate pair \((\mathcal{M}, \bar{E})\), there is a set \( \mathcal{I} \) of factorial terms such that

\[
\mathcal{I} = \bigcup_{\mathcal{T} \in \mathcal{I}} \bar{E}_I.
\]

Thus it makes sense to say that \( \mathcal{T} \) is an ineligible factorial term if \( \mathcal{T} \) \( \in \mathcal{I} \). Moreover, \( \mathcal{I} \) can be calculated explicitly, as the union of the sets \( \mathcal{I}(I, J) \setminus \{\mathcal{F}(I)\} \) for \( \mathcal{F}(I) \) \( \in \bar{E} \) and \( \mathcal{F}(J) \) \( \in \mathcal{M} \), where \( \mathcal{I}(I, J) = \{ \mathcal{F}(K) : I \triangle J \subseteq K \subseteq (I \cup J) \} \), and \( I = \{ i \in I : n_i > 2 \} \), so that \( \mathcal{I}(I, J) = \{ \mathcal{F}(K) : I \triangle J \subseteq K \subseteq (I \cup J) \} \) if \( g \geq 3 \). Here \( I \triangle J \) denotes the symmetric difference between \( I \) and \( J \), which is \( (I \setminus J) \cup (J \setminus I) \). Proposition 4.3 shows that, in three important special cases,

\[
\mathcal{I} = \{ \mathcal{F}(I \triangle J) : \mathcal{F}(I) \in \bar{E}, \ \mathcal{F}(J) \in \mathcal{M}, \ I \neq J \}.
\]

Proposition 4.2. Let \( I \) and \( J \) be subsets of \{1, \ldots, h\}. Then

\[
\bar{E}_I - \bar{E}_J = \bigcup_{\mathcal{T} \in \mathcal{I}(I, J)} \bar{E}_I.
\]

Thus the first step to determine the set \( I \) of ineligible characters is to determine the set \( \mathcal{I} \) of ineligible factorial terms. If any of the conditions in Proposition 4.3 is satisfied, this is done by identifying the sets \( K = I \triangle J \) for \( \mathcal{T} \) \( \in \mathcal{I} \) in \( \bar{E} \) and \( \mathcal{T} \) \( \in \mathcal{M} \). Otherwise, the sets \( \mathcal{I}(I, J) \) of ineligible factorial terms must be used, for \( \mathcal{T} \) \( \in \bar{E} \) and \( \mathcal{T} \) \( \in \mathcal{M} \). This step can be performed before the decomposition into pseudofactors. In a second step, the ineligible characters can be deduced from Eq. (2).

Example 4. For sake of brevity, we give an example simpler than Example 3. Suppose that there are three factors \( A, B, C \), that

\[
\mathcal{M} = \{ \mu, A, B, C, A.B, B.C \} \quad \text{and} \quad \bar{E} = \{ A, B, C, A.B \}.
\]

Table 1 gives for each model term \( \mathcal{T} \) \( \in \mathcal{M} \) and each estimate term \( \mathcal{T} \) \( \in \bar{E} \) the associated ineligible factorial term \( \mathcal{T} \triangle \mathcal{T} \). Therefore \( \mathcal{I} \) includes all non-mean terms in this table, that is \( \{ A, B, C, A.B, B.C \} \).

Example 5. Suppose that there are two factors \( A, B \) and we choose \( \mathcal{M} = \{ \mu, A, B \} \) and \( \bar{E} = \{ B \} \). Because it does not contain \( A \), the model \( \mathcal{M} \) is not complete, and because it does not contain \( \mu \), \( \bar{E} \) is not complete. Proposition 4.2 implies that \( \mathcal{I} = \{ A, B \} \) if factor \( B \) has two levels, whereas \( \mathcal{I} = \{ A, B, A.B \} \) if factor \( B \) has three or more levels. Here are some possibilities.
(a) If factors \( A \) and \( B \) have two levels, then \( \mathcal{M} = \{0, B, A + B\} \) and \( \mathcal{I} = \{A, B\} \). So the only ineligible factorial terms are the main effects of \( A \) and \( B \). We could construct a design by confounding the character \( A \) with \( B \).

(b) If factor \( A \) has four levels and factor \( B \) has two levels, then \( \mathcal{M} = \{0, B, A_1 + B, A_2 + B, A_1 + A_2 + B\} \) and \( \mathcal{I} = \{B, A_1, A_2, A_1 + A_2\} \). We could construct a design confounding any one of \( A_1 + B, A_2 + B, A_1 + A_2 + B \) and the interaction \( A.B \).

(c) If factors \( A \) and \( B \) have three levels, then \( \mathcal{M} = \{0, B, 2B, A + 2B, 2A + B, 2A + 2B\} \) and \( \mathcal{I} = \{A, 2A, 2B, A + 2B, A + 2B, 2A + B, 2A + 2B\} \). So the ineligible factorial terms are the main effects of \( A \) and \( B \) and the interaction \( A.B \). If we confound either part of the \( A.B \) interaction then \( B \) is confounded with the other part and so cannot be estimated.

(d) If factor \( A \) has two levels and factor \( B \) has four levels, then \( \mathcal{M} = \{0, B, 2B, A + 2B, A + B, A + B_1 + B_2\} \) and \( \mathcal{I} = \{A, B_1, B_2, B_1 + B_2, A + B_1, A + B_2, A + B_1 + B_2\} \). Again, no non-zero character is eligible for confounding.

As this example shows, if the model and the estimate-set are both incomplete, the ineligible factorial terms may depend on the numbers of levels of the factors.

### 4.2. Hierarchy constraints

In practice, besides constraints of ineligibility, it may be necessary to satisfy hierarchy constraints between factors, such as those shown in Examples 2 and 3. It is assumed that all constraints are of the form \( F_{i1} \land \ldots \land F_{ik} \approx F_{i0} \) or can be deduced from such constraints. This assumption is satisfied in most practical cases. Recall that \( \mathcal{P}(i) \) denotes the pseudofactors \( A_j \) that decompose the genuine factor \( F_j \), or more precisely the set of indices of these pseudofactors.

**Proposition 4.4.** Let \( F_{i0}, F_{i1}, \ldots, F_{in} \) be \( k + 1 \) genuine factors. Let \( j = \mathcal{P}(i_1) \cup \cdots \cup \mathcal{P}(i_k) \), which is the set of pseudofactors that decompose \( F_{i1}, \ldots, F_{in} \). For a regular design with design key matrix \( \Phi \), the following three conditions are equivalent:

(i) \( F_{i1} \land \ldots \land F_{in} \approx F_{i0} \);

(ii) \( \bigwedge_{j \in I} A_j \approx A_{j0} \), for all \( j_0 \in \mathcal{P}(i_0) \);

(iii) if \( j_0 \in \mathcal{P}(i_0) \) then the column \( \bar{A}_{j_0} \) of \( \Phi \) is a linear combination of the columns \( \bar{A}_j \), for \( j \in J \).

Proposition 4.4 shows that each hierarchy constraint between genuine factors generates one or more hierarchy constraints \( (A_{j1} \land \ldots \land A_{jl} \approx A_{j0}) \) between pseudofactors, each of which must be satisfied during the design search. In the sequel, it will be assumed that the pseudofactors are ordered so that, for any such constraint, \( j_0 \) is greater than \( j_1, \ldots, j_l \). The set of all coarser pseudofactors \( A_{j0} \) involved in such constraints, denoted by \( \mathcal{H}_+ \), results from the decomposition of the coarser factors in the original constraints. The set of all pseudofactors in the finer part of the constraint \( (A_{j1} \land \ldots \land A_{jl} \approx A_{j0}) \) will be denoted by \( \mathcal{H}_- \), where, depending on the context, \( z \) may refer to the coarser pseudofactor \( A_{j0} \) or to its index \( j_0 \).

**Example 3** (Continued). Recall that there is a unique hierarchy constraint \( P \land Q \approx A \). It follows that \( \mathcal{H}_+ = \{A\} \) and \( \mathcal{H}_- = \{P, Q\} \).

### 4.3. Extension to multi-stratum experiments

Multi-stratum experiments provide many examples of mixed models, where some factorial terms are assumed to be random and centred rather than fixed. For example, the block factors in Examples 2 and 3 are assumed to be random while the treatment factors are assumed fixed. If the proper randomisation is applied, the block effects can indeed be considered random and centred.

When the block factors form a poset block structure on the experimental units, each stratum is defined by a suitable subset of the block factors. In Example 3 these subsets are \( \{P\} \), \( \{P, Q\} \) and \( \{P, Q, U\} \). In Example 2 they are \( \{R\} \), \( \{C\} \), \( \{R, C\} \) and \( \{R, C, Z\} \), where \( Z \) is an implicit two-level factor for the units within each cell.

Let \( J \) be the set of indices for such a stratum subset. Then the stratum can also be identified to the subset \( \bar{g}(J) \) of random block factorial terms \( \mathcal{F}(K) \), for \( K \subseteq J \) and \( K \not\subseteq J \) for all stratum subsets \( J' \subseteq J \). In Example 3 these subsets of factorial terms are \( \{P\} \), \( \{P, Q\} \) and \( \{U, P, U, Q, U\} \). In Example 2 they are \( \{R\} \), \( \{C\} \), \( \{R, C\} \) and \( \{Z, R, Z, C, Z, R.C.Z\} \). Note that the factorial terms which include the units factors \( U \) or \( Z \) correspond to residual error terms and are not considered below.

If the treatment factorial term \( \mathcal{F}(I) \) is subject to the hierarchy constraint \( \bigwedge_{j \in I} F_j \leq \bigwedge_{j \in J} F_j \), then \( \mathcal{F}(I) \) should, if possible, be estimated in the stratum defined by \( I \). To do this, we define a model-set \( \mathcal{M}_I \) and estimate-set \( \mathcal{E}_I \) associated with this stratum, as follows. The model-set \( \mathcal{M}_I \) is made by removing from \( \mathcal{M} \) all the block factorial terms in \( \bar{g}(J') \) for any stratum subset \( J' \) with \( J' \supseteq J \). The estimate-set \( \mathcal{E}_I \) contains all treatment factorial terms \( \mathcal{F}(I) \) in \( \mathcal{E} \) which are subject to the hierarchy constraint \( \bigwedge_{j \in I} F_j \leq \bigwedge_{j \in J} F_j \), but are not subject to any constraint \( \bigwedge_{j \in K} F_j \leq \bigwedge_{j \in I} F_j \) for which \( K \) is a subset of the (indices of the) block factors defining a stratum and \( K \not\subseteq J \).

For each stratum, the set \( I_{i,j} \) is derived from \( \mathcal{E}_I \) and \( \mathcal{M}_I \) by Eq. (3) if \( \mathcal{M}_I \) is complete or if \( \mathcal{E}_I \) is complete or if all factors involved in \( \mathcal{M}_I \) have two levels, and otherwise is derived from \( \mathcal{E}_I \) and \( \mathcal{M}_I \) by using Proposition 4.2. The union \( I \) of these sets \( I_{i,j} \) is the full set of ineligible terms, from which the full set \( I \) of ineligible characters must be deduced.
Example 3 (Continued). For simplicity, we use subscripts 1 and 2 instead of \{P, Q, U\} and \{P, Q\}. We have

\[
\underline{I}_1 = \{A, B, C, D, \ A.B.A.C, A.D, B.C, B.D, C.D, \\
A.B.C, A.B.D, A.C.D, B.C.D, A.B.C.D, \\
B.P.Q, C.P.Q, D.P.Q, A.B.P.Q, A.C.P.Q, A.D.P.Q, B.C.P.Q, B.D.P.Q, C.D.P.Q\}
\]

and

\[
\]

The set of model-based ineligible factorial terms is given by \(\underline{I}_{1,2} = \underline{I}_1 \cup \underline{I}_2 = \underline{I}_1 \cup \{A, A.P\}\). Note that the interactions \(A.Q\) and \(A.P.Q\) are absent from \(\underline{I}_{1,2}\), whereas they would be included among the ineligible terms if the latter were based on \(\underline{M}\) and \(\underline{E}\).

4.4. Ineligibility due to combinatorial requirements

In some situations, it is required that all the combinations of levels of some factors are in the experiment, whatever the model and estimate constraints are. Then all the terms including the corresponding pseudofactors have to be included in the ineligible set. For instance, in a row-and-column design, all combinations of levels of the row and column factors must be present. If the rows are defined by a factor \(R\), the columns by a factor \(C\), then \(I\) must include the factorial terms \(R.C\) and \(R.C\) whatever the model specifications, and \(I\) will include the corresponding characters.

Example 3 (Continued). The experimental units consist of all levels combinations of \(P, Q\) and \(U\). To ensure that the design has all these combinations, the set of ineligible terms must become \(\underline{I} = \underline{I}_{1,2} \cup \{P, Q, P.Q, U, P.U, Q.U, P.Q.U\}\).

5. Search for key matrices of elementary designs

5.1. Main steps

Section 4 showed that the first step in the search for a key matrix solution is the determination of the set \(\underline{I}\) of ineligible factorial terms. Further technical details are given in Section 5.2. The next step is to deduce from \(\underline{I}\) a reduced set \(R\) of representative ineligible treatment characters, as described in Section 5.3. The third step is the search for one or more key matrices \(\Phi\) satisfying condition (iii) of Proposition 4.4 (hierarchies) and Eq. (5) of Definition 5.1 (ineligibility). This can be done by the backtrack algorithm described in Section 5.4.

5.2. Determining the ineligible factorial terms

When Eq. (3) holds, the elements of \(\underline{I}\) can be identified by representing each main effect or interaction \(\mathcal{F}(I)\) by a vector of dimension \(h\) over \(C_h\), with ith coordinate equal to 1 if \(i \in I\), to 0 otherwise. With this representation, if \(\mathbf{x}\) represents \(\mathcal{F}(I)\) and \(\mathbf{z}\) represents \(\mathcal{F}(J)\), then the vector representing \(\mathcal{F}(I \triangle J)\) is simply \(\mathbf{x} + \mathbf{z}\), whose ath coordinate is 1 if \(x_a = z_a\) and is 0 otherwise.

Example 4 (Continued). The associated vectors \(\mathbf{x}, \mathbf{z}\) and \(\mathbf{x} + \mathbf{z}\) for \(\mathbf{x} \neq \mathbf{z}\) are

- for \(\underline{M}(\mathbf{x})\) :
  \[
  \begin{pmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  0 \\
  1 \\
  0 \\
  1 \\
  0 \\
  1 \\
  1 \\
  0 \\
  0 \\
  0 \\
  0
  \end{pmatrix}
  \]

- for \(\underline{E}(\mathbf{z})\) :
  \[
  \begin{pmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  1 \\
  0 \\
  1 \\
  0 \\
  0 \\
  1 \\
  0 \\
  0 \\
  0 \\
  1
  \end{pmatrix}
  \]

- for \(\underline{I}(\mathbf{x} + \mathbf{z})\) :
  \[
  \begin{pmatrix}
  1 \\
  1 \\
  0 \\
  0 \\
  0 \\
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
  1 \\
  0 \\
  0 \\
  1
  \end{pmatrix}
  \]

5.3. Reduced set of ineligible characters

Once the full set \(\underline{I}\) of ineligible factorial terms has been determined, the ineligible characters can be deduced as the union of the sets \(\mathcal{E}_I\), for \(I \in \underline{I}\). If \(I\) is the set of ineligible characters, the homomorphism \(\varphi\) must satisfy

\[
\varphi(A) \neq 0 \quad \text{for every character } A \in I.
\]

However, if \(A\) and \(B\) in \(I\) are such that \(A\) is an integer multiple of \(B\), that is \(A = kB\), then the inequality \(\varphi(A) \neq 0\) clearly implies \(\varphi(B) \neq 0\). In the search for \(\varphi\), the inequality (4) has therefore to be checked only for an adequately chosen subset \(\mathcal{R}\) of \(I\) called a reduced ineligible set.
**Definition 5.1** (Reduced Set). A reduced ineligible set \( \mathcal{R} \) is any subset of \( \mathcal{I} \) such that condition (4) is equivalent to the apparently weaker condition

\[
\varphi(A) \neq 0 \quad \text{for every } A \in \mathcal{R}.
\]  

We now indicate how such a reduced ineligible set can be selected. Let \( \langle A \rangle \) be the cyclic subgroup generated by \( A \). The relation

\[
\langle A \rangle = \langle B \rangle \iff \exists \delta, \delta' \in \mathbb{N} \text{ such that } A = \delta B \text{ and } B = \delta' A
\]

defines an equivalence relation on \( \mathcal{I} \). Clearly it is enough to check inequality (4) for only one representative in each equivalence class.

When \( T^* \) and \( U^* \) are elementary abelian \( p \)-groups, all non-zero characters have order \( p \), the non-zero equivalence classes contain \( p - 1 \) characters, and the representatives are often chosen as those whose first non-zero coordinate is 1. In Section 5.4, the representatives are chosen like this if there is only one non-zero coordinate, but otherwise chosen to have last non-zero coordinate −1.

**Example 3** (Continued). In this example, \( p = 2 \) so that each equivalence class has a single element: hence the set of ineligible characters cannot be reduced.

**Example 6.** If factors \( A \) and \( B \) both have three levels and \( \mathcal{I} \) contains their two-factor interaction then a possible reduced set \( \mathcal{R} \) contains the characters \( A - B (= A + 2B) \) and \( 2A - B (= 2A + 2B) \) for this term but neither \( A + B \) nor \( 2A + B \).

### 5.4. Elementary backtrack search

Searching for the key matrix \( \Phi \) is equivalent to searching for its columns \( \tilde{A}_1, \ldots, \tilde{A}_k \) among the set \( U^* \) of unit characters. For the homomorphism \( \varphi \) to satisfy (5), these characters must satisfy:

\[
\text{for every } a_1A_1 + \cdots + a_kA_k \in \mathcal{R}, \quad a_1\tilde{A}_1 + \cdots + a_k\tilde{A}_k \neq 0.
\]  

(7)

In addition, because of the hierarchy constraints, some columns must be linear combinations of other ones of smaller index, as shown in Proposition 4.4. More precisely, for the indices \( j \) in the subset \( \mathcal{H}_j \), the columns \( \tilde{A}_j \) must satisfy

\[
\tilde{A}_j = \sum_{k \in \mathcal{H}_{<j}} a_k\tilde{A}_k
\]  

(8)

for some values \( a_k \). Note that all indices \( k \) in \( \mathcal{H}_{<j} \) are strictly smaller than \( j \).

The columns \( \tilde{A}_j \) can be found successively by the backtrack search presented in Algorithm 1 below. The process may either end at the first success or continue until there is no more admissible \( \tilde{A}_1 \) to select.

Algorithm 1 involves non-trivial calculations only when determining the set \( a_k\tilde{A}_k \) in Step 1. A unit character is considered admissible for column \( j \) if it satisfies the inequalities (7) involving it and the previous columns \( \tilde{A}_1, \ldots, \tilde{A}_{j−1} \). Let \( \mathcal{R}_j \) be the subset of characters \( a_1A_1 + \cdots + a_jA_j \in \mathcal{R} \) having \( a_j \) as the last non-zero coordinate. The inequalities in (7) to consider when searching for \( \tilde{A}_j \) are all those involving a character in \( \mathcal{R}_j \). They can be written:

\[
\text{for every } a_1A_1 + \cdots + a_{j−1}A_{j−1} - A_j \in \mathcal{R}_j, \quad \tilde{A}_j \neq a_1\tilde{A}_1 + \cdots + a_{j−1}\tilde{A}_{j−1},
\]  

(9)

which includes \( \tilde{A}_j \neq 0 \) if \( A_j \in \mathcal{R} \). The admissible characters satisfying (9) are looked for among \( U^* \) unless \( j \) belongs to \( \mathcal{H}_+ \).

In that case, the hierarchy constraints (8) allow the search to be restricted to the subgroup of \( U^* \) generated by the columns in \( \mathcal{H}_{<j} \).

Optimising the backtrack algorithm is a complex task which is not the subject of this paper. In Appendix B of the Supplementary material (see Appendix A), however, we describe a few tricks implemented in planor (Kobilinsky, 2005) to make the backtrack search run faster.

### 5.5. End of the search

It is sometimes necessary to go beyond a success to find other solutions, for instance if the whole set of solutions is searched for, or if the backtrack column search is part of a backtrack process among Sylow components as will be described in Section 7.

When the number of factors involved increases, the time taken by the search may become very long, especially if there is no solution or only a small number of solutions. So it can be necessary to stop a search which takes too much time. In that case, which must clearly be distinguished from a true failure, it is useful to know the index of the last column reached as this indicates the kind of experimental design obtainable in a reasonable time.
Algorithm 1 Backtrack search

1. \( \text{jprev} \leftarrow 0 \) and \( \text{j} \leftarrow 1 \)
2. \( \text{while} \ \text{j} > 0 \ \text{do} \)
3. \( \{\text{Step 1: determine or update the admissible set}\} \)
4. \( \text{if} \ \text{jprev} < \text{j} \ \text{then} \) (forward case)
5. \( \) determine the set \( \text{aA}_j \) of currently admissible unit characters for column \( \text{j} \) of \( \Phi \)
6. \( \) else (backward case)
7. \( \) delete the current character in column \( \text{j} \) from \( \text{aA}_j \)
8. \( \) end if
9. \( \{\text{Step 2: determine and make the next move}\} \)
10. \( \text{if} \ \text{aA}_j \) is empty then
11. \( \) \( \text{j} \leftarrow \text{j} - 1 \) \{no solution for column \( \text{j} \), so move backward\}
12. \( \) else
13. \( \) set column \( \text{j} \) to the first element in \( \text{aA}_j \)
14. \( \) if \( \text{j} < s \) then
15. \( \) \( \text{j} \leftarrow \text{j} + 1 \) \{solution for column \( \text{j} \), so move forward\}
16. \( \) else (\( \text{j} = s \) so that all columns have been found)
17. \( \) save the current key matrix in the solution set
18. \( \) either stop or continue to find more solutions
19. \( \) \( \) end if
20. \( \) end if
21. \( \text{jprev} \leftarrow \text{j} \)
22. \( \) \( \) end while

Example 3 (Continued). For completeness, we give the first solution for \( \Phi \) found by the \text{planor} algorithm:

\[
\Phi = \begin{pmatrix}
\hat{r}_1 & \hat{r}_2 & \hat{q} & \hat{u}_1 & \hat{u}_2 & \hat{a} & \hat{b} & \hat{c} & \hat{d} \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1
\end{pmatrix}
\]

In its current implementation, \text{planor} finds 9216 solutions, including solutions obtained from others by permutation of \( B, C, D, \) or \( P_1 \) and \( P_2 \), or \( U_1 \) and \( U_2 \).

It took \text{planor} 0.073 s to find the first solution and 2.193 s to complete an exhaustive search, using a laptop computer (7.7 GHz Intel Core i7-5600U CPU@2.60 GHz x 4 processor; Linux Ubuntu operating system; R version 3.3.0; \text{planor} version 1.0.1). See Appendix C.3 of the Supplementary material for more details (see Appendix A).

6. Generalised regular factorial designs

This section extends the previous methods to cover designs involving more than one prime. The underlying theory is given by \text{Bailey} (1977, 1985), \text{Kobilinsky} (1985), \text{Kobilinsky and Monod} (1995) and \text{Pistone and Rogantin} (2008).

6.1. Pseudofactors

Let \( p_1, \ldots, p_l \) denote distinct prime numbers. The experimental units are now identified with the elements of a product group \( U = (C_{p_1})^{r_1} \times \cdots \times (C_{p_l})^{r_l} \), and the genuine combinations with the elements of a product group \( T = (C_{p_1})^{s_1} \times \cdots \times (C_{p_l})^{s_l} \). Note that the theory extends to prime powers but, for simplicity, we avoid this level of generality here. The number \( N \) of units and the number \( n \) of genuine combinations factorise into \( N = N_1 \times \cdots \times N_l \) and \( n = n_1 \times \cdots \times n_l \), where \( N_k = p_k^{s_k} \) and \( n_k = p_k^{r_k} \).

As in Section 3, the treatment pseudofactors associated with this decomposition of \( T \) are a finer decomposition of the genuine factors \( \hat{r}_i \). Let \( r = \sum_k r_k \) and \( s = \sum_k s_k \). In what follows, we denote the \( r \) unit pseudofactors by \( V_i \), for \( i = 1, \ldots, r \) and the \( s \) treatment pseudofactors by \( A_j \), for \( j = 1, \ldots, s \). We denote by \( \pi(V_i) \) and \( \pi(A_j) \) the numbers of levels of pseudofactors \( V_i \) and \( A_j \). Occasionally, to stress the structure induced by the different primes, we denote the unit pseudofactors at \( p_k \) levels by \( V_{(k,l)} \) and the treatment pseudofactors at \( p_k \) levels by \( A_{(k,l)} \), for \( k = 1, \ldots, l \), for \( i = 1, \ldots, r_k \) and for \( j = 1, \ldots, s_k \). In the examples, though, we use \( p_k \) rather than \( k \) as the first index between square brackets, because it improves clarity. Unless specified otherwise, the pseudofactors are ordered in the natural lexicographic order induced by their double index, which puts together all the pseudofactors associated with the same prime.
To define the pseudofactors properly, all levels are embedded into the same cyclic group \( C_M \) where \( M = \prod_{k=1}^{l} p_k \). Unit and treatment pseudofactors are considered as mappings into that cyclic group \( C_M \). That is, if \( \mathbf{u} = (u_1, \ldots, u_r) \) is a unit in \( U \) and \( \mathbf{t} = (t_1, \ldots, t_s) \) a genuine combination in \( T \), then
\[
V_i(\mathbf{u}) = \frac{M}{\pi(V_i)} u_i \quad \text{and} \quad A_j(t) = \frac{M}{\pi(A_j)} t_j. \tag{10}
\]

**Example 1 (Continued).** The two primes involved in this example are 2 and 3, so that all genuine factors are decomposed into pseudofactors at two or three levels. We have \( M = 6, s_2 = 5 \) and \( s_3 = 2 \), and the pseudofactors are \( A_{[2,1]}, A_{[2,2]}, A_{[2,3]}, A_{[2,4]}, A_{[2,5]}, A_{[3,1]} \) and \( A_{[3,2]} \), where the first index denotes the prime and the second index runs from 1 to \( s_2 \) or \( s_3 \). The association between factors and pseudofactors is given by
\[
F_1 = A_{[2,1]} \wedge A_{[3,1]}; \quad F_2 = A_{[2,2]} \wedge A_{[3,2]}; \quad F_3 = A_{[3,3]} \quad \text{and} \quad F_4 = A_{[2,4]} \wedge A_{[2,5]}.
\]
An equivalent notation keeps the lexicographic order of the pseudofactors but identifies them with a single index. In the \( \mathcal{P}(.) \) notation, the association reads
\[
\mathcal{P}(F_1) = \{A_1, A_6\}; \quad \mathcal{P}(F_2) = \{A_2, A_3\}; \quad \mathcal{P}(F_3) = \{A_7\} \quad \text{and} \quad \mathcal{P}(F_4) = \{A_4, A_5\}.
\]
Consider, for example, the combination identified with \( \mathbf{t} = (1, 0, 1, 1, 2, 1) \) \( \in (C_2)^5 \times (C_3)^2 \). Following Eq. (10), we have \( A_{[2,1]}(\mathbf{t}) = 3, A_{[2,2]}(\mathbf{t}) = 0, A_{[2,3]}(\mathbf{t}) = 3, A_{[2,4]}(\mathbf{t}) = 3, A_{[2,5]}(\mathbf{t}) = 0, A_{[3,1]}(\mathbf{t}) = 4 \) and \( A_{[3,2]}(\mathbf{t}) = 2 \).

An option involving prime powers is to decompose the factors as little as possible. Then the pseudofactors are \( A'_{[2,1]}, A'_{[2,2]}, A'_{[2,3]}, A'_{[3,1]} \) and \( A'_{[3,2]} \), at respectively 2, 4, 4, 3 and 3 levels, with \( M = 12 \), so that
\[
F_1 = A'_{[2,1]} \wedge A'_{[3,1]}; \quad F_2 = A'_{[2,2]}; \quad F_3 = A'_{[3,2]} \quad \text{and} \quad F_4 = A'_{[2,3]}.
\]
This is not implemented in *planor* but discussed briefly in Section 8.

**Example 2 (Continued).** The two primes involved are 2 and 3 once again. Using the double index notation, we have:
\[
C = A_{[2,1]}; \quad R = A_{[3,1]}; \quad D = A_{[2,2]}; \quad E = A_{[2,3]} \quad \text{and} \quad A = A_{[3,2]}.
\]
However, we will rather use \( C, D, E, R \) and \( A \) to denote the pseudofactors, which are confounded with factors in this example.

6.2. Characters

The characters from \( T \) into \( C_M \) are all the linear combinations \( A = a_1A_1 + \cdots + a_rA_r \) of the treatment pseudofactors, with \( a_i \in \mathbb{C}_{\pi(A_i)} \). They belong to the group \( T^* \) which is the dual of \( T \), so that \( T^* \cong (\mathbb{C}_p)^{s_2} \times \cdots \times (\mathbb{C}_p)^{s_r} \). The characters and dual of \( U \) are defined and represented in the same way.

As in Section 3, each character \( \chi \) of \( T^* \) is associated with a pseudofactorial effect, denoted by \( \epsilon_\chi(A) \), which belongs to a unique pseudofactorial term in the ANOVA decomposition of the genuine effects. This term is identified by the non-zero coefficients of the character \( \chi \). We use the same definitions and interpretations as before for the character subsets \( E_i, \hat{E}_i, E_i \) and \( \hat{E}_i \).

6.3. Generalised regular factorial designs and key matrices

Definition 3.1 can be generalised to the more general setting of the present section (Kobilinsky and Monod, 1995; Pistone and Rogantin, 2008). The design \( d \) should satisfy an appropriate generalisation of Eq. (1). If \( \mathbf{u} \in U \) then \( \pi(A_j)d(\mathbf{u}) = 0 \) (mod \( M \)), by (10), for \( j = 1, \ldots, s \). When \( \mathbf{u} = \mathbf{0} \) then \( A_j(d(\mathbf{u})) = \alpha_j \), and so \( M \) must divide \( \pi(A_j)\alpha_j \). Now consider the unit \( \mathbf{u} \) defined by \( u_i = 1 \) and \( u_k = 0 \) if \( k \neq i \). Then \( A_j(d(\mathbf{u})) = \phi_{ij}M/\pi(V_i) + \alpha_j \), from (10). Hence \( \pi(A_j)\phi_{ij}M/\pi(V_i) = 0 \) (mod \( M \)) and so \( \pi(V_i) \) divides \( \pi(A_j)\phi_{ij} \). If \( \pi(V_i) \neq \pi(A_j) \) then \( \pi(V_i) \) divides \( \phi_{ij} \) and so \( \phi_{ij}M/\pi(V_i) = 0 \) (mod \( M \)).

**Definition 6.1 (Regular Design).** A factorial design \( d \) with \( U \) and \( T \) as sets of units and genuine combinations respectively is called regular if there are coefficients \( \phi_{ij} \) and \( \alpha_j \) in \( C_M \) such that, for \( j = 1, \ldots, s \),
\[
A_j \circ d = \phi_{ij}V_1 + \cdots + \phi_{ij}V_t + \cdots + \phi_{ij}V_r + \alpha_j, \tag{11}
\]
where \( V_1, \ldots, V_r \) are the unit pseudofactors, \( A_1, \ldots, A_s \) are the treatment pseudofactors, and the following two conditions are satisfied:
\[
M \text{ divides } \pi(A_j)\alpha_j \quad \text{for all } j = 1, \ldots, s, \tag{12}
\]
\[
\pi(V_i) \text{ divides } \pi(A_j)\phi_{ij} \quad \text{for all } i = 1, \ldots, r \text{ and } j = 1, \ldots, s. \tag{13}
\]
Fix \( j \), and put \( p = \pi(A_j) \). If \( t = d(u) \) and (11)-(13) hold then
\[
t_j = \frac{p}{M} A_j(t) = \sum_{i=1}^{\pi} \frac{p}{M} \phi_{ij} \frac{M}{\pi(V_i)} + \frac{p}{M} \alpha_j = \sum_{i} \phi_{ij} + \beta_j \pmod{p},
\]
where \( \beta_j = \frac{pc_i}{M} \pmod{p} \) and the summation in \( \sum' \) is restricted to those \( i \) for which \( \pi(V_i) = p \), and \( \phi_{ij} \) is interpreted modulo \( p \) if \( \pi(V_i) = p \).

**Proposition 6.1.** A factorial design \( d \) is regular if and only if the combination \( t = (t_1, \ldots, t_r)^\top \) allocated to unit \( u = (u_1, \ldots, u_r)^\top \) satisfies \( t = \Phi^\top u + t_0 \), where the calculation of \( t_j \) is performed modulo \( \pi(A_j) \),
\[
\Phi = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1s} \\ \vdots & \ddots & \vdots \\ \phi_{l1} & \cdots & \phi_{ls} \end{pmatrix},
\]
\[
t_0 = (\beta_1, \ldots, \beta_r)^\top , \phi_{ij} = 0 \text{ if } \pi(V_i) \neq \pi(A_j) \text{, and, for } j = 1, \ldots, s, \beta_j \in C_{\pi(A_j)} \text{ and } \phi_{ij} \in C_{\pi(A_j)}.
\]
In particular, \( \Phi \) is block diagonal:
\[
\Phi = \text{diag}(\Phi_1, \ldots, \Phi_s),
\]
where the block \( \Phi_k \) corresponds to the prime \( p_k \).

**Definition 6.2 (Key Matrix).** The matrix \( \Phi \) is called the key matrix of \( d \).

Concerning the characters, we have the same relationships as in Section 3.4. The mapping \( \psi: U \to T \) defined by \( u \mapsto t = \Phi^\top u \) is a group homomorphism from \( U \) into \( T \). If the dual of the homomorphism \( \psi \) is denoted by \( \varphi \) and if \( \varphi(A) = B \) (with \( A \in T^* \) and \( B \in U^* \)), then we have \( b = \varphi a \), where \( a \) and \( b \) are the vectors of coefficients of \( A \) and \( B \).

6.4. Decomposition into Sylow subgroups

For \( k = 1, \ldots, l \), the elements of \( U \) of order \( p_k \) or 1 form a subgroup \( \tilde{U}_k \) isomorphic to \((C_{p_k})^s\). These subgroups are called the Sylow subgroups of \( U \). By the fundamental theorem of abelian groups, \( U \) is the direct sum \( \tilde{U_1} \oplus \cdots \oplus \tilde{U_l} \). If \( r_k = 0 \) then \( \tilde{U}_k = \{0\} \). Similarly, \( T \) has Sylow subgroups \( \tilde{T}_k \) isomorphic to \((C_{p_k})^s\) for \( k = 1, \ldots, l \), and \( T = \tilde{T_1} \oplus \cdots \oplus \tilde{T_l} \).

Likewise, the duals are direct sums of those of \( U \) and \( T \): that is, \( U^* = \tilde{U}_1^* \oplus \cdots \oplus \tilde{U}_l^* \) and \( T^* = \tilde{T}_1^* \oplus \cdots \oplus \tilde{T}_l^* \). For instance, the character \( A \) in \( T^* \) associated with the element \((\tilde{A}_1, \ldots, \tilde{A}_l)\) of \( \tilde{T}_1 \oplus \cdots \oplus \tilde{T}_l \) is the mapping \( (\tilde{a}_1, \ldots, \tilde{a}_l) \mapsto \tilde{A}_1(\tilde{a}_1) + \cdots + \tilde{A}_l(\tilde{a}_l) \), provided all the characters take their values in the common cyclic group \( C_M \). We can write \( A = (A_1, \ldots, A_l) \) or \( A = \tilde{A}_1 + \cdots + \tilde{A}_l \).

The decomposition into Sylow subgroups corresponds to the block diagonal decomposition of \( \Phi \) in Eq. (14), because \( \tilde{\Phi}_k \) is the restriction of \( \Phi \) to \( \tilde{U}_k \) and \( \tilde{T}_k \).

**Definition 6.3 (Primary Components).** The character \( \tilde{A}_k \) is called the \( p_k \)-primary component of the character \( A \).

**Definition 6.4 (Sylow Components).** The diagonal blocks \( \tilde{\Phi}_1, \ldots, \tilde{\Phi}_l \) of the matrix \( \Phi \) associated with the distinct primes \( p_1, \ldots, p_l \), are called the Sylow components of \( \Phi \).

For the homomorphism \( \varphi \), we have \( \varphi(A) = \varphi(\tilde{A}_1) + \cdots + \varphi(\tilde{A}_l) \), and Proposition 6.1 implies that
\[
\varphi(A) = \varphi(\tilde{A}_1) + \cdots + \varphi(\tilde{A}_l),
\]
where each \( \tilde{\varphi}_k \) is the homomorphism from \( \tilde{T}_k^* \) to \( \tilde{U}_k^* \) associated with the matrix \( \tilde{\Phi}_k \). The search for a matrix \( \Phi \) meeting the requirements can thus be decomposed into the search for its Sylow components. The Sylow component \( \tilde{\Phi}_k \) does not appear explicitly in \( \Phi \) if there is no treatment pseudofactor with \( p_k \) levels \((r_k \neq 0 \text{ and } s_k = 0)\) or if no unit pseudofactor has \( p_k \) levels \((r_k = 0 \text{ and } s_k \neq 0)\). In the latter case, the regular designs associated with \( \Phi \) give a constant value to genuine factors having \( p_k \) levels, and this is usually prohibited unless the design is part of a larger one.

Section 7.5.1 shows that the Sylow components can often be searched for independently. Section 7.5.3 provides a backtrack search method otherwise.

7. Search for key matrices of generalised regular designs

7.1. Main steps

The search for a key matrix follows the same main steps as those in Section 5.1. The first step is to determine the set \( \bar{I} \) of ineligible factorial terms. Let \( I \) and \( J \) be different subsets of \( \{1, \ldots, h\} \). If \( i \in I \cap J \) and \( n_i \) is not prime then \( \mathcal{P}(i) \) contains
at least two indices, and so \( \tilde{E}_I - \tilde{E}_J \) contains some characters with at least one non-zero coefficient \( a_j \) for some \( j \) in \( \mathcal{P}(i) \), and also some characters for which \( a_j = 0 \) for all \( j \) in \( \mathcal{P}(i) \). As in Proposition 4.2,

\[
\tilde{E}_I - \tilde{E}_J = \bigcup_{\mathcal{F}(K) \in \mathcal{J}(I,J)} \tilde{E}_K.
\]

Hence Proposition 4.3 remains true in this more general setting, and so \( \mathcal{J} \) can be determined as in Section 4.1.

Concerning the second step, a reduced ineligible set of characters \( \mathcal{R} \) can be deduced from \( \mathcal{J} \) by the following steps.

(a) Deduce from \( \mathcal{J} \) a set \( \mathcal{J}_p \) of ineligible pseudofactorial terms when decomposing the factors into pseudofactors. This set can be reduced as explained in Section 7.3.

(b) Deduce from \( \mathcal{J}_p \) the equivalence classes to be considered for \( \mathcal{R} \).

(c) Select one representative character in each class and eliminate representatives having a proper multiple in \( \mathcal{J} \).

In Examples 1 and 2, the models are complete and so the first main step is based on Proposition 4.3.

Example 1 (Continued). Proceeding as in Example 4 (Section 4.1.2), we find the ineligible set

\[
\mathcal{J} = \{F_1, F_2, F_3, F_4, F_1,F_2, F_1,F_3, F_1,F_4, F_2,F_3, F_2,F_4, F_3,F_4, F_1,F_2,F_3, F_1,F_3,F_4\}.
\]

Example 2 (Continued). The set \( \mathcal{M} \) is complete, so Proposition 4.3 shows that the terms in the ineligible set specified by \( \mathcal{M} \) and \( \mathcal{E} \) are those given in Table 2. Furthermore, the design must contain all the level combinations of factors \( C \) (columns) and \( R \) (rows), so the factorial terms \( C \) and \( R \) are also ineligible (see Section 4.4). Therefore

\[
\]

7.2. Main principle of the reduction

Section 5.3 explained how the set \( \mathcal{J} \) of ineligible characters can be reduced for elementary regular designs. The basic principle is that, if characters \( A \) and \( B \) are ineligible and there is a \( \delta \in \mathbb{N} \) such that \( A = \delta B \), then it is sufficient to check that \( \psi(A) \neq 0 \). Thus the character \( B \) can be omitted from the set of ineligible characters. It follows that, to be parsimonious, a reduced ineligible set \( \mathcal{R} \) must include only one representative per equivalence class defined by (6).

The same principle applies to generalised regular designs, but it is possible to go further. Indeed the relation

\[
\langle A \rangle \subseteq \langle B \rangle \iff \exists \delta \in \mathbb{N} \text{ such that } A = \delta B \quad (17)
\]

defines a partial order on equivalence classes and it is clear that (4) has to be checked only for representatives of minimal classes. The class of \( A \) is minimal if and only if \( \langle A \rangle \) contains no proper subgroup \( \langle B \rangle \) with \( B \in \mathcal{J} \). A reduced ineligible set \( \mathcal{R} \) can thus be obtained by picking one representative in each equivalence class and avoiding representatives having a proper multiple in \( \mathcal{J} \).

7.3. Reduction of the set of ineligible pseudofactorial terms

Suppose that the interaction \( F_1,F_2 \) is ineligible and that \( F_1 = A_1 \wedge A_2, F_2 = B_1 \wedge B_2 \). The character set \( \tilde{E}_{(1,2)} \) of the interaction \( F_1,F_2 \) is the union of the character sets of the nine pseudofactor interactions in \( \mathcal{P}(F_1,F_2) \), where

\[
\mathcal{P}(F_1,F_2) = \{A_1, B_1, A_2, B_1, A_1,A_2,B_1, A_1,B_2, A_2,B_2, A_1,A_2,B_2, A_1,B_1,B_2, A_2,B_1,B_2, A_1,A_2,B_1,B_2.\}.
\]

We call a term such as \( A_1 \) or \( A_1,B_1 \) a pseudofactorial term. Note that a pseudofactorial term is ineligible if and only if it is part of an ineligible factorial term.

In some cases, only a few of the pseudofactor interactions such as the nine above need to be considered when determining \( \mathcal{R} \). This property is related to the Sylow decomposition of \( T \) and was not relevant for the elementary regular designs. It motivates the construction of a reduced set \( \mathcal{J}_P \) of ineligible pseudofactorial terms, intermediate between the set \( \mathcal{J} \) of ineligible factorial terms and the reduced set \( \mathcal{R} \) of ineligible characters.

<table>
<thead>
<tr>
<th>Set ( \mathcal{E} )</th>
<th>Model ( \mathcal{M} )</th>
<th>C</th>
<th>R</th>
<th>C.R</th>
<th>A</th>
<th>D</th>
<th>E</th>
<th>D.A</th>
<th>E.A</th>
<th>D.E</th>
<th>E.A</th>
<th>D.E</th>
</tr>
</thead>
</table>

Table 2

Terms in the ineligible set (Example 2), obtained by Proposition 4.3 (redundant terms and the mean are crossed out).
In the pseudofactor notation $A_{k,l}$ introduced in Section 6.1, each pseudofactorial term can be expressed as a product $\prod_{k \in K} \prod_{j \in J} A_{k,j}$, where the index $k$ varies over a subset $K$ of $\{1, \ldots, l\}$ and, for each $k, j$, is a non-empty subset of $\{1, \ldots, s_k\}$. The associated characters in $T^*$ are the linear combinations $\sum_{k \in K} \sum_{j \in J} a(k,j)A_{k,j}$ for which every coefficient $a(k,j)$ is non-zero. The set of these characters is the sum $\bigoplus_{k \in K} E_k$ of the pseudofactorial sets $E_k$ of characters associated with the pseudofactorial terms $\prod_{j \in J} A_{k,j}$.

**Definition 7.1** (Support). The support of the character $(\hat{A}_1, \ldots, \hat{A}_l)$ is the set $\{k : 1 \leq k \leq l \text{ and } \hat{A}_k \neq 0\}$. The support of the pseudofactorial term $\prod_{k \in K} \prod_{j \in J} A_{k,j}$ is the set $K$.

**Proposition 7.1.** Let $\prod_{k \in K} \prod_{j \in J} A_{k,j}$ be an ineligible pseudofactorial term. If $L$ is any proper subset of $K$ then the pseudofactorial term $\prod_{k \in L} \prod_{j \in J} A_{k,j}$ is different from $\mu$ and different from $\prod_{k \in K} \prod_{j \in J} A_{k,j}$. If there is any such subset $L$ such that $\prod_{k \in L} \prod_{j \in J} A_{k,j}$ is ineligible, then the ineligible set of characters can be reduced by removing all characters associated with $\prod_{k \in K} \prod_{j \in J} A_{k,j}$.

**Corollary 7.1.** Consider any two pseudofactors that decompose the same factor $F_1$ and have different prime numbers of levels. Any pseudofactorial term that includes both pseudofactors can be omitted from $\mu$.

A more thorough elimination can proceed according to Algorithm 2 below, where $\text{ipt} I$ is the initial set $I_p$ deduced directly from $I$, $\text{ipt} R$ denotes the reduced set under construction, and $\text{ipt} q, \text{ipt} K$ are temporary subsets of $\text{ipt} I$.

**Algorithm 2** Reduction of ineligible pseudofactorial terms

1. $\text{ipt} I \leftarrow$ complete set of ineligible pseudofactorial terms
2. $\text{ipt} R \leftarrow \emptyset$
3. for $q = 1 \ldots l - 1$ do
   1. $\text{ipt} q \leftarrow$ subset of elements in $\text{ipt} I$ with support of size $q$
   2. $\text{ipt} R \leftarrow \text{ipt} R \cup \text{ipt} q$
   3. $\text{ipt} I \leftarrow \text{ipt} I \setminus \text{ipt} q$
4. for each pseudofactorial term $pft$ in $\text{ipt} q$ do
   1. determine the support $L$ of $pft$
   2. $\text{ipt} K \leftarrow$ subset of elements in $\text{ipt} I$ whose restriction to the support $L$ equals $pft$
   3. $\text{ipt} I \leftarrow \text{ipt} I \setminus \text{ipt} K$
5. end for
6. end for
7. return $\text{ipt} R$

**Example 1** (continued). The factorial terms in $I_p$ are expanded as functions of the pseudofactors. In this process,

- $F_1$ gives $A_{2,1}, A_{3,1}, (A_{2,1}A_{3,1})$, $A_{2,2}, A_{3,2}, A_{2,2}A_{3,2}$,
- $F_2$ gives $A_{2,1}, A_{3,1}, A_{2,2}, A_{3,2}, A_{2,1}A_{3,2}$, $A_{2,1}, A_{3,1}, A_{2,2}, A_{3,1}A_{3,2}$, $A_{2,1}, A_{3,1}, A_{2,2}, A_{3,2}$, $A_{2,1}, A_{3,1}, A_{2,2}, A_{3,2}, A_{2,1}A_{3,1}A_{3,2}$, $A_{2,1}, A_{3,1}A_{3,2}$,

and so on, where the terms between parentheses involve different primes for the same genuine factor and so may be omitted immediately, by **Corollary 7.1**.

Then the algorithm starts with the ineligible pseudofactorial terms with only one non-zero primary component (support size $q = 1$). Here only 22 of the 31 pseudofactorial terms with support $\{2\}$ are ineligible:

- $A_{2,1}, A_{2,2}, A_{2,3}, A_{2,4}, A_{2,5}, A_{2,1}A_{2,2}$, $A_{2,1}A_{2,3}$, $A_{2,1}A_{2,4}$, $A_{2,1}A_{2,5}$, $A_{2,2}A_{2,3}$, $A_{2,2}A_{2,4}$, $A_{2,2}A_{2,5}$, $A_{2,3}A_{2,4}$, $A_{2,3}A_{2,5}$, $A_{2,4}A_{2,5}$, $A_{2,1}A_{2,2}A_{2,3}$, $A_{2,1}A_{2,2}A_{2,4}$, $A_{2,1}A_{2,2}A_{2,5}$, $A_{2,1}A_{2,2}A_{2,3}A_{2,4}$, $A_{2,1}A_{2,2}A_{2,3}A_{2,5}$, $A_{2,1}A_{2,2}A_{2,4}A_{2,5}$, $A_{2,1}A_{2,2}A_{2,3}A_{2,4}A_{2,5}$, $A_{2,1}A_{2,2}A_{2,3}A_{2,4}A_{2,5}$, and $A_{2,1}A_{2,3}A_{2,4}A_{2,5}$.

However, all three of the pseudofactorial terms with support $\{3\}$, namely $A_{3,1}, A_{3,2}$ and $A_{3,1}A_{3,2}$, are ineligible. When considered as subsets of pseudofactors, all pseudofactorial terms with support $\{2,3\}$ include one or more of these pseudofactorial terms with support $\{3\}$. **Proposition 7.1** shows that they can be eliminated.

Since this first reduced set of ineligible elements includes only elements with one non-zero primary component, the same is true of any reduced ineligible set deduced from it. Section 7.5.1 shows that this very often occurs in practice and that it allows us to make the search separately for each prime. But it is not always true, as shown by **Example 2**.
Table 3
Representatives of equivalence classes in the ineligible set \( I_\varphi \) in Example 2.

<table>
<thead>
<tr>
<th>Support</th>
<th>Representative characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>{2}</td>
<td>( C, D, E, D + E )</td>
</tr>
<tr>
<td>{3}</td>
<td>( R, A )</td>
</tr>
<tr>
<td>{2, 3}</td>
<td>( C + D + R + 2A, C + D + 2R + 2A, C + E + R + 2A, C + E + 2R + 2A )</td>
</tr>
</tbody>
</table>

Example 2 (Continued). The pseudofactorial terms are confounded with the factorial ones in this example so they are given in (16). Those with support of size one may include \( C, D \) or \( E \) for \( p_1 = 2 \), or \( R \) or \( A \) for \( p_2 = 3 \), which yields \( D, E, C, D, E, R \) and \( A \). Among the other pseudofactorial terms, we can eliminate \( D.A, D.R.A, D.E.A, E.A, E.R.A \) and \( C.R \), which have \( D, E, C \) or \( D.E \) as 2-primary component, and \( C.D.A \) and \( C.E.A \), which have \( A \) as 3-primary component. The remaining terms are \( C.D.R.A \) and \( C.E.R.A \). So the reduction of pseudofactorial terms leads to the set
\[
\]

Section 3.3 shows that the subsets of characters associated with the factorial terms \( C.D.R.A \) and \( C.E.R.A \) are
\[
\{ C + D + R + A, C + D + 2R + A, C + D + R + 2A, C + D + 2R + 2A \}
\]
and
\[
\{ C + E + R + A, C + E + 2R + A, C + E + R + 2A, C + E + 2R + 2A \}
\]
respectively. Picking one representative in each equivalence class gives the ten elements in Table 3, four of which have two non-zero primary components. Each equivalence class has one element if the coefficients of the three-level factors \( R \) and \( A \) are both 0, or two otherwise. If factors \( R \) and \( A \) are both involved, the representative selected is the one whose coefficient of \( A \) is \(-1 = 2 \pmod 3\).

7.4. Reduction of the set of ineligible characters

The first step in getting the reduced ineligible set is to select one representative in each equivalence class for the relation (6). This is easy if there is some canonical way of selecting unambiguously the representative in each class. The following proposition shows that a canonical representative can be formed by picking the canonical representative of each primary component.

Proposition 7.2. If \( A = (\tilde{A}_1, \ldots, \tilde{A}_l) \) then the equivalence class \( A \) of \( A \) is the product of the equivalence classes \( A_1, \ldots, A_l \) of its primary components \( \tilde{A}_1, \ldots, \tilde{A}_l \); that is, \( A = A_1 \times \cdots \times A_l \).

When looking for representatives of minimal classes, the following proposition is useful.

Proposition 7.3. Let \( A = (\tilde{A}_1, \ldots, \tilde{A}_l) \) and \( B = (\tilde{B}_1, \ldots, \tilde{B}_l) \) be the Sylow decompositions of elements \( A \) and \( B \) of \( T^* \). Then \( (A) \subseteq (B) \) if and only if \( (\tilde{A}_k) \subseteq (\tilde{B}_k) \) for \( k = 1, \ldots, l \).

In other words, the class of \( A \) is contained in the class of \( B \) in the sense defined by (17) if and only if, for each \( k \leq l \), the class of \( \tilde{A}_k \) is contained in that of \( \tilde{B}_k \); that is, there exists an integer \( \delta_k \) such that \( \tilde{A}_k = \delta_k \tilde{B}_k \).

Given an ineligible pseudofactorial term \( \prod_{k \in K} \prod_{j \in \mathcal{J}_k} A_{k[j]} \), we seek a set of representatives of the equivalence classes in the corresponding set \( E \) of characters, by which we mean a set containing exactly one element in each equivalence class. As in Section 7.3, \( E = \bigoplus_{k \in K} E_k \), where \( E_k \) is the set of characters associated with \( \prod_{j \in \mathcal{J}_k} A_{k[j]} \).

Proposition 7.4. For each \( k \in K \), let \( E_k \) be a set of representatives of the equivalence classes in \( \tilde{E}_k \). Put \( E = \bigoplus_{k \in K} E_k \). Then \( E \) is a set of representatives of the equivalence classes in \( \bigoplus_{k \in K} \tilde{E}_k \).

Thus the search for \( \varphi \) is reduced to the search for the primary homomorphisms \( \tilde{\varphi}_k \) for each prime \( p_k \). Denote by \( I_k \) the set of ineligible characters \( A \) whose support is \( [k] \). Then \( \tilde{\varphi}_k \) must satisfy
\[
\tilde{\varphi}_k(A) \neq 0 \quad \text{for all } A \text{ in } I_k.
\]
for \( k = 1, \ldots, l \). Sometimes this necessary condition is also sufficient for condition (4) to be satisfied; sometimes it is not.

We discuss the two cases in Section 7.5.

7.5. Dependencies between Sylow components of the key matrix

7.5.1. A condition leading to independent searches for each Sylow component

For \( k = 1, \ldots, l \), denote by \( R_k \) the set of all non-zero \( p_k \)-primary components \( \tilde{A}_k \) of the characters in \( R \).
Definition 7.2. A subset of characters in $T^*$ is thin if all of its elements have support of size one.

Proposition 7.5. If $\mathcal{R}$ is thin then condition (5) on $\varphi$ is equivalent to the conjunction of the $l$ conditions

$$\tilde{\varphi}_k(A_k) \neq 0 \quad \text{for every } A_k \in \tilde{\mathcal{R}}_k,$$

for $k = 1, \ldots, l$.

Hence if $\mathcal{R}$ is thin then searching for $\varphi$ satisfying (5) is equivalent to searching separately (and independently) for the primary homomorphisms $\tilde{\varphi}_k$ satisfying (18). In practice it is easy to check directly whether $\mathcal{R}$ is thin. But a question naturally arises: is this often satisfied in practice? Proposition 7.6 gives a positive answer by giving a mild condition under which $\mathcal{R}$ is thin. On the contrary, Example 7 illustrates a practical situation in which it is not.

A subset $\delta$ of a group is said to be closed under integer multiplication if

$$A \in \delta \implies \delta A \in \delta \quad \text{for every integer } \delta.$$

It is easy to show that the subsets closed under integer multiplication are unions of subgroups.

In practice, a set like $\mathcal{M}$ defining the model is often a union of subgroups of $T^*$ and is thus closed under integer multiplication. This is always true when $\mathcal{M}$ is complete. As to the set $\mathcal{E}$ of effects to estimate, it is usual that, if it contains an interaction, it also contains all effects marginal to it except for the mean. For instance, if it contains $A, B, C$, it also contains the main effects $A, B, C$ and the two-factor interactions $A, B, A, B, C$. Under these assumptions, $\mathcal{E} \cup \{0\}$ and $\mathcal{M}$ are both closed under integer multiplication, and so is the difference $\mathcal{E} - \mathcal{M}$, which is $\mathcal{I} \cup \{0\}$.

Proposition 7.6. If $\mathcal{I} \cup \{0\}$ is closed under integer multiplication then there exists a thin reduced ineligible set $\mathcal{R}$. In particular, this holds if $\mathcal{E} \cup \{0\}$ and $\mathcal{M}$ are both closed under integer multiplication.

This result can be generalised easily to ineligible sets of characters derived from one or more sets of the form $\mathcal{I}$, described in Section 4.3: if, for every subset $\mathcal{I}$ defining a stratum, the subsets $\mathcal{E}_j \cup \{0\}$ and $\mathcal{M}_j$ are closed under integer multiplication, then the reduced set $\mathcal{R}$ is thin.

Proposition 7.6 gives as a particular case the following classical result (Bailey, 1985): the design is of resolution $R$ if all its Sylow components are.

7.5.2. Counter-examples

There are, however, situations, like in some criss-cross experiments, when $\mathcal{E}$ includes an interaction but not the main effects of the corresponding factors and when $\mathcal{R}$ is not thin. Examples 2 and 7 were constructed to illustrate this situation and its different consequences.

Example 2 (Continued). As shown in Section 7.3, the reduced set contains characters of support size 2 and so $\mathcal{R}$ is not thin. However, this example still allows for separate solutions of the Sylow components of $\Phi$, as we now explain.

The unit pseudofactors are $V_{[2,1]}, V_{[2,2]}$, and $V_{[3,1]}$. There is no loss of generality in putting $\tilde{\varphi}_2(C) = V_{[2,1]}$ and $\tilde{\varphi}_3(R) = V_{[3,1]}$. For example, one possibility for $\Phi$ is

$$V_{[2,1]} \begin{pmatrix} c & d & e & r & a \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

It is clear that $\tilde{\varphi}_2(D)$ and $\tilde{\varphi}_2(E)$ must be two of $V_{[2,1]}, V_{[2,2]}$ and $V_{[2,1]} + V_{[2,2]}$. Since the character $A$ is in $\mathcal{R}$, $\tilde{\varphi}_3(A)$ must be $V_{[3,1]}$ or $2V_{[3,1]}$, whatever $\tilde{\varphi}_3$ is. If $\tilde{\varphi}_3(A)$ is $V_{[3,1]}$ then $\varphi(C + D + R + 2A) = \varphi(C + D)$ and $\varphi(C + E + R + 2A) = \varphi(C + E)$, so $\tilde{\varphi}_2(D)$ and $\tilde{\varphi}_2(E)$ must both be different from $\tilde{\varphi}_2(C)$. If $\tilde{\varphi}_3(A)$ is $2V_{[3,1]}$ then exactly the same is true due to $\varphi(C + D + R + A)$ and $\varphi(C + E + R + A)$.

In Example 2, the solutions for $\tilde{\varphi}_2$ do not depend on the solution for $\tilde{\varphi}_3$ or vice versa. Here is a similar example with 36 units where the choice for the two-level factors depends on the choice previously made for the three-level ones.

Example 7. This is a small modification of Example 2. Now there are six units in each of the six cells, and the factor $A$ is no longer constrained to be coarser than $R$. The sets $\mathcal{M}, \mathcal{E}$ and $\mathcal{I}$ are unchanged, and $\mathcal{M}$ is complete.

The unit pseudofactors are $V_{[2,1]}, V_{[2,2]}, V_{[3,1]}$ and $V_{[3,2]}$. There is no loss of generality in putting $\tilde{\varphi}_3(C) = V_{[2,1]}$ and $\tilde{\varphi}_3(R) = V_{[3,1]}$. Then $\tilde{\varphi}_2(D)$ and $\tilde{\varphi}_2(E)$ must be two of $V_{[2,1]}, V_{[2,2]}$ and $V_{[2,1]} + V_{[2,2]}$, while $\tilde{\varphi}_3(A)$ can be any non-zero combination of $V_{[3,1]}$ and $V_{[3,2]}$. If $\tilde{\varphi}_3(A)$ is $V_{[3,1]}$ then $\varphi(C + D + R + 2A) = \varphi(C + D)$ and $\varphi(C + E + R + 2A) = \varphi(C + E)$, so $\tilde{\varphi}_2(D)$ and $\tilde{\varphi}_2(E)$ must both be different from $\tilde{\varphi}_2(C)$. If $\tilde{\varphi}_3(A)$ is not a multiple of $V_{[3,1]}$ then there is no such constraint on $\tilde{\varphi}_2(D)$ and $\tilde{\varphi}_2(E)$. 
In this case, the search cannot be made independently in the Sylow components. There are the following three fundamentally different possibilities for $\Phi$.

\[
\begin{bmatrix}
\tilde{v}_{[2,1]} & 1 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[2,2]} & 0 & 1 & 1 & 0 & 0 \\
\tilde{v}_{[3,1]} & 0 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[3,2]} & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\quad \begin{bmatrix}
\tilde{v}_{[2,1]} & 1 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[2,2]} & 0 & 0 & 0 & 0 & 1 \\
\tilde{v}_{[3,1]} & 0 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[3,2]} & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\quad \begin{bmatrix}
\tilde{v}_{[2,1]} & 1 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[2,2]} & 0 & 0 & 0 & 0 & 0 \\
\tilde{v}_{[3,1]} & 0 & 0 & 0 & 0 & 1 \\
\tilde{v}_{[3,2]} & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

More generally, counter-examples arise when, say, $A = \tilde{A}_1 + \tilde{A}_2$ belongs to $R$ but neither $\tilde{A}_1$ nor $\tilde{A}_2$ does. This may happen if an interaction has to be estimated but not the associated main effects. In that case, it may well be necessary and sufficient that either $\tilde{\psi}_1(\tilde{A}_1) \neq 0$ or $\tilde{\psi}_2(\tilde{A}_2) \neq 0$. So there may be solutions with $\tilde{\psi}_1(\tilde{A}_1) \neq 0$ and solutions with $\tilde{\psi}_1(\tilde{A}_1) = 0$. Thus the equivalence in Proposition 7.5 is not satisfied.

7.5.3. Backtrack search in the non-independent case

When $R$ is not thin, as in Example 7, the Sylow components of $\varphi$ and $\Phi$ can be found by backtrack search. Assume that $\tilde{\psi}_j$ has already been defined for $j = 1, \ldots, k - 1$. Let $R_{[k]}$ be the subset of elements $A = (\tilde{A}_1, \ldots, \tilde{A}_k, 0, \ldots, 0)$ in $R$ having $\tilde{A}_k$ as last non-zero primary component. If $A \in R_{[k]}$ then $\varphi(A) = (\tilde{\psi}_1(\tilde{A}_1), \ldots, \tilde{\psi}_k(\tilde{A}_k), 0, \ldots, 0)$ and the choice of $\tilde{\psi}_k$ must ensure that $\varphi(A) \neq 0$. If there is any index $j$ between 1 and $k - 1$ such that $\tilde{\psi}_j(\tilde{A}_j) \neq 0$, then $\varphi(A) \neq 0$ whatever the choice of $\tilde{\psi}_k$. Such elements therefore need not be considered in the search for $\tilde{\psi}_k$, which can proceed with reduced ineligible set $R_{[k]}$ consisting of those elements $A$ in $R_{[k]}$ for which $\tilde{\psi}_j(\tilde{A}_j) = 0$ for $j = 1, \ldots, k - 1$. If it succeeds and $k < l$, it goes on to find $\tilde{\psi}_{k+1}$. If it fails and $1 < k$, it goes back and tries to find another choice for $\tilde{\psi}_{k-1}$. The search finally fails if it goes back to $k = 1$ and fails to find another $\tilde{\psi}_1$. It finally succeeds if it reaches $k = l$ and finds an admissible $\tilde{\psi}_l$.

In any case, the elementary step in the search for $\varphi$ is the search for the primary homomorphisms $\tilde{\psi}_k$ for each prime $p_k$.

8. Discussion

Quite apart from the computational aspects, this paper shows great unity between different types of factorial design: fractional or not; one prime or many; blocked, split-plot, row-column, criss-cross, and so on. The approach using one or more model-estimate pairs $(M, C)$ gives a unified framework. The set of ineligible factorial terms is at the centre of this framework, since it synthesises all the constraints associated with the users’ specifications. The other central component is the design key, which determines the combinatorial and statistical properties of the design. Indeed the design problem essentially consists of finding a design key adapted to the set of ineligible factorial terms.

A few remarks must be made from a statistical point of view. Of course, once an initial design has been generated, it then needs to be randomised. Since the two steps are quite independent, we only focused on the first one in this paper. Another point is that we made no distinction between the key matrices, provided they are solutions to the design specifications of Section 2. To cope with finer criteria such as minimum aberration or maximum estimation capacity (see e.g. Mukerjee and Wu, 2006), the approach developed here gives the possibilities (up to computational constraints) to get all solutions and then select the best ones either according to such a criterion or by looking in detail at properties of the designs more relevant to the application at hand. An efficient alternative for a user of $R$ is to use the FrF2 R package (Grömping, 2014), which makes better use of such considerations but is restricted to factors at two levels.

The framework could be even more general. For example, if a factor has four levels, it is possible to associate it with the cyclic group $C_4$ rather than using two pseudofactors with two levels each; similarly for other primes and other powers. Several authors have extended the theory to this more general setting, and it was implemented in the initial version of planar. However, the work of Voss (1988, 1993) suggests that there is no practical benefit from the more general framework.

The algorithmic approach presented here to generate designs is based on backtracking, which aims at a complete exploration of the possible solutions. The drawback is that the computational burden becomes too hard when the number of factors or the degree of fractionating becomes too high. So there is clearly a need to improve the speed of the algorithm. There are many directions to do so, but we want to stress two of them.

Cheng and Tsai (2013) have shown how templates may be used for the design key in certain situations. Such a template enables us to fix one or more columns in the matrix $\Phi$. For instance, in the matrix given for Example 3 in Appendix B of the Supplementary material (see Appendix A) we lose nothing by making the columns for $A$, $B$ and $C$ the same as those for $Q$, $U_1$ and $U_2$ respectively.

The search could also be accelerated by making use of symmetries between factors or pseudofactors, with respect to the design specifications. To do so efficiently, it might be better to implement the search in a language like GAP (2016), which is expressly designed to cut down searches in this way.

The R package planar is available on the CRAN (Monod et al., 2012). It deals with the whole class of generalised regular factorial designs presented here. In addition to generating such designs, it can randomise them appropriately if the block factors and their hierarchy relationships define an orthogonal block structure. Application to the main three examples of
the paper and to one higher dimensional one is presented in Appendix C of the Supplementary material (see Appendix A). A more detailed presentation will be the subject of another paper.

Acknowledgements

This paper is dedicated to the first author, André Kobilinsky, who died before the present version was completed. André conducted most of this research and turned it into the first version of the planor software as early as 1995. The R package planor has been developed much more recently with the helpful support of Annie Bouvier (INRA, UR MalAGE). The second and third authors are grateful to the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK. A decisive step in the writing was their joint work there during the 2011 INI programme on Design and Analysis of Experiments. Support was also provided by the French Research Agency (ANR), project Escapade (ANR-12-AGRO-0003).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2016.09.003.

References

Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs

Ehsan Masoudi a,∗, Heinz Holling a, Weng Kee Wong b

a Department of Psychology, University of Münster, Fliednerstr. 21, 48149, Germany
b Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA 90095-1772, USA

ARTICLE INFO

Article history:
Received 28 February 2016
Received in revised form 24 June 2016
Accepted 26 June 2016
Available online 1 July 2016

Keywords:
Approximate design
D-optimality
Evolutionary algorithm
Fisher information matrix
General equivalence theorem
Local search

ABSTRACT

Finding optimal designs for nonlinear models is complicated because the design criterion depends on the model parameters. If a plausible region for these parameters is available, a minimax optimal design may be used to remove this dependency by minimizing the maximum inefficiency that may arise due to misspecification in the parameters. Minimax optimal designs are often analytically intractable and are notoriously difficult to find, even numerically. A population-based evolutionary algorithm called imperialist competitive algorithm (ICA) is applied to find minimax or nearly minimax D-optimal designs for nonlinear models. The usefulness of the algorithm is also demonstrated by showing it can hybridize with a local search to find optimal designs under a more complicated criterion, such as standardized maximin optimality.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

A wide class of evolutionary algorithms has been increasingly used to solve hard optimization problems in engineering, bioinformatics, computer science and finance. Of particular interest is the class of nature-inspired algorithms motivated from the influence of biology and the life sciences. Examples of such algorithms frequently used in the statistics literature are simulated annealing (SA) and genetic algorithms (GA). There are more recent and potentially more powerful ones such as differential evolution (DE), particle swarm optimization (PSO), imperialist competitive algorithm (ICA) and cuckoo search (CS) that have not yet been well tested for solving statistical problems. Our interest in this paper is application of one such algorithm to construct optimal experimental designs. The usefulness of implementing such designs in terms of cost saving and accurate statistical inferences is detailed in Atkinson (1996). An introduction to this subfield of optimal experimental designs is available in Berger and Wong (2009) and real applications of optimal designs can be found in Berger and Wong (2005).

A common appeal of such algorithms is that they are mainly assumptions free, fast, easy to implement and are broadly applicable to different types of constrained or unconstrained optimization problems. Consequently, they have good potential to optimize complicated functions with many variables regardless whether the objective function is differentiable or not. A common feature among these algorithms is that there require tuning parameters and if they are well chosen, the algorithm finds the optimum very fast. If the tuning parameters are poorly chosen, the algorithm does not give satisfactory answers. These algorithms do not usually have a firm theoretical basis, such as proof of its convergence to the optimum. However, these algorithms have been used successfully in many applied fields to solve real, complicated and high dimensional
optimization problems where traditional formulations or methods fail. Our view is that the lack of proof of convergence should not hinder their use in statistics; for some problems such as the ones we work with, there is a theory for verifying whether the generated design is optimum, and if it is not, theory is available to assess its proximity to the optimum without knowing the optimum. We next briefly review selected applications of such algorithms for finding optimal designs in the literature.

Simulated annealing was first proposed by Kirkpatrick et al. (1983) and Meyer and Nachtsheim (1988) appeared to be among the first to use SA and constructed exact D-optimal designs for both finite and continuous design spaces. Haines (1987) applied SA to construct exact D-, I-, and G-optimal designs for polynomial regression models. Atkinson (1992) discussed starting values for SA with a focus on optimal design construction and recommended segmenting the search to a maximum number of evaluations. Other applications of SA include Zhou (2008), who found exact minimax D-optimal design on discrete design spaces, and Wilmut and Zhou (2011), who constructed D-optimal minimax two-level fractional factorial designs using a sequential algorithm. Woods (2010) implemented SA to obtain exact optimal designs for binary models under the optimum-in-average criterion. This algorithm is one of the very few that can be shown to converge to the optimum.

GA was proposed by Holland and John (2000) and has been applied to search for exact optimal designs mainly for linear models. An early proponent is Montepiedra et al. (1998), who found exact optimal designs for polynomial models; others include Heredia-Langner et al. (2003), Drain et al. (2004) and Mandal et al. (2015). Hamada et al. (2001) used GA to find near-optimal Bayesian experimental designs for linear and nonlinear regression and dichotomous data. A most recent review on the application of GA to solve optimal design problems is given in Lin et al. (2015).

Particle swarm optimization (PSO) algorithm proposed by Eberhart and Kennedy (1995) has emerged to be a popular tool for solving real world optimization problems. In PSO terminology, each possible solution is called a “particle”. Similar to many evolutionary algorithms, PSO is initialized with a population of random particles, called a “swarm”. Each particle starts to fly through the problem space with its own “velocity” that is being updated in every iteration according to the particle’s best position and the best global position over the swarm. PSO has been used to find several types of optimal designs for different problems. For example, Qiu et al. (2014) applied PSO to find locally D- and c-optimal designs for the compartmental, logistic and double exponential models and comparing PSO performance with the differential evolution algorithm proposed by Storn and Price (1997). Wong et al. (2015) demonstrated the usefulness of using PSO to find various types of optimal designs by applying it to solve several types of optimal design problems for different mixture models defined over a regular or irregular simplex. In addition, Chen et al. (2014) modified PSO to find minimax optimal designs for the logistic and enzyme kinetic models. Such optimal designs are notoriously difficult to find because the design criterion is non-differentiable and involves two layers of optimization.

The main goal of this paper is to investigate the capability of the Imperialist Competitive Algorithm (ICA) for finding optimal designs. Our work appears to be the first to use ICA for a statistical application. ICA is a meta-heuristic evolutionary algorithm inspired from the socio-political process of humans and proposed in Atashpaz-Gargari and Lucas (2007). In this sense, it is different from the above mentioned nature-inspired algorithms, which are inspired by animal behavior. ICA has been successfully applied in engineering subfields such as industrial, civil, mechanical, electronic, petroleum and computer engineering; see Hosseini and Al Khaled (2014) for a review. Our interest in ICA is in part due to recent reports from the engineering literature that suggest ICA can outperform some widely used evolutionary nature-inspired algorithms including PSO. For example, Hamel et al. (2012) compared performances of ICA with PSO to optimize some famous test functions like the De Jong’s, Rastrigin’s and Hartmann’s functions in nondestructive Eddy-Current Testing (ECT) problems. They reported that when the objective function has five or fewer parameters, ICA and PSO techniques performed almost the same. However, when the number of parameters was increased, ICA found the solution faster than PSO and with more accuracy (see Table 3 of Hamel et al., 2012 for details).

In what is to follow, we focus on finding minimax type of optimal designs using ICA. These are hard design problem because they involve solving nested multi-level optimization problems, the optimality criterion is not differentiable and there is no algorithm that we know of that can be shown to converge to such optimal designs in a general nonlinear regression setup. We show how to modify ICA with a perturbed move, coupled with a local search procedure to find the optimal designs more effectively.

In the next section, we review the statistical setup and theory for finding optimal designs. We present details and implementation information for the ICA in Section 3. In Section 4 we demonstrate how ICA may be applied to find minimax optimal designs for the power logistic model. Section 5 modifies ICA to find standardized maximin D-optimal designs for the log-linear, exponential and enzyme kinetic models. Section 6 provides a discussion.

2. Background and minimax optimal designs

Throughout we focus on approximate designs proposed by Kiefer in late 1950s. His subsequent work and numerous applications of approximate designs ideas are now voluminously documented in Kiefer (1985). An approximate design $\xi$ is a probability measure defined on a user-selected design space $X$. Let $S$ be the space of all such designs on $X$ and let $\xi$ be an approximate design with $k$ support points at $x_1, x_2, \ldots, x_k$ from $X$ with corresponding weights $w_1, \ldots, w_k$, $w_i > 0$, $\sum_{i=1}^{k} w_i = 1$. This means that when we have a pre-determined total number of observations for the study, say $N$, we take approximately i.e. $Nw_1$, number of observations at $x_1$, subject to $Nw_1 + \cdots + NW_k = N$. 

In the next section, we review the statistical setup and theory for finding optimal designs. We present details and implementation information for the ICA in Section 3. In Section 4 we demonstrate how ICA may be applied to find minimax optimal designs for the power logistic model. Section 5 modifies ICA to find standardized maximin D-optimal designs for the log-linear, exponential and enzyme kinetic models. Section 6 provides a discussion.
Given a statistical model and a design criterion $\psi$, an optimal design optimizes the criterion by finding the best $k, \mathbf{x}$‘s and corresponding $w_i$’s among all designs on $\Xi$. In practice, $\psi$ is formulated as a function of the Fisher information matrix (FIM) because under some regularity conditions, the variance–covariance matrix of MLE is asymptotically equal to the inverse of FIM (Lehmann and Casella, 1998). A design that optimizes the criterion over $\Xi$ is called $\psi$-optimal. Some examples are $D$- and $A$-optimal designs that minimize over $\Xi$, the negative of the logarithm of the determinant of FIM and the trace of the inverse of FIM, respectively. A further example is $E$-optimal design that minimizes the maximum eigenvalue of the inverse of FIM. For further discussion on these criteria see, for example, Silvey (1980) or Atkinson et al. (2007).

Let $Y$ be a univariate response variable and $x$ be the set of covariates in the model. Let $E(Y) = f(\mathbf{x}, \theta)$ and $f$ is a known function apart from the model parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$. Assuming errors are independent with means zero, a direct calculation shows FIM of a $k$-point design $\xi$ is

$$M(\xi, \theta) = \sum_{i=1}^{k} w_i l(\mathbf{x}_i, \theta),$$

where

$$l(\mathbf{x}_i, \theta) = \frac{1}{\text{var}(Y_i)} \nabla f(\mathbf{x}_i, \theta) \nabla f(\mathbf{x}_i, \theta)^T$$

and

$$\nabla f(\mathbf{x}_i, \theta)^T = \left( \frac{\partial f(\mathbf{x}_i, \theta)}{\partial \theta_1}, \frac{\partial f(\mathbf{x}_i, \theta)}{\partial \theta_2}, \ldots, \frac{\partial f(\mathbf{x}_i, \theta)}{\partial \theta_p} \right).$$

Obviously, FIM depends on $\theta$ for nonlinear models and the information matrix is singular if $k < p$. For this reason, our work here assumes $k \geq p$. One simple approach to solve this dependency is to replace the unknown parameters with initial estimates from a similar study or a pilot study (Abdelbasit and Plackett, 1983). The optimal design obtained by this method is called a locally optimal design (Chernoff, 1953). Usually, locally optimal designs are not robust with respect to misspecification of the parameters, see for example, Dette and Wong (1999), and Dette et al. (2006), Dette et al. (2010).

Alternatively, one may elicit a prior distribution for the unknown parameters and find a Bayesian optimal design. Chaloner and Larntz (1989) computed Bayesian $D$-, $A$- and $c$-optimal designs for the two-parameter logistic model by assuming a uniform distribution for the parameters. This approach is a good option if there is reliable prior information and effective software for averaging out the criterion over the prior distribution. Their findings were that more support points are generally needed when we have less informative prior distributions.

Sitter (1992) proposed a minimax approach which assigns a region of uncertainty $\Theta \subset \mathbb{R}^p$ to the unknown parameters. A minimax optimal design is obtained by minimizing the worst case of the criterion on $\Theta$ over $\Xi$. It follows that minimax optimal designs are “conservative” in that they protect against the worst scenario. Let $\psi$ be a convex functional and let $\xi^*$ be defined by

$$\xi^* = \arg \min_{\xi \in \Xi} \max_{\theta \in \Theta} \psi(\xi, \theta).$$

The design $\xi^*$ is called a minimax $\psi$-optimal design. If $\psi(\xi, \theta) = -\log |M(\xi, \theta)|$, then $\xi^*$ is a $D$-minimax optimal design. Throughout, the terms “inner problem” and “outer problem” are used for optimization over $\Theta$ and $\Xi$, respectively. Finding minimax optimal designs is generally analytically intractable and even with numerical methods, the optimization can still be difficult problem. This is because we have a bi-level optimization problem; an inner problem to maximize over a continuous set $\Theta$ and an outer problem to minimize the maximum value of the inner problem over all approximate designs $\Xi$. In addition, the minimax criterion is not differentiable suggesting that optimization method employed must not rely on derivatives.

There are at least three numerical methods for finding a minimax type of optimal design. Early attempts to find such optimal designs for a regression model were generally ad hoc. For example, Berger et al. (2000) and King and Wong (2000, 2004) applied a Fedorov–Wynn type exchange algorithm as part of their algorithm to search for minimax optimal designs for the logistic and power logistic models. The algorithm begins with a randomly generated design with a non-singular information matrix and proceeds iteratively by adding a point to the current design to create a new design. The procedure is repeated until this sequence meets the conditions of a user-specified stopping rule. A problem with such an approach is that as the iteration proceeds, the generated designs have increasingly many support points clustered around the true support points of the optimal design (Mandal et al., 2015). Guess work is required to periodically collapse the accumulated points into a smaller number of points and restart the algorithm. Sometimes, this results in a very slow convergence, especially when the $\Theta$ space is large.

Semi-infinite programming (SIP) based algorithms were recently used to find minimax optimal designs for nonlinear models (Duarte and Wong, 2013). This algorithm requires (i) that the region of uncertainty be discretized and (ii) a powerful algebraic modeling language to find globally optimal solutions for nonlinear programs (NLPs). SIP-based algorithms typically uses the solver QNLP from the General Algebraic Modeling System language (GAMS; Brooke et al., 1996) to find the global optimum of NLPs. Most graduate training programs in statistics exclude SIP based algorithms in the curricula and so such methods are less used in solving optimization problems in the field of statistics.

Most recently, Chen et al. (2014) investigated the ability of PSO to solve minimax optimal design problems. In their nested-PSO algorithm, each particle is a candidate for the optimal design and so each has support points and corresponding weights. The swarm is a population of these particles randomly generated at the beginning. After the initialization step,
each particle flies through the design space with its own velocity to optimize the given criterion. The criterion value at an iteration is called the fitness of the particle and it varies as the particle changes its position. The velocity, location and fitness value of each particle is updated every iteration according to its current best position and the global best position found by the swarm. This third way of finding minimax optimal designs seems to be most flexible to date as the method is assumptions free, including not having to depend on whether the criterion is differentiable or convex over the space of information matrices. Like many other metaheuristic algorithms, there is no guarantee that PSO finds the optimum, but there is a good volume of empirical evidence in the literature that PSO does, and if they do not, they tend to get close to the optimum quickly, and frequently finds the optimum after a few re-runs. What distinguishes among these algorithms are how fast they get to the proximity of the optimum, their success rates in finding the optimum after a few re-runs, their ease of use and the types of problems they are particularly good at solving them efficiently. They each have their own features; for example, some get close to the optimum quickly but linger on thereafter (exploitation) and others are good to find the optimum when they are close to the optimum only (exploitation). A specially attractive feature of PSO frequently cited in the literature is that the default values for the tuning parameters tend to work very well.

Our setup allows us to use theory to confirm the optimality of the generated design. Our criterion is convex in the space of all information matrices and we can use directional derivative arguments to arrive at an equivalence theorem to check whether a design is optimal among all designs on the given design space $\chi$. Each convex design criterion gives rise to a different equivalence theorem but they generally have the same form. For the problem at hand, suppose that the mean nonlinear regression function has a $p \times 1$ vector of model parameters $\theta$ belonging to an uncertainty region $\Theta$. A design $\xi^*$ is minimax $D$-optimal among all designs on $\chi$ if and only if there exists a probability measure $\mu^*$ on $\Theta$ such that the following inequality holds for all $x \in \chi$,

$$c(x, \mu^*, \xi^*) = \int_{A(\xi^*)} \text{tr} M^{-1}(\xi^*, v) J(x, v) \mu^* d(v) - p \leq 0,$$

with equality in (4) at all support points of $\xi^*$. The set $A(\xi^*)$ is sometimes called the “answering set” of $\xi^*$ and the measure $\mu^*$ is a subgradient of the non-differentiable criterion evaluated at $M(\xi^*, v)$. Of course if the criterion is differentiable, $\mu^*$ becomes degenerate which explains why $D$- or $c$-optimal designs under a differentiable criterion are easier to find than minimax types of optimal designs.

The properties of the subgradient and how to find the subgradient efficiently remain evasive to date and present a key problem in solving this type of minimax optimal design problems. If the generated design is optimal, failure to find a subgradient that satisfies the equivalence theorem will mean we continue the search. In particular, there is no theoretical rule on how to choose the number of points in $A(\xi^*)$ as support for the measure $\mu^*$ and they would have to be found by trial and error. To this end, we first find all the local maxima of Problem (3) by a local search with different starting points and then pick the ones nearest to the global minimum subject to a tolerance of 0.005. Further details on numerical methods for solving such problems (3) and (4) are available in King and Wong (2004). We note that if $\chi$ is one or two dimensional, one may plot $c(x, \xi^*, \mu^*)$ versus $x \in \chi$ and visually inspect whether the graph meets the conditions in the equivalence theorem. If it does, the design $\xi^*$ is minimax optimal; otherwise it is not optimal. All of the ICA generated designs in this paper have been validated by the equivalence theorem. Due to space consideration, only a selected number of such plots, called derivative plots for short, are shown in this paper.

When the design is not optimal, the plot also reveals its proximity to the optimum without knowing the latter. We measure the closeness of a design $\xi$ to the minimax optimal design using its minimax $D$-efficiency defined by

$$d_{\text{eff}} = \left( \frac{\max_{\theta \in \Theta} - \log |M(\xi^*, \theta)|}{\max_{\theta \in \Theta} - \log |M(\xi, \theta)|} \right)^{1/p},$$

where $\xi^*$ is the minimax $D$-optimal design. Using argument similar to Atwood (1969), we obtain a lower bound for the minimax $D$-efficiency of a design $\xi$ without knowing $\xi^*$. This lower bound is $p/(p + \max_{x \in \chi} c(x, \mu^*, \xi))$, where $\mu^*$ is the probability measure defined on $A(\xi)$ that maximizes $c(x, \mu, \xi)$ over all probability measures $\mu$ and $x$, is any arbitrary support point of $\xi$. This lower bound can serve as a stopping criterion for the algorithm. If one sets this lower bound to be 0.95, for example, then the algorithm will stop when the generated design has 95% minimax $D$-efficiency or higher.

### 3. The Imperialist Competitive Algorithm (ICA)

This section describes the basic concepts behind ICA and we show they may be used to search for optimal designs. ICA begins with a random population of solutions called countries. Countries within ICA are analogous to particles in PSO. The population of countries is divided into some sub-populations called empires. Each empire contains one imperialist and some colonies. These empires start to evolve simultaneously by moving all colonies toward the imperialists through an operation...
called assimilation. In the ensuing competition step, evolution among these empires is achieved by the possession of the weakest colonies by the most powerful imperialists. Imperialist competitions direct the search process toward the best imperialist or the optimum (Kaveh and Talatahari, 2010a).

To fix ideas, suppose we have a given nonlinear mean regression function (apart from unknown values of the model parameters) defined on a known design space $\chi$. Our problem is to find a minimax $D$-optimal design for estimating the model parameters using ICA. Let $\text{pos}_i$ denote the $i$th solution (country) defined by the concatenation of $\text{pos}_i(x)$ and $\text{pos}_i(w)$, where $\text{pos}_i(x)$ and $\text{pos}_i(w)$ denote $x = (x_1, \ldots, x_n)$ and $w = (w_1, \ldots, w_k)$, respectively. If $n$ is the number of model explanatory variables, the $\text{pos}_i(x)$ and $\text{pos}_i(w)$ are of length $k \times n$ and $k$, respectively. The design optimization problem has a total of $v = k \times (n + 1)$ decision variables. To satisfy the equality constraints $\sum_{i=1}^{k} w_i = 1$, we could use a penalty function or only search within the feasible region. We tried both and noted there was not much difference between the two; accordingly, we selected the latter strategy by normalizing $\text{pos}_i(w)$ to only search in the space of all possible designs. The choice $k$, the number of support points, is user-selected and should be chosen so that the number is least equal to $p$, the number of parameters in the model; otherwise the information matrix of the design has a singular information matrices. In practice, one executes the ICA with $k = p$ and increase the value of $k$ by one until the ICA-generated design meets (or nearly meets) the conditions in the equivalence theorem and optimality is confirmed (or nearly confirmed).

For a minimax optimization problem, the inner optimization problem is the first to tackle and requires finding the values of $\theta$ to maximize a given function $\psi(\xi, \theta)$ for a given design $\xi$. Obviously the solution depends on $x$ and $w$ through $\xi$. We used the function $\text{directL}$ from package nloptr to solve the inner optimization problem. This Package is an R (R Core Team, 2014) interface to NLopt (Johnson, 2014), which is a free/open-source library for nonlinear optimization. $\text{directL}$ is a deterministic search algorithm based on systematic division of the search domain into smaller and smaller hyperrectangles (Gablonsky and Kelley, 2001). ICA is not dependent on $\text{directL}$ and any other optimization function that can assure finding the global optimum of the inner problem in a reasonable time can be used. The discussion about our reason to choose $\text{directL}$ is postponed to Section 6.

We now describe how ICA works stepwise in following subsections.

3.1. Initialization

The user first selects the number of countries $N_{\text{count}}$, number of imperialists $N_{\text{imp}}$ and starts ICA with an initial population of $N_{\text{count}}$ randomly generated countries. Each country is analogous to a particle in PSO or a chromosome in GA and represents a candidate for the optimal design with support points and weights for a non-singular design. As part of the inner optimization, ICA computes a cost value of each country. Then $N_{\text{imp}}$ “imperialists” are selected from countries with lower costs and the remaining countries form the “colonies”. The number of colonies is equal to $N_{\text{col}} = N_{\text{count}} - N_{\text{imp}}$. Next, initial empires are formed by randomly dividing the colonies among the imperialists such that the number of colonies of an empire is proportional to its “power”. If $c_i$ is the cost of the imperialist $i$, the power of the imperialist $i$ is defined by

$$\begin{align*}
P_i &= \begin{cases} 
1.3 \left( \max_{1 \leq j \leq N_{\text{imp}}} \{c_j\} \right) - c_i & \text{if } \max_{1 \leq j \leq N_{\text{imp}}} \{c_j\} > 0 \\
0.7 \left( \max_{1 \leq j \leq N_{\text{imp}}} \{c_j\} \right) - c_i & \text{if } \max_{1 \leq j \leq N_{\text{imp}}} \{c_j\} \leq 0
\end{cases} 
\end{align*}
$$

and the number of initial colonies of the empire $i$ is given by

$$\begin{align*}
N_{\text{col}} = \left\lfloor \frac{P_i}{N_{\text{imp}}} \right\rfloor \sum_{j=1}^{N_{\text{col}}} \frac{P_j}{N_{\text{imp}}} \right\rfloor
\end{align*}
$$

In a minimax optimal design problem $c_j$ is equal to $\max_{\theta \in \Theta} \psi(\xi_j, \theta)$, while in a locally optimal design problem $c_j = \psi(\xi_j, \theta_0)$, where $\xi_j$ is the design corresponding to the position of the $j$th imperialist and $\theta_0$ is the initial value specified by the practitioner when finding locally optimal design. Clearly this scheme allocates more colonies to the more powerful empires. After creating the initial empires, ICA enters the evolution stages, which comprise “assimilation”, “revolution”, “imperialist update”, “uniting similar empires” and “imperialists competition”. The algorithm terminates when user-stopping rules are met.

3.2. Assimilation within an empire

Assimilation of an empire is achieved by moving all the colonies of the empire toward the imperialist. Let $\text{pos}_i^{\text{col}}$ and $\text{pos}_i^{\text{imp}}$ denote one colony and its corresponding imperialist position, respectively. The new position of the colony at the $(t + 1)$th iteration is

$$\text{pos}_{i}^{\text{col}} = \text{pos}_{i}^{\text{col}} + \beta \times \delta \otimes (\text{pos}_{i}^{\text{imp}} - \text{pos}_{i}^{\text{col}}),$$

where $\beta$ is a coefficient, $\delta$ is a random vector, and $\otimes$ represents element-wise multiplication. The coefficient $\beta$ is a step size parameter that controls the distance the colonies move in each iteration. The random vector $\delta$ is typically drawn from a uniform or normal distribution to ensure that the colonies explore the search space. The element-wise multiplication $\otimes$ allows each component of the difference vector to be scaled independently, which can help in balancing exploration and exploitation.
where $\delta$ is a vector of length $\nu$ whose elements are random values from uniform distribution between 0 and 1, and $\otimes$ denotes element-by-element multiplication between two vectors of length $\nu$. Here $\beta$ is called “assimilation coefficient”. A value larger than 1 for $\beta$ causes the colonies to get closer to the imperialist from both sides. To increase the chance of escaping from local optima, Lin et al. (2013) proposed a perturbed assimilation that allows countries moving away from $\text{pos}_t^{\text{imp}}$:

$$\text{pos}_{t+1}^{\text{col}} = \text{pos}_t^{\text{col}} + (\beta \times \delta - 1) \otimes (\text{pos}_t^{\text{imp}} - \text{pos}_t^{\text{col}}).$$

The recommended value for $\beta$ is 4 (Lin et al., 2013). To pre-empt pre-mature convergence, different versions of assimilation movements were suggested, see for example, Kaveh and Talatahari (2010a), Kaveh and Talatahari (2010b) and Talatahari et al. (2012). In our version, we only change the position of the country if the cost of the new position given by Eq. (9) is lower than the cost of the previous position. This strategy has marginally increased the speed of the algorithm.

As with PSO and other evolutionary algorithms, search particles or countries may become outlying and take on unacceptable positions. Let $\text{pos}_j$ be the $j$th dimension of $\text{pos}_{t+1}$, and $L_j$ and $U_j$ be the lower and upper bounds of $\text{pos}_j$, respectively. The original ICA defined $\text{pos}_j = \max(\min(\text{pos}_j, U_j), L_j)$ to restrict the out-of-bound $\text{pos}_j$ to their nearest boundary. However, this is not applicable for a design problem when $\text{pos}_j$ is the position of the weight. In addition, this method of boundary handling can generate singular designs by assigning more than one level to $L_j$ and $U_j$. This means, from a design point of view, the original strategy reduces the number of support points and can generate singular designs. To overcome this, we implement the following strategy proposed by Lin et al. (2013) for correcting out-of-bound positions.

$$\text{pos}_j = \begin{cases} 
2U_j - \left( \text{pos}_j - \frac{\text{pos}_j - U_j}{U_j - L_j} (U_j - L_j) \right) & \text{if } \text{pos}_j > U_j \\
2L_j - \left( \text{pos}_j + \frac{L_j - \text{pos}_j}{U_j - L_j} (U_j - L_j) \right) & \text{if } \text{pos}_j \leq L_j.
\end{cases}$$

This method also allows us to freely choose a large value for $\beta$ and improve the exploration of the algorithm without worrying about getting stuck at the boundary of the search space.

### 3.3. Revolution within an empire

Revolution brings a sudden random change in the position of some colonies within each empire in the search space. Revolution may increase the exploration and prevent the pre-mature convergence. In this step, the number of colonies to be replaced is equal to the product of the “revolution rate” and the number of colonies of that empire. Typically, the initial value for the revolution rate is set to 0.3 and then gradually reduces in subsequent iterations by multiplying a parameter called “damp ratio”. The value for damp ratio is set to 0.99 (Lin et al., 2013). A large value for the revolution rate reinforces the exploration, while a small value encourages exploitation.

### 3.4. Imperialist update

A colony may reach a lower cost than its corresponding imperialist’s cost after moving to a new position in an assimilation or revolution step. To update the empire, the colony will be swapped with its corresponding imperialist if it is in a better position than the imperialist.

### 3.5. Uniting similar empires

If the distance between two imperialists is smaller than the product of the “uniting threshold” and the largest distance in the search space, then ICA unites these two empires. The value for the “uniting threshold” is 0.02 (Lin et al., 2013).

### 3.6. Imperialistic competition

During the competition among the imperialists, the powerful empires take possession of the colonies of the weaker empires. ICA assigns a possession probability to each of the empires. In this scheme, the powerful empires have better chances to possess colonies from the weakest empire. The probability of the possession for empire $i (i = 1, \ldots, N_{\text{imp}})$ is calculated as follows

$$\text{Prob}_i = \frac{E_i}{N_{\text{imp}} \sum_{j=1}^{N_{\text{imp}}} E_j},$$

where $E_i$ is the energy of the empire $i$. This energy can be calculated as

$$E_i = E_{\text{assim}} + E_{\text{rev}} + E_{\text{comp}}.$$
where

\[ E_i = Q_i - \max_{1 \leq j \leq N_{\text{imp}}} \{Q_j\} \tag{12} \]

and

\[ Q_i = c_i - \zeta \frac{1}{N_i} \sum_{r=1}^{N_i} d_{r,i}, \tag{13} \]

where \( E_i \) is the “power” of empire \( i \), \( Q_i \) is the “total cost” of empire \( i \), \( d_{r,i} \) is the cost of colony \( r \) of empire \( i \) and \( N_i \) is the number of colonies in empire \( i \). The value of \( \zeta \) is fixed and set equal to 0.1 (Atashpaz-Gargari and Lucas, 2007). When an empire has no more colonies, it is eliminated. Competition in every generation can cause pre-mature convergence. To overcome this, we use the “competition pressure” (default value 0.11 (Atashpaz-Gargari, 2008)) to control the probability of competition in each iteration (Lin et al., 2013). ICA performs a competition if and only if a randomly generated number between 0 and 1 exceeds 0.11.

The imperialist competition step concludes the stage wise performance of ICA. Algorithm 1 is the pseudo code of ICA we used to generate all designs in this paper. The ICA continues until some user-specified stopping rule is met. The stopping rule can be a maximum number of (outer) iterations or when all empires have collapsed and only one remains. In a design problem, Atwood’s efficiency lower bound for a design can also be used as a stopping rule. However, it is based on an equivalence theorem and so it can be time-expensive because ascertaining the efficiency lower bound of the ICA-generated design \( \xi_i \) at the \( i \)th iteration requires two other optimization problems to find \( \mu_i^* \) the probability measure on \( \mathcal{A}(\xi_i) \) that maximizes \( c(\mathbf{x}_{\xi_i}, \mu, \xi_i) \) over all probability measures \( \mu \), where \( \mathbf{x}_{\xi_i} \) is any support point of \( \xi_i \) and also \( \max_{x \in \chi} c(x, \mu_i^*, \xi_i) \). We chose maximum number of iterations as the stopping rule in ICA.

It is common practice to hybrid an algorithm with another to improve performance. For example, Lin et al. (2013) and Hosseini et al. (2014) proposed to apply a local search in ICA to improve the quality of the solutions. Local search is a technique to search the neighborhood of a solution to improve its quality. Our experience is that for optimal design problems, adding a local search strategy to ICA can be very efficient because after some iterations, the algorithm finds a nearly optimal design. A local search can also help ICA search the neighborhood around the current best designs and enable escape from a design it found pre-maturely. Performing local search for all of the countries can be very time-consuming, especially in our case in which every cost evaluation calls another optimization process. Based on Lin et al. (2013), we applied a local search for imperialists, before an assimilation step. Lin et al. (2013) provides technical details and the pseudo code of the local search.

3.7. Tuning parameters

Table 1 lists the ICA tuning parameters and their recommended values. ICA has several tuning parameters but only a few of them seem to influence the convergence of the algorithm. Lin et al. (2013) tested ICA on six famous benchmark problems and showed that once (9) and (10) are implemented, the revolution step and the uniting similar empires step do not seem to have much impact on the final solutions. In the current version, we used the default values in the original proposal and also note that changes in the values of these parameters do not seem to alter the computational time noticeably. The important tuning parameters in Table 1 are the assimilation coefficient \( \beta \), the total number of countries \( N_{\text{count}} \) and the number of imperialists \( N_{\text{imp}} \).
Table 1
ICA tuning parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Step</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_{\text{count}})</td>
<td>Initialization</td>
<td>–</td>
</tr>
<tr>
<td>(N_{\text{imp}})</td>
<td>Initialization</td>
<td>(%10) of (N_{\text{count}})</td>
</tr>
<tr>
<td>Assimilation coefficient</td>
<td>Assimilation</td>
<td>4</td>
</tr>
<tr>
<td>Revolution rate</td>
<td>Revolution</td>
<td>0.3</td>
</tr>
<tr>
<td>Damp ratio</td>
<td>Revolution</td>
<td>0.99</td>
</tr>
<tr>
<td>Uniting threshold</td>
<td>Uniting</td>
<td>0.02</td>
</tr>
<tr>
<td>(\xi)</td>
<td>Competition</td>
<td>0.1</td>
</tr>
<tr>
<td>Competition pressure</td>
<td>Competition</td>
<td>0.11</td>
</tr>
</tbody>
</table>

A large value for \(\beta\) encourages exploration, while a small value assists the exploitation. In our examples, the recommended value of \(\beta = 4\) seems to be a good balance between exploitation and exploration so that it prevents premature convergence due to over-exploitation or slow convergence because of over-exploration of the search space. In a more advanced method a fuzzy system may be used to adaptively adjust the tuning parameter of the assimilation step (see Khaled and Hosseini, 2014). \(N_{\text{count}}\) should be set based on the size of the design space and the region of uncertainty. We chose \(N_{\text{imp}}\) to be \(\%10\) of \(N_{\text{count}}\). By this strategy, we only need to choose the total number of countries for each problem. A small value of \(N_{\text{imp}}\) may result in having only one empire in the earliest iterations and a pre-mature convergence. Our first strategy to solve a pre-mature convergence is to increase \(N_{\text{count}}\).

Depending on the method that solves the inner problem, the minimax version of ICA may require some extra tuning parameters. In our implementation, only one tuning parameter, the maximum number of function evaluations \(\text{maxeval}\) that comes from the turning parameters of \(\text{direct}\) was added. A large value for \(\text{maxeval}\) increases the chance of escaping from local optima and helps to return the global optimum of the inner problem in \(\text{direct}\); on the other hand, it inflates the computational time. A conservative strategy is to set a large value, say 2000. Another strategy is to set \(\text{maxeval}\) to a smaller value, say 500, and after finding the output design, say \(\xi^*\), check whether increasing the value of \(\text{maxeval}\) will return a better solution of \(\max_{\theta \in \Theta} \psi(\xi^* , \theta)\) than the current one or not. If a better solution is found, the output design should not be trusted and the algorithm should be repeated with a higher value of \(\text{maxeval}\). This checking procedure makes sense to be run only on the output design because the number of local optima of the inner problem usually increases as the algorithm approaches the optimal design.

In what is to follow, we show ICA produces minimax and standardized maximin \(D\)-optimal designs for several nonlinear models that agree with those published in the recent literature obtained using different methods. All ICA-generated designs in this paper were generated using an AMD eight-core processor running 64-bits Windows 7 operating system with 3.50 GHz. ICA was written in R version 3.1.2. To speed up the algorithm, all of the Fisher information matrices were written in C++ and called within R using Rcpp package (Eddelbuettel et al., 2011).

4. Minimax \(D\)-optimal designs for the power logistic model

The power logistic model is given by

\[
f(x, \theta) = \left(\frac{1}{1 + \exp(-b(x - a))}\right)^s, \tag{14}\]

where \(\theta = (a, b)\). In the minimax approach, we suppose \(\theta\) belongs to the region of uncertainty \(\Theta = [a^L, a^U] \times [b^L, b^U]\), where \(a^L, a^U, b^L, b^U\) are known limits of the lower and upper bounds for \(a\) and \(b\). This model is sometimes called the 3-parameter logistic model and is more flexible than the simple logistic model where \(s = 1\). When \(s \neq 1\), heteroscedasticity and skewness can be accommodated by the power logistic model but not the simple logistic model. Finding optimal designs for the simple logistic model is also much simpler because the \(D\)-optimal design is symmetric around point \(a_M = (a^L + a^U)/2\) and one can use this property to reduce the dimension of the search space by half (King and Wong, 2000). Chen et al. (2014) used this property and applied the nested-PSO to find minimax optimal designs for the simple logistic model.

There have been previous attempts to find \(D\)-minimax optimal designs for the power logistic model. For example, King and Wong (2004) used a Fedorov–Wynn type algorithm to find several such optimal designs. More recently, Duarte and Wong (2013) used a semi-infinite programming (SIP) approach to search for minimax \(D\)-optimal designs for the power logistic model. Both gave similar optimal designs and both assumed the power parameter is fixed and the other two parameters are assumed to lie inside known intervals.

Table 2 displays minimin \(D\)-optimal designs \(\xi^*\) obtained from ICA for selected choices for \(\theta\) when \(s = 0.2\) and \(s = 1\). Column 1 lists the set \(\Theta\) and Column 2 lists each support point one on top of the other with its corresponding weight in parentheses. Column 3 shows the two-dimensional elements in (3) along with the weight distribution for \(\mu^*\) in (4). The rest of the columns in the table display the minimin \(D\)-criterion value, \(\max_{\theta \in \Theta} - \log |M(\xi^*, \theta)|\). Atwood’s \(D\)-efficiency lower bound, the maximum number of (outer) iterations and the CPU time (in seconds) required to find the optimal designs. Here, we set \(\text{maxeval}\) equal to 400; for \(s = 1\), we set \(N_{\text{count}}\) equal to 40 and increased it to 500 for \(s = 0.2\). A general observation is that the number of support points increases as the size of \(\Theta\) increases corresponding to larger uncertainty in the model.
Table 2
ICA-generated minimax D-optimal designs for the power logistic model when \( s = 0.2 \) and \( s = 1 \).

<table>
<thead>
<tr>
<th>( \Theta )</th>
<th>( \zeta^* )</th>
<th>( A(\zeta^*) )</th>
<th>Minimax D-criterion value</th>
<th>Atwood’s efficiency lower bound</th>
<th>Maximum number of iterations</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s = 0.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>([0, 0] \times [1, 1])</td>
<td>(-7.8769(0.5000)) (0.06949(0.5000))</td>
<td>(0.0000; 1.0000) (1.0000)</td>
<td>4.5202</td>
<td>1.0000</td>
<td>100</td>
<td>2.04</td>
</tr>
<tr>
<td>([0, 0] \times [1, 1.5])</td>
<td>(-6.4145(0.5000)) (0.0494(0.5000))</td>
<td>(0.0000; 1.0000) (0.5000)</td>
<td>4.5457</td>
<td>0.9998</td>
<td>40</td>
<td>323.60</td>
</tr>
<tr>
<td>([0, 0] \times [1, 3])</td>
<td>(-4.3814(0.4998)) (0.0211(0.5002))</td>
<td>(0.0000; 1.0000) (0.5000)</td>
<td>4.7120</td>
<td>0.9890</td>
<td>80</td>
<td>643.46</td>
</tr>
<tr>
<td>([0, 1] \times [1, 1])</td>
<td>(-7.3876(0.5000)) (0.5631(0.5000))</td>
<td>(0.0000; 1.0000) (0.5000)</td>
<td>4.5641</td>
<td>0.9999</td>
<td>60</td>
<td>474.00</td>
</tr>
<tr>
<td>([0, 1] \times [1, 1.5])</td>
<td>(-5.5512(0.5000)) (0.1676(0.5000))</td>
<td>(1.0000; 1.0000) (0.3333) (0.3333) (0.3333) (0.3333)</td>
<td>4.6295</td>
<td>0.9863</td>
<td>100</td>
<td>686.72</td>
</tr>
<tr>
<td>([0, 1] \times [1, 3])</td>
<td>(-3.4314(0.4533)) (0.0781(0.3466)) (1.0817(0.2001))</td>
<td>(0.0000; 1.0000) (0.3827) (0.3827) (0.3827) (0.3827)</td>
<td>4.9015</td>
<td>0.9968</td>
<td>200</td>
<td>1680.90</td>
</tr>
<tr>
<td>([0, 3] \times [1, 1])</td>
<td>(-7.0440(0.4371)) (0.1500(0.3323)) (2.8090(0.2306))</td>
<td>(0.0000; 1.0000) (0.5348) (0.5348) (0.5348) (0.5348)</td>
<td>4.8213</td>
<td>0.9985</td>
<td>200</td>
<td>1927.99</td>
</tr>
<tr>
<td>([0, 3] \times [1, 1.5])</td>
<td>(-4.5515(0.4100)) (0.2130(0.3723)) (2.8075(0.2177))</td>
<td>(0.0000; 1.5000) (0.5053) (0.5053) (0.5053) (0.5053)</td>
<td>4.9346</td>
<td>0.9937</td>
<td>200</td>
<td>1554.22</td>
</tr>
<tr>
<td>([0, 3] \times [1, 3])</td>
<td>(-2.7284(0.3593)) (0.1915(0.3216)) (1.3726(0.1250)) (2.1959(0.0636)) (2.9292(0.1303))</td>
<td>(0.0000; 1.5000) (1.5000) (1.5000) (1.5000) (1.5000)</td>
<td>5.2650</td>
<td>0.9687</td>
<td>2000</td>
<td>18579.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( s = 1 )</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>([0, 2] \times [1, 1.25])</td>
<td>(-0.8273(0.3466)) (1.0000(0.3068)) (2.8273(0.3466))</td>
<td>(0.0000; 1.2500) (0.5000) (0.5000) (0.5000) (0.5000)</td>
<td>3.3641</td>
<td>1.0000</td>
<td>50</td>
<td>46.51</td>
</tr>
<tr>
<td>([0, 3.5] \times [1, 1.25])</td>
<td>(-0.8811(0.2746)) (1.7500(0.4508)) (4.3811(0.2746))</td>
<td>(0.0000; 1.2500) (0.5000) (0.5000) (0.5000) (0.5000)</td>
<td>3.6582</td>
<td>1.0000</td>
<td>100</td>
<td>88.53</td>
</tr>
<tr>
<td>([0, 2] \times [1, 3])</td>
<td>(-0.3388(0.2591)) (0.8604(0.2409)) (1.1396(0.2409)) (2.3388(0.2591))</td>
<td>(0.0000; 3.0000) (0.5000) (0.5000) (0.5000)</td>
<td>3.9846</td>
<td>1.0000</td>
<td>100</td>
<td>107.76</td>
</tr>
<tr>
<td>([0, 3] \times [1, 2.5])</td>
<td>(-0.5193(0.2430)) (0.7320(0.2570)) (2.2680(0.2570)) (3.5193(0.2430))</td>
<td>(0.0000; 2.5000) (0.4814) (0.4814) (0.4814)</td>
<td>4.2254</td>
<td>1.0000</td>
<td>300</td>
<td>283.64</td>
</tr>
<tr>
<td>([0, 3] \times [1, 3])</td>
<td>(-0.3806(0.2221)) (0.6240(0.2363)) (1.5000(0.0833)) (2.3760(0.2363)) (3.8060(0.2221))</td>
<td>(0.0000; 3.0000) (0.3873) (0.3873) (0.3873)</td>
<td>4.4081</td>
<td>0.9927</td>
<td>500</td>
<td>488.86</td>
</tr>
<tr>
<td>([0, 3.5] \times [1, 3.5])</td>
<td>(-0.3511(0.1801)) (0.5993(0.2109)) (1.3856(0.1090)) (2.1144(0.1090)) (2.9007(0.2109)) (3.8511(0.1801))</td>
<td>(0.0000; 3.5000) (0.3570) (0.3570)</td>
<td>4.7654</td>
<td>0.9991</td>
<td>500</td>
<td>463.43</td>
</tr>
</tbody>
</table>

parameter values. This finding is consistent with results reported in King and Wong (2004) and Duarte and Wong (2013). Another finding is that when the set \( \Theta \) is large, it is harder to find the optimal designs. For example, the largest size of \( \Theta \) considered in our work for \( s = 0.2 \) is \( \Theta = [0, 3] \times [1, 3] \) and for this case, the ICA found a design only after 2000 iterations and more than 18 000 s in CPU computing time. From Atwood’s efficiency lower bound, this ICA-generated design has at least 96% efficiency, which is close to the corresponding value of 96.87% for the SIP-generated design for this example. The ICA generated design has an efficiency of 99.93% relative to the SIP-based generated design, suggesting the two generated designs are nearly identical. Note that we are not near enough to the true optimal designs for both generated designs to
(a) ICA evolution plot for $b_U = 1.5$. (b) Derivative plot for $b_U = 1.5$.

(c) ICA evolution plot for $b_U = 3$. (d) Derivative plot for $b_U = 3$.

Fig. 1. ICA evolution plots (left) and the corresponding directional derivative plots (right) of the design criterion evaluated at the ICA-generated designs for the power logistic model when $s = 0.2$ and $\Theta = [0, 3] \times [1, b_U]$ with $b_U = 1.5$ (top) and $b_U = 3$ (bottom). The derivative plot in (b) confirms the optimality and the plot in (d) confirms the near optimality of the ICA generated designs.

have a precise conclusion about the true number of support points of the optimal designs. When $s = 1$, ICA converges very quickly to the minimax $D$-optimal design.

Fig. 1 shows examples of ICA evolution plots and derivative plots for the ICA-generated designs when $s = 0.2$, $\Theta = [0, 3] \times [1, b_U]$ and $b_U = 1.5$ or $b_U = 3$. The derivative confirms optimality of the ICA-generated designs for $b_U = 1.5$ and nearly so for case when $b_U = 3$. In the evolution plot, the solid red line is the cost value of the best imperialist and the dotted blue line is the mean of all imperialists’ costs. The evolution plots show how ICA’s best imperialist quickly moves toward the optimal design after only a few iterations.

5. Standardized maximin $D$-optimal designs

Minimax optimal designs are essentially equivalent to maximin optimal designs. These optimal designs can have very different criterion values depending on the nominal set of parameter values. Accordingly, it is desirable to standardize the criterion and control for the potentially widely varying magnitude of the criterion.
A design $\xi^*$ is a standardized maximin $D$-optimal design (Dette, 1997) if

$$
\xi^* = \arg \max_{\xi \in \mathcal{D}} \inf_{\theta \in \Theta} \left[ \left( \frac{|M(\xi, \theta)|}{|M(\xi^*, \theta)|} \right)^{\frac{1}{p}} \right],
$$

where $\xi^*$ is the locally $D$-optimal design for the parameter $\theta$. The general equivalence theorem for verifying whether a design is standardized maximin is obtained by replacing $A(\xi^*)$ in Eq. (3) by $N(\xi^*)$, where

$$
N(\xi^*) = \left\{ \nu \in \Theta \mid \left( \frac{|M(\xi^*, \nu)|}{|M(\xi^*, \theta)|} \right)^{\frac{1}{p}} = \inf_{\theta \in \Theta} \left( \frac{|M(\xi^*, \nu)|}{|M(\xi^*, \theta)|} \right)^{\frac{1}{p}} \right\}. \tag{16}
$$

Often, analytical solutions for standardized maximin optimal designs are available for simple models. For complicated models, a numerical approach is required to find the standardized maximin optimal designs. In this section, we study the standardized maximin $D$-optimal designs for the log-linear, exponential growth, enzyme kinetic and Michaelis–Menten models. Note that the analytical solutions for locally $D$-optimal designs are available for all these models.

### 5.1. Log-linear model

The log-linear model is frequently used in dose-finding studies and is given by

$$
E(y) = \theta_0 + \theta_1 \log(x + \theta_2), \tag{17}
$$

where $y$ is the continuous outcome, $x \in \chi = [0, x_0]$ and $x_0$ is known (Bretz et al., 2010). The design problem is to determine the optimal number of dose levels ($k$) for the study, the optimal doses $x_i$, $i = 1, \ldots, k$ in $\chi$ and the optimal proportion of patients $w_i$ at $x_i$, $i = 1, \ldots, k$. Previous work showed that the locally $D$-optimal design for this model is supported at three points, with two at the endpoints of the design interval, with the remaining dose dependent only on the value of $\theta_2$ (Dette et al., 2010).

Table 3 shows ICA generated standardized maximin $D$-optimal designs for $\theta_2 \in \Theta = [1, \theta_2^{U}]$ when the dose range is $\chi = [0, 150]$. Column 4 lists the standardized maximin $D$-criterion value of the design $\xi^*$, which is

$$
\inf_{\theta_2 \in \Theta} \left[ \left( \frac{|M(\xi^*, \theta)|}{|M(\theta_2, \theta)|} \right)^{\frac{1}{p}} \right].
$$

The values we used for $N_{\text{count}}$ and $\text{maxeval}$ to generate the standardized maximin optimal designs were 600 and 1000, respectively, and the maximum number of (outer) iterations was set to 200. Our results suggest that the standardized maximin $D$-optimal designs require up to four dose levels when $\theta_2^{U}$ is large.

### 5.2. Exponential growth model

The two-parameter exponential growth model is commonly used for analyzing crop yield and is given by

$$
E(Y) = a + \exp(-bx).
$$

A direct calculation shows that the locally $D$-optimal design is independent of the nominal value of $a$ and is equally supported at $x = 0$ and $x = 1/b$ (Dette and Neugebauer, 1997). Duarte and Wong (2013) used SIP-based methods to generate...
Table 4
ICA-generated standardized maximin D-optimal designs on the design space $\chi = [0, 1]$ for the exponential growth model for selected uncertainty regions $\Theta = [1, b^U]$.

<table>
<thead>
<tr>
<th>$b^U$</th>
<th>$\xi^*$</th>
<th>$N(\xi^*)$</th>
<th>Standardized maximin D-criterion value</th>
<th>Standardized maximin D-efficiency lower bound</th>
<th>Maximum number of iterations</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000(0.5000)</td>
<td>1.0000(0.5000)</td>
<td>(1.0000; 1.0000)(1.0000)</td>
<td>1.0000</td>
<td>1.0000</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>0.0000(0.5000)</td>
<td>0.6931(0.5000)</td>
<td>(1.0000; 1.0000)(0.5000)</td>
<td>0.9421</td>
<td>0.9977</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>0.0000(0.4323)</td>
<td>0.2823(0.3068)</td>
<td>(1.0000; 1.0000)(1.3711)</td>
<td>0.8260</td>
<td>0.9989</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>0.0000(0.3784)</td>
<td>0.4144(0.2283)</td>
<td>(1.0000; 1.0000)(0.3867)</td>
<td>0.7809</td>
<td>0.9874</td>
<td>1000</td>
</tr>
<tr>
<td>50</td>
<td>0.0000(0.2645)</td>
<td>0.1303(0.1876)</td>
<td>(1.0000; 1.0000)(0.1979)</td>
<td>0.7319</td>
<td>0.9748</td>
<td>1000</td>
</tr>
</tbody>
</table>

The standardized maximin D-optimal designs for this model. Here, we assume that the user-selected region of uncertainty for $b$ has the form $\Theta = [1, b^U]$ and $b^U$ is known. Our goal is to find an approximate design $\xi^*$ on user-specified design space $\chi = [0, 1]$ that maximizes $\inf_{b \in \Theta} \left( \frac{|M(b, b^U)}{W(b, b^U)} \right)^{1/2}$.

Table 4 displays ICA generated standardized maximin designs for selected values of $b^U$. Clearly, if $b^U = 1$, the reported design is the locally $D$-optimal design. To find the ICA-generated standardized maximin designs for $b^U = 2, 5, 10$, we set $N_{\text{count}}$ and $\text{maxeval}$ equal to 200 and 600, respectively. When $b^U$ is larger, say $b^U = 50$, i.e. the uncertainty space $\Theta$ is larger, more iterations are required to find the optimal design and so to avoid the premature convergence, we increased the value of $N_{\text{count}}$ to 500. We were able to verify that all the ICA-generated designs are in good agreement with the SIP-based generated designs. As expected, the standardized maximin optimal designs have more support points as $b^U$ gets larger.

5.3. Enzyme kinetic models

Enzyme inhibition kinetic models are applied in pharmaceutical studies to study the rate of enzyme kinetics reaction (Bogacka et al., 2011). Let $S$ be the substrate concentration and $I$ be the inhibitor concentration. There are generally four basic ways to model the mean velocity of the reaction rate using these two explanatory variables:

Competitive inhibition model:

$$\eta = \frac{V S}{K_m \left(1 + \frac{I}{K_i}\right) + S};$$  \hspace{1cm} (19)

Noncompetitive inhibition model:

$$\eta = \frac{V S}{(K_m + S) \left(1 + \frac{I}{K_m}\right)};$$  \hspace{1cm} (20)

Uncompetitive inhibition model

$$\eta = \frac{V S}{K_m + S \left(1 + \frac{I}{K_m}\right)};$$  \hspace{1cm} (21)

Mixed inhibition model:

$$\eta = \frac{V S}{K_m \left(1 + \frac{I}{K_i}\right) + S \left(1 + \frac{I}{K_m}\right)}.$$

Here $V$ denotes the maximum velocity of the enzyme, $K_i$ and $K_m$ are the dissociation constants and $K_m$ is the Michaelis–Menten constant. Bogacka et al. (2011) provided closed-form formulae for the locally $D$-optimal designs of these models on $\chi = [S^U, I^U] \times [I^U, I^U]$, where $S^U$, $S^U$, $I^U$ and $L^U$ are user-specified. The formulae do not depend on parameter $V$. 

Apart from the uninteresting parameter $V$, the assumed region of plausible values of $\Theta$ for each model has the following form: $\Theta = [K_m^L, K_m^U] \times [K_i^L, K_i^U]$ for the competitive and noncompetitive models; $\Theta = [K_m^L, K_m^U] \times [K_i^L, K_i^U]$ for the uncompetitive model and $\Theta = [K_m^L, K_m^U] \times [K_i^L, K_i^U] \times [K_i^L, K_i^U]$ for the mixed type inhibition model.

Standardized maximin $D$-optimal designs for these models were found using nested-PSO and SIP-based methods, respectively, in Chen et al. (2014) and Duarte and Wong (2013). To compare results from ICA, we assumed the same set-up with $\chi = [0, 30] \times [0, 60]$ and $K_i^L = K_i^U = 4, K_m^L = K_m^U = 5, K_i^L = 2$ and $K_i^U = 3$. The ICA-generated standardized maximin $D$-optimal designs are given in Table 5 and they agree quite closely with those reported in Chen et al. (2014) and Duarte and Wong (2013). For this example, the values we used for $N_{count}$ and maxeval were 100 and 300, respectively. Each of the ICA-generated designs has also been verified by the equivalence theorem. As examples, Fig. 2 displays the plots of the directional derivatives of the criterion evaluated at the ICA-generated designs for the competitive and the mixed inhibition models and they confirm the optimality of the ICA-generated designs.

We note that when there is no inhibition, the mean velocity is modeled by the popular Michaelis–Menten function given by

$$E(Y) = \frac{VS}{K_m + S}, \quad S \in [0, S_0].$$

where $S_0$ is user-specified. The locally $D$-optimal design for this reduced model is equally supported at $S = \frac{K_m S_0}{2K_m + S_0}$ and $S = S_0$ (Rasch, 1990) and analytical standardized maximin optimal designs within the class of equally weighted saturated designs were reported in Dette and Biedermann (2003). Since there is no guarantee that the resulting optimal design found within this class is globally optimal, they then employed the Nelder–Mead method to find the standardized maximin optimal designs. We applied ICA and generated standardized maximin $D$-optimal designs. Table 6 shows the optimal designs for $\Theta \in [100, K_i^U]$ and $S_0 = 2000$ and they are all consistent with those reported in Dette and Biedermann (2003). The values we used to implement ICA to find these optimal designs were $N_{count} = 60$ and maxeval = 300. Unlike the two-dimensional enzyme kinetic models, it is interesting to observe that the numerical results show that the standardized maximin optimal designs are not always equally weighted for the Michaelis–Menten model. This happens when the optimal design has three points, as in the last case in Table 6 when $K_m^U = 2000$. 

**Table 5**
ICA-generated standardized maximin $D$-optimal designs for the competitive, noncompetitive, uncompetitive and mixed inhibition models.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\xi^*$</th>
<th>$N(\xi^*)$</th>
<th>Standardized maximin $D$-criterion value</th>
<th>Standardized maximin $D$-efficiency lower bound</th>
<th>Maximum number of iterations</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competitive</td>
<td>(3.4432, 0.0000), (30.0000, 0.0000), (30.0000, 18.8954)</td>
<td>18.8954, 0.3337</td>
<td>0.9836</td>
<td>1.0000</td>
<td>200</td>
<td>986.93</td>
</tr>
<tr>
<td>Noncompetitive</td>
<td>(3.4429, 0.0000), (30.0000, 0.0000), (30.0000, 2.4495)</td>
<td>2.4495, 0.3333</td>
<td>0.9916</td>
<td>0.9999</td>
<td>200</td>
<td>824.27</td>
</tr>
<tr>
<td>Uncompetitive</td>
<td>(3.4451, 0.0000), (30.0000, 0.0000), (30.0000, 2.4495)</td>
<td>2.4495, 0.3333</td>
<td>0.9958</td>
<td>0.9991</td>
<td>400</td>
<td>1735.49</td>
</tr>
<tr>
<td>Mixed</td>
<td>(3.4614, 0.0000), (4.2801, 1.4216), (30.0000, 0.0000), (30.0000, 4.0373)</td>
<td>4.0373, 0.2500</td>
<td>0.9900</td>
<td>0.9991</td>
<td>1000</td>
<td>6212.32</td>
</tr>
</tbody>
</table>

**Table 6**
ICA-generated standardized maximin $D$-optimal designs for the Michaelis–Menten model for selected uncertainty regions $\Theta \in [100, K_i^U]$.

<table>
<thead>
<tr>
<th>$K_m^U$</th>
<th>$\xi^*$</th>
<th>$N(\xi^*)$</th>
<th>Standardized maximin $D$-criterion value</th>
<th>Standardized maximin $D$-efficiency lower bound</th>
<th>Maximum number of iterations</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>(177.8260, (0.5000), (0.5000))</td>
<td>(100.0000, (0.5000), (0.5000))</td>
<td>0.8816</td>
<td>1.0000</td>
<td>60</td>
<td>106.28</td>
</tr>
<tr>
<td>1000</td>
<td>(115.0339, (0.2461), (0.2891), (0.4648))</td>
<td>(100.0000, (0.4811), (0.0518), (0.0518))</td>
<td>0.8134</td>
<td>0.9984</td>
<td>200</td>
<td>349.81</td>
</tr>
<tr>
<td>2000</td>
<td>(110.7724, (0.2362), (0.3197), (0.4441))</td>
<td>(100.0000, (0.4241), (0.0943), (0.0943))</td>
<td>0.7924</td>
<td>0.9995</td>
<td>250</td>
<td>427.18</td>
</tr>
</tbody>
</table>
6. Discussion

In this paper, we proposed and tested ICA as an alternative evolutionary algorithm to find minimax and standardized maximin optimal designs for nonlinear models. We applied the ICA algorithm and generated minimax optimal designs for the power logistic model, standardized maximin $D$-optimal designs for the log-linear model, exponential model and several inhibition models useful in enzyme kinetic studies. These are hard problems because the design criterion is non-differentiable and requires nested optimizations. To make the algorithm more effective, we hybridized ICA with a local search and the overall results suggest that such an approach was effective for solving our design problems. All our results agree with those reported in the literature where they were obtained using PSO and the more complex SIP-based methods. It is likely that our proposed hybrid ICA algorithm can also generate optimal designs for more general problems.

We conclude with two important remarks. First, unlike many other algorithms, ICA algorithm does not require the design space or the region of uncertainty to be discretized. This is an advantage because this means that the search for the support points of the optimal design is not be restricted to the grid points. Second, the inner optimization problem can be a very tricky problem to solve and advanced optimization methods may either fail to find all the locations of the global optima or
they require a huge amount of time to do so. As an example, consider the standardized maximin optimal design $\xi_0$ found in Duarte and Wong (2013) for the growth exponential model when $b \in [1, 50]$, $a = 1$ and $\chi = [0, 1]$:  

$$
\xi_0 = \begin{bmatrix}
0 & 0.0290 & 0.1232 & 0.3561 & 0.9416 \\
0.2699 & 0.1988 & 0.1839 & 0.1932 & 0.1542 
\end{bmatrix}.
$$

(24)

The inner optimization problem is to find $\inf_{b \in [1,50]} f(b)$ where $f(b) = \left( \frac{M(b_0 \mid b)}{M(b_0 \mid b)} \right)^b$ and $\xi_0$ is the locally $D$-optimal design when the nonlinear parameter has nominal value $b$. This function has five local minima at

$$
\begin{align*}
B_1 &= 1.00000, \quad f(b_1) = 0.732162 \\
B_2 &= 2.80671, \quad f(b_2) = 0.732112 \\
B_3 &= 7.22539, \quad f(b_3) = 0.732163 \\
B_4 &= 18.49032, \quad f(b_4) = 0.732208 \\
B_5 &= 50.00000, \quad f(b_5) = 0.731978 
\end{align*}
$$

(25)

with the global minimum attained at $b_5 = 50$. To solve the problem in R, we employed optimize from the stats package, psoptim from package pso (Bendtsen, 2012), gosolnp from the package Rsolnp (Ghalanos and Theussl, 2014), and isres and directL from the package nloptr to create an interface to the Nlopt library. Note that many of the implemented box-constraint optimization algorithms (at least in R) cannot be used naively because they put all out-of-range support points at the endpoints of the design interval and consequently produce singular designs. Table 7 reports the CPU times. Interestingly, optimize was trapped in a local optimum for all 100 repetitions. The current implementation of psoptim is very slow and requires an unreasonable amount of CPU time even when we have only 80 particles. gosolnp and directL are the best among them, but directL is approximately 10 times faster. This was our reason for choosing directL to solve the inner optimization problem.

The nested-PSO algorithm employed by Chen et al. (2014), as from its name, used another PSO algorithm to find the optima of the inner problem. In the R environment, we can use psoptim to solve this problem to implement the nested-PSO algorithm. However, the CPU times in Table 7 suggest that psoptim is much slower and less accurate than other alternatives, such as directL, gosolnp and isres for solving the inner problem. Since the speed of a minimax multistart algorithm like nested-PSO or ICA highly depends on solving the inner problem, a R implementation of the nested-PSO algorithm would likely appear much slower than ICA.

In summary, ICA produced results very similar to those available from the literature obtained from a handful of methods, such as the nested-PSO and SIP-based algorithms. Our conclusion is that ICA offers good promise for finding optimal designs for more complicated design problems, such as multiple-objective optimal designs and different types of Bayesian optimal designs. Our future plan is to apply ICA to tackle these and related challenging design problems and also disseminate reliable and develop user-friendly computer codes as a R package to help practitioners efficiently find optimal designs for non-linear models using ICA.

Acknowledgments

The authors would like to thank three anonymous referees and an Associate Editor for helpful comments and suggestions on an earlier version of the article. The research of Wong reported in this paper was partially supported by a grant award R01GM107639 from the National Institute of General Medical Sciences of the National Institutes of Health. The contents in this paper are solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References


Designing combined physical and computer experiments to maximize prediction accuracy

Erin R. Leatherman a,*, Angela M. Dean b, Thomas J. Santner b

a West Virginia University, United States
b The Ohio State University, United States

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form 18 July 2016
Accepted 19 July 2016
Available online 25 July 2016

Keywords:
Augmented nested LHD
Calibration
Gaussian process Kriging interpolator
IMSPE
Latin hypercube design
Simulator

ABSTRACT

Combined designs for experiments involving a physical system and a simulator of the physical system are evaluated in terms of their accuracy of predicting the mean of the physical system. Comparisons are made among designs that are (1) locally optimal under the minimum integrated mean squared prediction error criterion for the combined physical system and simulator experiments, (2) locally optimal for the physical or simulator experiments, with a fixed design for the component not being optimized, (3) maximin augmented nested Latin hypercube, and (4) I-optimal for the physical system experiment and maximin Latin hypercube for the simulator experiment. Computational methods are proposed for constructing the designs of interest. For a large test bed of examples, the empirical mean squared prediction errors are compared at a grid of inputs for each test surface using a statistically calibrated Bayesian predictor based on the data from each design. The prediction errors are also studied for a test bed that varies only the calibration parameter of the test surface. Design recommendations are given.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Combinations of observations from a physical system and a deterministic computer simulator of that system have been used, for example, to calibrate the simulator statistically, to optimize the physical system, and to achieve other objectives (see, for example, Higdon et al., 2004, 2008; Leatherman et al., 2014b). This paper considers settings where such data come from a physical experiment that varies the values of the inputs to the physical system and a computer experiment that varies the input values to the simulator code. The goal is to determine the initial design of a combined simulator and physical experiment with the objective of most accurately predicting the mean of the physical system using a statistically calibrated simulator.

For both physical and computer experiments there has been much research on the optimal design of experiments using intuitively defined criteria. For example the I-optimality criterion, which minimizes the integrated mean squared prediction error (MSPE) for a specified regression model, is a design metric for physical experiments that emphasizes prediction accuracy (Studden, 1977; Hardin and Sloane, 1993).

Designs that minimize the integrated MSPE (IMSPE) have also been proposed for computer experiments. However, because the analysis of simulator output ordinarily is based on non-parametric Kriging predictors, most simulator designs constructed in the literature are locally optimal, corresponding to a specification of correlation and other model parameters...

* Correspondence to: Department of Statistics, PO Box 6330, Morgantown, WV 26506, United States. Fax: +1 304 293 2272.
E-mail address: erleatherman@mail.wvu.edu (E.R. Leatherman).

http://dx.doi.org/10.1016/j.csda.2016.07.013
0167-9473/© 2016 Elsevier B.V. All rights reserved.
(although the designs of Leatherman et al., 2016, are constructed for a weighting of the parameter values). More often, the initial design of a computer experiment uses a “space-filling” criterion resulting in, for example, minimax designs (Johnson et al., 1990), minimum average reciprocal distance designs (Audze and Eglais, 1977; Welch, 1985; Liefvendahl and Stocki, 2006), and lattices, nets, and uniform designs (Niederreiter, 1978, 1992; Fang and Wang, 1994; Owen, 1995).

Previous work on the design of combined physical and simulator experiments includes studies of how to take follow-up runs. For example, Ranjan et al. (2011) and Williams et al. (2011) focus on batch sequential design optimization but use standard space-filling designs as the initial physical and simulator designs. The initial observations are used to estimate model parameters, and additional design points are added to improve the design’s measure of goodness, in particular to provide the maximum IMSPE reduction in Ranjan et al. (2011) and the maximum generalized expected improvement for global fit in Williams et al. (2011).

Using a calibrated Bayesian predictor for the mean of a physical system, this paper compares the accuracy of local IMSPE-optimal designs for combined physical and simulator experiments with maximin augmented nested Latin hypercube designs (MmANLHD) and other designs. The comparisons are based on the empirical mean squared prediction error (EMSEP) in a large test bed of examples. Section 2 describes the model used to relate the simulator experiment output and the physical system output. Section 3 gives the formulas for the MSPE and IMSPE and defines local IMSPE-optimal designs while Section 3.2 defines MmANLHDs. Sections 4.1 and 4.2 give algorithms for constructing local minimum IMSPE designs and MmANLHDs, respectively. Section 5 presents a study of the prediction accuracy of the initial combined designs. Fourteen designs are selected in Section 5.1 to compare across 18 corresponding physical and simulator test-bed families, where the test beds are described in Section 5.2. Section 5.4 compares the designs’ prediction accuracy across the surfaces using the EMSEP criterion defined in Section 5.3. An additional comparison of prediction accuracy is made in Section 6 where test beds are formed from stationary GP draws with \( \theta \neq 0.5 \times \mathbf{1}_d \), when the design used to collect training data for prediction is locally optimal for \( \theta = 0.5 \times \mathbf{1}_d \). A brief summary of the conclusions is given in Section 7.

2. Modeling combined simulator and physical outputs

In the simulator code let \( \mathbf{x}^i \) and \( t \) denote a \( d_x \times 1 \) vector of control inputs and a \( d_t \times 1 \) vector of calibration inputs, respectively. Control inputs can be ‘set’ by the researcher in both the physical experiment as well as in simulator runs. Calibration inputs can be varied in the simulator runs but are fixed and unknown in the associated physical experiment; for example, while the material properties of meniscal tissue are fixed values in a biomechanical cadaver study of stresses in the knee, a finite element simulator may regard these values as inputs. Let \( \mathbf{x}^p \) be a \( d_x \times 1 \) vector of control inputs in the physical experiment and \( \mathbf{\theta} \) be the true, but unknown, \( d_t \times 1 \) vector of calibration parameters. Assume the input space of the control variables is rectangular but transformed so that \( \mathbf{x}^c, \mathbf{x}^p \in [0, 1]^d_x \), while the input space of the calibration variables is also rectangular but transformed so that \( t, \mathbf{\theta} \in [0, 1]^d_t \). Finally let \( y^s(\mathbf{x}^c, t) \) and \( y^p(\mathbf{x}^p) \) denote the outputs from the simulator and physical experiments when run at \( \mathbf{x}^c, t \) and \( \mathbf{x}^p \), respectively.

Adopting the model of Kennedy and O’Hagan (2001), denoted KO hereafter, this paper regards the simulator output \( y^s(\mathbf{x}^c, t) \) as a draw from the Gaussian Process (GP)

\[
y^s(\mathbf{x}^c, t) = \sum_{i=1}^{k} f_i(\mathbf{x}^c, t) \beta_i + Z(\mathbf{x}^c, t) = f^T(\mathbf{x}^c, t) \beta + Z(\mathbf{x}^c, t),
\]

where \( f(\mathbf{x}^c, t) = (f_1(\mathbf{x}^c, t), f_2(\mathbf{x}^c, t), \ldots, f_k(\mathbf{x}^c, t))^T \) are known regression functions, \( \beta = (\beta_1, \beta_2, \ldots, \beta_k)^T \) is a vector of unknown regression coefficients, \( \mathbf{w}^T \) denotes the transpose of \( \mathbf{w} \), and \( Z(\cdot, \cdot) \) is a zero mean, stationary GP over \([0, 1]^{d_x+d_t}\) with process variance \( \sigma_z^2 \) and separable Gaussian correlation function:

\[
\text{Cor} \left( Y(\mathbf{x}^c_1, t_1), Y(\mathbf{x}^c_2, t_2) \right) = R_Z \left( (\mathbf{x}^c_1, t_1) - (\mathbf{x}^c_2, t_2) \right) \rho_{Z}\)
\[
= \prod_{j=1}^{d_x} R_Z(\mathbf{x}^c_1 - \mathbf{x}^c_2)^2 \prod_{j=1}^{d_t} R_Z(t_1 - t_2)^2,
\]

where \( \mathbf{x}^c_{ij} \) and \( t_{ij} \) are the jth inputs of \( \mathbf{x}^c_i \) and \( t_i \), respectively, \( i = 1, 2 \). The parameter \( \rho_{Z,j} \in [0, 1] \) is the correlation between outputs at inputs \( (\mathbf{x}^c_1, t_1) \) and \( (\mathbf{x}^c_2, t_2) \) that differ only in the jth input by half the range of this input.

Let \( \zeta(\mathbf{x}^p) = \zeta(\mathbf{x}^p, \mathbf{\theta}) \) denote the mean of the physical system run at input \( \mathbf{x}^p \). The output \( y^p(\mathbf{x}^p) \) is modeled as the sum of \( \zeta(\mathbf{x}^p) \) and a zero mean measurement error \( \epsilon(\mathbf{x}^p) \), so that \( y^p(\mathbf{x}^p) = \zeta(\mathbf{x}^p) + \epsilon(\mathbf{x}^p) \) where additional assumptions regarding \( \epsilon(\mathbf{x}^p) \) are stated below. The KO model assumes that the simulator, even when run at the true value of \( \mathbf{\theta} \), need not perfectly represent the underlying physical process because the mathematical model uses simplified physics or biology. Following KO, denote the simulator model bias (discrepancy) as

\[
\delta(\mathbf{x}^p) = \zeta(\mathbf{x}^p) - y^p(\mathbf{x}^p, \mathbf{\theta}),
\]
and assume that $\delta(x^p)$ can be regarded as a realization of $\Delta(x^p)$ which is a stationary, zero-mean GP over $[0,1]^d$ with process variance $\sigma^2_\delta$ and separable Gaussian correlation function

$$R_\delta(x_i^p - x_j^p \mid \rho_{s,t}) = \prod_{j=1}^{d_p} \rho_{s,t}^{-4(x_i^p - x_j^p)^2},$$

(3)

where $x_{i,j}$ is the $j$th element of input $x_i^p$, $i = 1, 2$, and $\rho_{s,t} \in [0,1]$ is interpreted analogously to $\rho_{x,i}$. The zero-mean assumption for $\Delta(x^p)$ is interpreted as saying that there is no global trend in $\delta(x^p)$ requiring regression modeling.

As in KO, assume $y^p(x^p)$ can be regarded as a realization of

$$Y^p(x^p) = Y^s(x^p, \theta) + \Delta(x^p) + \epsilon(x^p),$$

(4)

where $Y^s(x^p, \theta), \Delta(x^p)$, and $\epsilon(x^p)$ are mutually independent with the distribution of $Y^s(x^p, \theta)$ defined through (1) and (2), $\Delta(x^p)$ has distribution described in the previous paragraph, and $\epsilon(x^p)$ is a white noise process with variance $\sigma^2_\epsilon$.

Suppose that $n_s$ simulator observations $y^s = (y^s(x_1^p, t_1), y^s(x_2^p, t_2), \ldots, y^s(x_{n_s}^p, t_{n_s}))^T$ result from running a set of inputs specified by the rows of the $n_s \times (d_s + d_t)$ simulator design matrix

$$X_s = \begin{bmatrix} x_1^p & x_2^p & \cdots & x_{n_s}^p \\ t_1 & t_2 & \cdots & t_{n_s} \end{bmatrix}^T.$$

Additionally, suppose that $n_p$ physical observations $y^p = (y^p(x_1^p), y^p(x_2^p), \ldots, y^p(x_{n_p}^p))^T$ are to be collected at inputs which are the rows of the $n_p \times d_p$ physical design matrix $X_p = [x_1^p, x_2^p, \ldots, x_{n_p}^p]^T$. The combined physical and simulator observations are denoted by the $(n_p + n_s) \times 1$ vector $y = [y^p]^T, (y^s)^T]^T$.

In the experimental setting of this paper, the goal is to predict the mean of the physical system output $\zeta(x_0)$ at $x_0$ based on the physical and simulator training data. Focusing on the prediction of $\zeta(x_0) = y^s(x_0, \theta) + \delta(x_0)$ eliminates the problem of unidentifiability of predicting separately $\delta(\cdot)$ and $\theta$. When $\theta$ is unknown, while $\Omega = (\theta, \sigma^2_\delta, \rho_\delta, \sigma^2_\epsilon, \rho_\epsilon, \sigma^2_\zeta)$ in (1)-(3) are known, Sec 3.3 of Santner et al. (2003) presents the argument that shows that the best linear unbiased predictor (BLUP) of $\zeta(x_0)$ is

$$\widehat{\zeta}_{\text{BLUP}}(x_0) = f_0^T \hat{\beta} + v_0^T \Sigma^{-1}_y (y - F \hat{\beta}),$$

(5)

where $f_0 = f(x_0, \theta)$ is the $k \times 1$ vector of known regressors at control input $x_0$ and calibration parameter $\theta$; $\hat{\beta} = (F^T \Sigma^{-1}_y F)^{-1} F^T \Sigma^{-1}_y$ is the $k \times 1$ general least squares estimator of $\beta$; $F$ is a $(n_p + n_s) \times k$ matrix of known regressors with the first $n_p$ rows defined by $f_i(x_i^p, \theta^s)$ for $1 \leq i \leq n_p$ and $1 \leq j \leq k$, and the last $n_s$ rows defined by $f_j(x_j^p, t_j)$ for $1 \leq i \leq n_s$ and $1 \leq j \leq k; \text{and} \: \Sigma_0 = [(v_0^p)^T, (v_0^s)^T]^T$ is the $(n_p + n_s) \times 1$ vector of covariances, with the ith element of $v_0^p$ being

$$
\text{Cov}(Y^s(x_i^p, \theta) + \Delta(x_i^p), Y^s(x_j^p, \theta) + \Delta(x_j^p))
= \sigma^2_\zeta R_\zeta ((x_i^p, \theta) - (x_j^p, \theta) \mid \rho_\zeta) + \sigma^2_\delta R_\delta (x_i^p - x_j^p \mid \rho_\delta), \quad \text{for} \quad i = 1, \ldots, n_p,
$$

while the jth element of $v_0^s$ is

$$\text{Cov}(Y^s(x_i^p, \theta) + \Delta(x_i^p), Y^s(x_j^p, t_j)) = \sigma^2_\zeta R_\zeta ((x_i^p, \theta) - (x_j^p, t_j) \mid \rho_\zeta), \quad \text{for} \quad j = 1, \ldots, n_s.$$

Also, the $(n_p + n_s) \times (n_p + n_s)$ covariance matrix

$$\Sigma_y = \begin{pmatrix} \Sigma_{Z_{pp}} & \Sigma_{Z_{ps}} \\ \Sigma_{Z_{sp}} & \Sigma_{Z_{ss}} \end{pmatrix} = \begin{pmatrix} \Sigma_\delta + \Sigma_\epsilon & 0 \\ 0 & \Sigma_\delta + \Sigma_\epsilon \end{pmatrix} = \Sigma_\zeta + \begin{pmatrix} \Sigma_\delta & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix},$$

(6)

where $\Sigma_{Z_{pp}} = (\sigma^2_\zeta R_\zeta ((x_i^p, \theta) - (x_j^p, \theta) \mid \rho_\zeta))$ is $n_p \times n_p$, $\Sigma_{Z_{ps}} = (\sigma^2_\zeta R_\zeta ((x_i^p, \theta) - (x_j^p, t_j) \mid \rho_\zeta))$ is $n_p \times n_s$, $\Sigma_{Z_{sp}} = (\sigma^2_\zeta R_\zeta ((x_i^p, t_j) - (x_j^p, t_j) \mid \rho_\zeta))$ is $n_s \times n_p$, $\Sigma_\delta = (\sigma^2_\delta R_\delta (x_i^p - x_j^p \mid \rho_\delta))$ is $n_p \times n_p$, and $\Sigma_\epsilon = \sigma^2_\epsilon I_{n_p}$. In Section 3.1, the BLUP of $\zeta(x_0)$ in (5) will be used to define IMSPE and the designs that are locally optimal for the IMSPE measure.

The predictors used in the simulation comparisons of Sections 5 and 6 do not assume that $[\beta, \Omega]$ is known. Instead, the (fully) Bayesian predictor

$$\widehat{\zeta}_{\text{BF}}(x_0) = E \left[ Y^s(x^p, \theta) + \Delta(x^p) \mid y \right] = E_{[\beta, \theta, \Omega]} \left( E \left( Y^s(x^p, \theta) + \Delta(x^p) \mid \beta, \Omega, y \right) \right)
= E_{[\beta, \theta, \Omega]} \left( f_0^T \beta + v_0^T \Sigma^{-1}_y (y - F \beta) \right),$$

(7)

is used where a prior for $[\beta, \Omega]$ is assumed. Unfortunately, the predictor (7) can be analytically demanding, and, therefore, to make predictions in Sections 5 and 6, (7) is computed based on draws from $[\beta, \Omega \mid y]$ that are constructed using the Markov Chain Monte Carlo algorithm that is implemented in the GPM/SA software of Gattiker (2008).
3. Designs for combined physical and simulator experiments

In their initial research on calibration, KO speculated on the design of combined physical and simulator experiments. They noted that the physical design is often "not a matter of choice". Next, they suggested that the simulator design should cover both the control input space and calibration input space well, and that a sequential design is a good way to ensure the coverage of the calibration input space. Their final observation is that there should be simulator design points that are 'close' to physical design points in order to determine the model bias.

While physical designs and simulator designs have been studied extensively in their own right, the combination of these designs has received much less attention in the literature. Kanjan et al. (2011) and Williams et al. (2011) studied combined follow-up designs. However, to the authors' knowledge, the effect of the initial combined design has not yet been presented in the literature.

The following subsections define two design criteria for initial combined designs. The first is the local minimum IMSPE criterion described in Section 3.1. This criterion is prediction-based; i.e., it is defined using the physical and simulator output models and the BLUP from Section 2. The second is the MmANLHD criterion that is defined in Section 3.2. This geometrically-based criterion yields space-filling simulator designs with corresponding physical designs whose points replicate some of the control input values from the simulator design. Thus, this criterion follows the KO suggestion to align simulator and physical design points. A third design is used in Section 5 for comparison; this commonly-used design consists of an l-optimal design for the physical experiment paired with an MmLHD for the simulator experiment.

3.1. Locally optimal designs for minimizing the integrated mean squared prediction error

Designs constructed using the local minimum IMSPE criterion are meant to provide small expected prediction errors of the mean of the physical system on average across the control input space. The predictor \( \hat{\xi}_{\text{blup}}(x_0) \) in (5) depends on \( y \), the physical and simulator designs \( X^p \) and \( X^s \) (through \( F, R_y(\cdot | \rho_z) \), and \( R_0(\cdot | \rho_0) \)), and on \( \Omega = (\theta, \sigma^2_y, \rho_z, \sigma^2_z, \rho_0, \sigma^2_0) \). For a given \( X^p, X^s, \) and \( \Omega \), the MSPE of \( \hat{\xi}_{\text{blup}}(\cdot) \) at \( x_0 \in [0, 1]^6 \) is

\[
\text{MSPE} (x_0, X^p, X^s \mid \Omega) = E \left( (\hat{\xi}_{\text{blup}}(x_0) - (Y^i(x_0, \theta) + \Delta(x_0)))^2 \mid \Omega \right),
\]

where \( f_0, F, \nu_0, \) and \( \Sigma_y \) are defined below (5) and the expectation (8) is taken with respect to \( (Y^i(x_0, \theta) + \Delta(x_0), (Y^p)^T, (Y^s)^T) \), where \( Y^p \) and \( Y^s \) are the GP model analogs of \( y^p \) and \( y^s \).

To avoid the numerical non-invertibility of \( \Sigma_y \) when simulator design points \( \{x_i^s\} \) are “too close” together, simulator designs that are constructed using (8) are restricted to be an element of the set of all possible \( n_s \)-run simulator designs in \( d_s + d_t \) inputs where no two design rows are within \( \epsilon \)-ball of diameter \( b = 10^{-3} \) of each other, denoted by \( D_{n_s, d_s+b}^p \) hereafter. Note that the replication of physical design points does not pose a similar problem because of the measurement error term \( \epsilon \) in the physical model (4). Thus, physical designs will be selected from \( D_{n_s, d_s}^p \), the set of all possible \( n_s \)-run physical designs in \( d_s \) inputs.

Given parameters \( \Omega = (\theta, \sigma^2_y, \rho_z, \sigma^2_z, \rho_0, \sigma^2_0) \), the IMSPE of the predictor \( \hat{\xi}_{\text{blup}}(\cdot) \) using design \( (X^p, X^s) \) is

\[
\text{IMSPE} (X^p, X^s \mid \Omega) = \int_{[0, 1]^6} \text{MSPE} (x_0, X^p, X^s \mid \Omega) \, dx_0
\]

\[
= \sigma^2_y + \sigma_z^2 - \text{trace} \left[ \begin{bmatrix} 0 & F \end{bmatrix} \Sigma_y^{-1} \right] \int_{[0, 1]^6} \left( f_0 \Sigma_y f_0^T \nu_0 \right) \, dx_0.
\]

Given \( \Omega \), an \( (X^p, X^s) \) that minimizes Eq. (9) over \( \{D_{n_s, d_s+b}^p, D_{n_t, d_t+b}^s\} \) is said to be a local IMSPE-optimal combined design (w.r.t. \( \Omega \)). Equivalently, factoring (9) into \( \sigma^2_{x^*} \) times

\[
\text{IMSPE}^* (X^p, X^s \mid \Omega^*) \equiv \frac{1}{\sigma^2_{x^*}} \text{IMSPE} (X^p, X^s \mid \Omega),
\]

shows that a local IMSPE-optimal combined design minimizes IMSPE* \( (X^p, X^s \mid \Omega^*) \) and depends only on \( \Omega^* = (\theta, \rho_z, \sigma^2_y / \sigma^2_z, \rho_0, \sigma^2_0 / \sigma^2_z) \). Notice that a local IMSPE-optimal design is independent of \( \beta \) but depends on the values of the regression functions at the training data inputs, through \( F \), and at the point to be predicted, through \( f_0 \).

The parameters \( \Omega^* \) needed to calculate IMSPE* typically are not known in advance of the experiment. The simulation study in Section 5.4 examines the prediction accuracy of a range of local IMSPE-optimal designs to determine whether there is a choice of \( x^* \) that allows for accurate empirical predictions for a variety of test-bed surfaces. The choices of \( \Omega^* \) that are used to construct locally optimal designs for the simulation study of Section 5 are based on the results of Leatherman et al. (2016) for the simulator-only setting.
3.2. Maximin augmented nested Latin hypercube designs

A second design criterion that will be considered in the simulation study of Section 5 is the MmANLHD criterion. Recall that the projections of an \((n_s \times d_s)\) LHD onto every one of the \(d_s\) axes has one design value on each grid point \((0, 1/(n_s - 1), \ldots, 1)\). The MmANLHDs described below are constructed from nested LHDs (NLHDs). NLHDs have the property that the smaller (physical) design points must coincide with a subset of the control inputs for the larger (simulator) design. Because the NLHDs are marginally non-collapsing and are selected to be space-filling, the MmANLHDs constructed from these designs inherit these properties.

First, an NLHD that maximizes the minimum inter-point distance over pairs of rows is selected. Assuming \(n_p \leq n_s\), the full NLHD is used for the control variables in the simulator experiment and the smaller (nested) design is used for the physical design. Then, augmentation is performed by adding columns to the simulator design matrix for the calibration inputs. Formally, let \(L_{n_p, d_p, n_s, d_s + d_t}\) denote the set of all \(n_p \times (d_s + d_t)\) LHDs whose first \(d_s\) columns form the specified maximin NLHD with designs sizes \((n_p, d_p)\) and \((n_s, d_s)\). Any design \(X \in L_{n_p, d_p, n_s, d_s + d_t}\) that maximizes

\[
\min_{x_{11}, x_{22} \in X} \sum_{i=1}^{d_s + d_t} (x_{1,i} - x_{2,i})^2,
\]

is said to be a *maximin augmented nested Latin hypercube design*. The MmANLHDs used for comparison in the simulation study of Section 5.4 were constructed starting with the maximin NLHDs posted on the website of van Dam et al. (2013).

4. Design construction algorithms

4.1. A global/local algorithm for constructing local IMSPE-optimal designs

To find the combined designs \((X^p \in D^p_{n_p, d_p}, X^s \in D^s_{n_s, d_s + d_t, b})\) that minimize IMSPE\(^*\) in (10) for \(b = 10^{-3}\) at a specific set of parameter values \(\Omega^*\), this paper uses particle swarm optimization (PSO) to identify a design that serves as the starting point for a gradient-based quasi-Newton (QN) search for the best design. A detailed description of this heuristic approach and an illustrative example is presented in Leatherman et al. (2014a).

Briefly, PSO begins with a set of \(N_{\text{des}}\) starting combined designs \((X^p, X^s)\) spread over the design space \(D^p_{n_p, d_p}, D^s_{n_s, d_s + d_t, b}\). Each design is iterated \(N_{\text{its}}\) times. At each iteration, a design is updated to a new design that is “between” the global best design among all combined designs generated thus far and the best design among those having started at the same \((X^p, X^s)\). For the examples in this paper, the PSO parameter settings followed the recommendations of Kennedy and Eberhart (1995) and Yang (2010), and the PSO algorithm was run with \(N_{\text{des}} = 4(n_p d_p + n_s (d_s + d_t))\) starting designs and \(N_{\text{its}} = 2 N_{\text{des}}\) iterations. The best design constructed by PSO was used as the starting design for a single run of a QN algorithm (as implemented in the MATLAB (MATLAB, 2015) code *fmincon*.m) to produce the final (approximate) local IMSPE-optimal design.

An example of a local IMSPE-optimal combined 10-run \(\times\) 2-d physical design and 15-run \(\times\) 3-d simulator design w.r.t. \(\Omega^* = (\theta = 0.5, \rho_2 = 0.25 \times 1_1, \sigma_2^2/\sigma_2^2 = 0.1, \rho_3 = 0.25 \times 1_2, \sigma_3^2/\sigma_2^2 = 0.01)\) and the constant mean, \(f(x, t) = 1\), is listed in Table 1. The physical design and the 2-d projection of the simulator design onto \((x_1, x_2)\) space is shown in the left

\[
\begin{array}{cccccc}
\text{Physical design} & & & & & \\
\hline
x_{1}^p & x_{2}^p & x_{1}^s & x_{2}^s & t_1 \\
0.4682 & 0.7144 & 1.0000 & 0.4026 & 0.4954 \\
0.7852 & 0.3414 & 1.0000 & 1.0000 & 0.5004 \\
0.3766 & 0.0795 & 1.0000 & 1.0000 & 0.0000 \\
0.1381 & 0.2439 & 0.3070 & 0.5737 & 0.4987 \\
0.7092 & 0.9092 & 0.0000 & 1.0000 & 0.5001 \\
0.9286 & 0.0883 & 0.0000 & 1.0000 & 0.0000 \\
0.0480 & 0.5000 & 0.3881 & 1.0000 & 0.4997 \\
0.1570 & 0.8472 & 1.0000 & 1.0000 & 1.0000 \\
0.4502 & 0.3779 & 0.0000 & 0.0000 & 0.4954 \\
0.8898 & 0.6930 & 0.5411 & 0.0000 & 0.0000 \\
0.6703 & 0.0010 & 0.0000 & 0.4932 \\
0.6330 & 0.5643 & 0.0004 & 0.4962 \\
0.5411 & 0.0010 & 0.0004 & 0.0019 \\
0.0000 & 1.0000 & 0.0000 & 0.0000 \\
0.6702 & 0.0000 & 0.4932 \\
\end{array}
\]
Fig. 1. Projections of the local IMSPE-optimal combined design in Table 1. Left panel: projection of the simulator design onto \((x_1, x_2)\) space. Right panel: scatterplot of the 3-d simulator design, with 2-d physical design shown on plane \(t_1 = 0\). The physical design shown as ‘+’ symbols; the simulator design shown as open circles.

panel of Fig. 1; this projection is visually space-filling. The right panel of Fig. 1 shows the full 3-d simulator design which includes the calibration parameter \(t_1\) on the vertical axis. The right panel also shows the 2-d physical design in the plane \(t_1 = 0\). Note that the eleventh and fifteenth point of the simulator design in Table 1 have a Euclidean distance of only 0.001005 between them. This distance is very near the simulator design minimum distance restriction that is permitted by the \(\epsilon\)-ball with radius \(10^{-3}\). Other authors who have found this near-replication of simulator points are Crary (2002) and Crary and Stormann (2015).

4.2. A Smart Swap algorithm for constructing MmANLHDS

Each MmANLHD used in the simulation study of Section 5.4 was constructed by appending columns to the design matrix of a fixed maximin NLHD with \(d_t\) inputs, using a modified version of the Smart Swap algorithm described by Moon et al. (2011). (This modification also allows the algorithm to be used to add inputs to any fixed LHD having points on a grid.) Specifically, for this paper, \(d_t\) additional columns are required for the input settings of the calibration parameters in the \(n_s\)-run simulator design (with nested \(n_p\)-run physical design). For the given NLHD, it is required that the resulting augmented design is an MmANLHD, i.e., it should have a maximum value of the minimum inter-point Euclidean distance (11).

The Smart Swap algorithm of Moon et al. (2011) iteratively improves a candidate design as follows: at each iteration, one of the design points involved in the minimum inter-point distance is selected at random and, if it improves the design, a coordinate of this design point is swapped with the corresponding coordinate of another point randomly selected from the design. Additionally, this coordinate swap occurs with a probability specified by the user when the swap produces an equivalent value of (11); this probability was set to 0.05 for the designs constructed for this paper. For the MmANLHDs in this paper, the swap is only applied to the last \(d_t\) columns of the candidate design matrix, since the first \(d_s\) columns are fixed.

One other minor modification must be applied to the original Smart Swap algorithm when the design is to be constructed on a grid (rather than design points randomly selected within the “cells” that are formed by the grid). Since, in this situation, multiple pairs of design points are likely to have the same value of minimum inter-point distance, the algorithm is modified to identify and list all such pairs.

Table 2 lists an MmANLHD of size \(10 \times 2\) for the physical experiment and \(15 \times 3\) for the simulator experiment. The design was constructed using the modified Smart Swap algorithm to augment a maximin NLHD from the website created by van Dam et al. (2013). The left panel of Fig. 2 shows the physical design with the 2-d projection of the simulator design onto \((x_1, x_2)\) space; this panel visually demonstrates the space-filling property of the \(d_s\) control inputs for the physical and simulator designs, separately, and the coinciding property of the two designs in these dimensions. The right panel of Fig. 2 shows the 3-d simulator design with the physical design on the plane at \(t_1 = 0\).

5. Comparison of design prediction accuracy

This section examines the prediction accuracy of specific physical and simulator designs when used to predict for a test bed of surfaces. The EMSPE, defined below in Section 5.3 for the Bayesian predictor \(\hat{\zeta}_{fn}(x_0)\) in (7), will be used to compare the prediction accuracy of the designs described in Section 5.1 for the surfaces described in Section 5.2.
Table 2
An MmANLHD with 10 × 2 physical and 15 × 3 simulator component designs that was constructed using the modified Smart Swap algorithm.

<table>
<thead>
<tr>
<th>Physical design</th>
<th>Simulator design</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^p_1 )</td>
<td>( x^s_1 )</td>
</tr>
<tr>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.0714</td>
<td>0.0714</td>
</tr>
<tr>
<td>0.2143</td>
<td>0.2143</td>
</tr>
<tr>
<td>0.2857</td>
<td>0.2857</td>
</tr>
<tr>
<td>0.3571</td>
<td>0.3571</td>
</tr>
<tr>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5714</td>
<td>0.5714</td>
</tr>
<tr>
<td>0.7143</td>
<td>0.7143</td>
</tr>
<tr>
<td>0.8571</td>
<td>0.8571</td>
</tr>
<tr>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. An MmANLHD with 10 × 2 physical and 15 × 3 simulator component designs. Left panel: projection of the simulator design onto \((x_1, x_2)\) space. Right panel: scatterplot of the 3-d simulator design, with 2-d physical design shown on plane \(t_1 = 0\). The physical design shown as ‘+’ symbols; the simulator design shown as open circles.

5.1. Combined physical and simulator designs studied

The prediction accuracy of eight design sizes \((n_p, d_x, n_s, d_s + d_t)\) for each of fourteen \((X^p, X^s)\) designs are compared in Section 5.4 in terms of EMSPE. The designs are described in the following paragraphs and are summarized in Table 3.

Four of the designs were constructed by minimizing IMSPE* \((X^p, X^s \mid \Omega^*)\) in (10) for a specific \(\Omega^* = (\theta, \rho_Z, \sigma^2_Z / \sigma^2, \rho_\delta, \sigma^2_\delta / \sigma^2_Z)\). In the simpler problem of predicting simulator output based on a set of simulator training data, Leatherman et al. (2016) found that local IMSPE-optimal designs constructed using “small” correlation parameter values yielded smaller empirical prediction errors than designs based on “larger” correlation parameter values. Guided by these observations, the correlation vectors \(\rho_Z\) and \(\rho_\delta\) were chosen to have a small common correlation value \(\rho\) for the local IMSPE-optimal designs used in the study of Section 5.4. Four \((\rho_Z, \rho_\delta)\) vector combinations were used in total: \((0.25 \times 1_{d_x+d_t}, 0.25 \times 1_{d_s})\), \((0.25 \times 1_{d_x+d_t}, 0.5 \times 1_{d_s})\), \((0.5 \times 1_{d_x+d_t}, 0.25 \times 1_{d_s})\), and \((0.5 \times 1_{d_x+d_t}, 0.5 \times 1_{d_s})\). The calibration parameter vector \(\theta\) is an element of \([0, 1]^d\) and, as a “naïve” selection, was set to be \(\theta = 0.5 \times 1_{d_s}\). Additionally, the variance ratios were selected to be \(\sigma^2_\delta / \sigma^2_Z = 0.1\) and \(\sigma^2_\epsilon / \sigma^2_Z = 0.01\), and the constant mean, \(f(x, t) = 1\), was used for the GP in (1).

For each of the four \(\Omega^*\) specified in the previous paragraph, the local IMSPE-optimal combined design \((X^p, X^s)\) was constructed using the PSO plus QN optimization algorithm described in Section 4.1. In the sections that follow, these four designs are denoted \(D^S_{0.25,0.25}, D^S_{0.25,0.5}, D^S_{0.5,0.25},\) and \(D^S_{0.5,0.5}\), where the superscript “PS” denotes that the optimality criterion
is applied to the combined (physical and simulator) design. The subscripts denote the common values of the simulator and the discrepancy correlation parameters, respectively (see also Table 3).

A second class of designs fixes the physical design $X^p$ and constructs a local IMSPE-optimal simulator design $X^s$ to minimize IMSPE*. Four of the designs included in this paper are such local IMSPE-optimal simulator designs used in conjunction with a given I-optimal physical design. The parameter values used to optimize IMSPE* are identical to those used to construct the combined local IMSPE-optimal designs $D^s$. The fixed I-optimal designs were constructed with JMP® (JMP, 1989–2007) assuming a cubic mean model that also included quadratic interaction terms, as this was the largest polynomial model that could be fit for all $n_x$ and $d_x$ studied. In the sections that follow, the local IMSPE-optimal simulator designs combined with the fixed I-optimal physical design are denoted $D_{0.25,0.25}^s$, $D_{0.25,0.5}^s$, $D_{0.5,0.25}^s$, and $D_{0.5,0.5}^s$, where “S” in the superscript indicates that only the simulator design is IMSPE optimal. As for the case of combined designs, the subscripts state the common value of the $\rho_2$ and $\rho_3$ parameters, respectively (see also Table 3).

Analogously, a third class of designs fixes the simulator design $X^s$ and constructs a local IMSPE-optimal physical design $X^p$ to minimize IMSPE*. Again, four designs were included in the EEMSPE studies. These ($X^p$, $X^s$) were determined by combining a given maximin LHD (MmLHD) for $X^s$ with a local IMSPE-optimal $X^p$. The parameter values used to optimize the physical design are identical to those for obtaining $D_{0.25,0.25}^s$. The MmLHDs were obtained from van Dam et al. (2013). In the sections that follow, these designs are denoted $D_{0.25,0.25}^p$, $D_{0.25,0.5}^p$, $D_{0.5,0.25}^p$, and $D_{0.5,0.5}^p$, where the superscript “P” indicates that only the physical design is IMSPE optimal. Again, the subscripts state the common value of the $\rho_2$ and $\rho_3$ parameters, respectively (see also Table 3).

In the two single design optimization scenarios above, the PSO plus QN algorithm was used to optimize the simulator design $X^s$ with a fixed physical design $X^p$, and vice versa. The PSO algorithm was initiated with $N_{\text{des}}$ starting designs taken from the appropriate design space, where $N_{\text{des}} = 4n_x(d_x + d_t)$ when optimizing $X^s$ with a fixed $X^p$, and $N_{\text{des}} = 4n_p d_x$ when optimizing $X^p$ with a fixed $X^s$. In both cases, the PSO employed $N_{\text{its}} = 2N_{\text{des}}$ iterations.

The final designs studied in the following sections are space-filling. The first design is an “off-the-shelf” design that combines an MmLHD for the simulator runs and an I-optimal design for the physical experiment observations. These MmLHD and I-optimal designs are the same designs used in the fixed simulator and fixed physical design settings, respectively, from above. This combined design is denoted I-opt + MmLHD in Table 3. The second design is an MmANLHD which is an intuitively more sophisticated version of the off-the-shelf design; the MmANLHD uses common inputs for the physical experiment and the control portion of the simulator inputs. This design was computed using the Smart Swap algorithm in Section 4.2 by augmenting the $(n_3 \times \text{grid})$ maximin NLHD from van Dam et al. (2013) having $n_p$ and $n_i$ points in $d_x$ dimensions. This design is denoted ANLHD in Table 3.

The designs described above were constructed using Linux compute machines having Dual Eight Core Xeon 2.7 E5–2680 processors with 384 GB RAM. The exception was the $(n_x, d_x, n_p, d_t, d_t + d_t) = (10, 2, 15, 3)$ combined local IMSPE-optimal design, which was constructed using a Linux compute machine with Dual Quad Core Xeon 2.66 E5430 processors with 32 GB RAM. For local IMSPE-optimal designs, Table 4 lists the average hours of computation time, over the four correlation

<table>
<thead>
<tr>
<th>Design type</th>
<th>Design label</th>
<th>Common correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>Simulator</td>
<td>$D_{0.25,0.25}^p$</td>
</tr>
<tr>
<td>Combined</td>
<td>IMSPE-opt</td>
<td>$D_{0.25,0.25}^p$</td>
</tr>
<tr>
<td>I-opt</td>
<td>IMSPE-opt</td>
<td>$D_{0.25,0.25}^s$</td>
</tr>
<tr>
<td>I-opt</td>
<td>MmLHD</td>
<td>$D_{0.25,0.25}^p$</td>
</tr>
<tr>
<td>I-opt + MmLHD</td>
<td>ANLHD</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3
Labels used to denote the designs compared in the simulation study of Section 5.4. The superscripts denote the portion of the combined design that was optimized using the local minimum IMSPE criterion. The subscripts denote the common correlation parameter values $\rho$ that were used in the IMSPE optimization, where $\rho_2 = \rho \times 1_{d_x+d_t}$ and $\rho_3 = \rho \times 1_{d_t}$. The subscripts are listed in the order $\rho_2$, $\rho_3$. |
Table 4
The average computation time (in hours), over the four correlation scenarios, needed to construct each design type and size. The computation time listed for ANLHD is for a single MmANLHD, where the starting maximin NLHD was given.

<table>
<thead>
<tr>
<th>(n_s, d_s, n_p, d_p + d_t)</th>
<th>Design type</th>
<th>( D^{52} )</th>
<th>( D^{50} )</th>
<th>( D^{47} )</th>
<th>ANLHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 2, 15, 3)</td>
<td>25</td>
<td>1.62</td>
<td>0.03</td>
<td>0.12</td>
<td>0.0009</td>
</tr>
<tr>
<td>(20, 2, 30, 5)</td>
<td>25</td>
<td>3.54</td>
<td>0.11</td>
<td>0.53</td>
<td>0.0057</td>
</tr>
<tr>
<td>(20, 4, 25, 5)</td>
<td>25</td>
<td>6.78</td>
<td>0.45</td>
<td>1.53</td>
<td>0.0032</td>
</tr>
<tr>
<td>(20, 4, 30, 6)</td>
<td>25</td>
<td>9.83</td>
<td>0.56</td>
<td>2.98</td>
<td>0.0252</td>
</tr>
<tr>
<td>(30, 3, 50, 5)</td>
<td>25</td>
<td>19.30</td>
<td>1.57</td>
<td>10.34</td>
<td>0.1069</td>
</tr>
<tr>
<td>(40, 4, 50, 5)</td>
<td>25</td>
<td>40.24</td>
<td>4.37</td>
<td>13.00</td>
<td>0.0253</td>
</tr>
<tr>
<td>(40, 4, 60, 6)</td>
<td>25</td>
<td>52.25</td>
<td>4.82</td>
<td>40.10</td>
<td>0.1934</td>
</tr>
</tbody>
</table>

4 Constructed using a different type of computer machine than for the other design types and sizes (see text).

Fig. 3. A surface \( \xi_{\text{test}}(\mathbf{w}) \) when \( d_s = 2, d_t = 1, \) and \( \theta = 0.25 \): (a) a 2-d simulator surface \( S_{0.25}^{\text{Krig}} \) evaluated at \( \theta = 0.25 \), (b) a Kriging discrepancy surface \( \beta_{0.25}^{\text{Krig}} \), (c) the mean surface \( \xi_{\text{test}}(\mathbf{w}) \) obtained by summing (a) and (b).

5.2. The test bed of physical and simulator surfaces

To compare the prediction accuracy of the designs of Section 5.1, a test bed of non-linear physical and simulator surfaces was created. Each mean surface of the physical observations \( \xi(\mathbf{x}^p) \) is constructed as the summation of a corresponding simulator response surface and a discrepancy (bias) response surface. When physical “observations” are made, an additional observation error is added to the mean physical response.

5.2.1. Simulator surfaces

The test beds of simulator surfaces used in the simulation study of Section 5.4 can be categorized in three groups. The first group of simulator surfaces uses the Kriging surfaces of Trosset (1999). These surfaces have the form

\[
y_{\text{test}}^i(\mathbf{w}) = \hat{\beta}_Z + \mathbf{r}_Z(\mathbf{w})^T \mathbf{R}_Z^{-1} \left( \mathbf{Y}^{500} - \mathbf{1}_{500} \hat{\beta}_Z \right), \quad \mathbf{w} \in [0, 1]^{d_s + d_t}.
\]

where \( \mathbf{Y}^{500} \) is a 500 × 1 vector drawn from a stationary GP based on inputs that form a 500 × (\( d_s + d_t \)) approximate MmLHD, denoted by \( \mathbf{L}_{d_s + d_t} \).

The sampled GP has mean \( \hat{\beta}_Z = 100 \), variance \( \sigma_Z^2 = 10 \), and Gaussian correlation function (2), where \( \rho_{ij} \) is either 0.25 × \( \mathbf{1}_{d_s + d_t} \) or 0.5 × \( \mathbf{1}_{d_s + d_t} \). For numerical stability, a nugget \( 10^{-6} \) was added to the diagonal of \( \sigma_Z^2 \mathbf{R}_Z \). In (12), \( \hat{\beta}_Z = \left( \mathbf{1}_{500} \mathbf{R}_Z^{-1} \mathbf{1}_{500} \right)^{-1} \mathbf{1}_{500} \mathbf{R}_Z^{-1} \mathbf{Y}^{500} \), \( \mathbf{r}_Z(\mathbf{w}) \) is the 500 × 1 vector of correlations having the \( i \)th component \( R_Z(x_i, \mathbf{w}) \) for \( x_i^T \in \mathbf{L}_{d_s + d_t} \), and \( \mathbf{R}_Z \) is the 500 × 500 matrix of correlations having \( (i, j) \)th element \( R_Z(x_i, x_j) \) for \( x_i^T, x_j^T \in \mathbf{L}_{d_s + d_t} \). For each \( \rho_{ij} \), 30 surfaces \( y_{\text{test}}^i(\mathbf{w}) \) were constructed. The collections of Kriging simulator surfaces are denoted \( S_{\text{Krig}}^{0.25} \) and \( S_{\text{Krig}}^{0.5} \), where the subscript specifies the common correlation value used to construct the surface. A representative simulator surface from \( S_{\text{Krig}}^{0.25} \) is plotted in Panel (a) of Fig. 3.
5.2.2. Discrepancy (bias) surfaces

Here surface from the sampled GP has mean \( \theta \) of 30 surfaces were formed, each by taking i.i.d. Uniform stationary activity near the center of \([0, 0, 0, 0, 0] \) and \([0, 0, 0, 0, 0, 0] \). No bias is required, \( B_{\text{Krig}} \) of Figs. 3–5. When no bias is required, \( \delta_{\text{test}}(w) \) is set to zero for \( w \in [0, 1]^{d_{x}} \).

The second group of simulator surfaces uses the exponential model of Sobol’ and Levitan (1999)

\[
y_{\text{test}}(w) = \exp(b^T w) - I_{d_{x} + d_{t}}, \quad w \in [0, 1]^{d_{x} + d_{t}},
\]

where \( I_{d_{x} + d_{t}} = \prod_{i=1}^{d_{x} + d_{t}} b_{i}^{-1} \), and \( b_{i} \) is the ith element of \( b \). The \( b_{i} \) parameters were chosen based on results from Loeppky et al. (unpublished). Thirty sets of \( b \) (and surfaces (13)) were constructed around two central \( b \) values. The central value of the first 30 \( b \) draws produced surfaces with inputs having equal effects, whose main effects contributed 50% of the overall variance, and whose calibration inputs contributed 75% of the main effect variance. The central value of the second set of 30 \( b \) draws produced surfaces whose inputs had unequal effects, whose main effects contributed 95% of the overall variance, and whose calibration inputs contributed 75% of the main effect variance. The simulator exponential surface families are denoted \( C_{E}^{SL} \) and \( C_{U}^{SL} \), where the subscript specifies whether the inputs have “E”qual or “U”nequal effects. A representative simulator test-bed surface from \( S_{U}^{SL} \) is plotted in Panel (a) of Fig. 4.

The third group of simulator surfaces uses a modification of the non-stationary function described by Xiong et al. (2007) (see also Ba and Joseph, 2012). Two test beds of non-stationary test surfaces are formed from

\[
y_{\text{test}}(w) = 10^{(d_{x} + d_{t})/2} \prod_{i=1}^{d_{x} + d_{t}} \left\{ \sin(a_{i}(u_{1} - b_{i})^{4}) \cos(2(u_{1} - b_{i})) + \frac{w_{i} - b_{i}}{2} \right\}, \quad w \in [0, 1]^{d_{x} + d_{t}}
\]

where \( a_{i} \) and \( b_{i} \) are varied to form each surface. Two families of test beds are formed. In the first test bed, the base family is defined by (14) and has non-stationary activity that occurs near the edges of the input domain \([0, 1]^{d_{x} + d_{t}} \); in the second test bed, a change of variable is made in the base function to \( y_{\text{test}}(w_{1} - 0.5, \ldots, |w_{d_{x} + d_{t}} - 0.5|) \) which results in non-stationary activity near the center of \([0, 1]^{d_{x} + d_{t}} \). Using these two formulations of (14), two non-stationary test-bed families of 30 surfaces were formed, each by taking i.i.d. Uniform(20, 35) draws \( a_{1}, a_{2}, \ldots, a_{d_{x} + d_{t}} \) and i.i.d. Uniform(0.5, 0.9) draws \( b_{1}, b_{2}, \ldots, b_{d_{x} + d_{t}} \). The families of such surfaces are denoted \( C_{\text{MIX}}^{\text{edge}} \) and \( S_{\text{MIX}}^{\text{mid}} \), respectively. A representative simulator test-bed surface from \( S_{\text{MIX}}^{\text{mid}} \) is shown in panel (a) of Fig. 5.

5.2.2. Discrepancy (bias) surfaces

All discrepancy (bias) surfaces are of the form

\[
\delta_{\text{test}}(w) = \hat{\beta}_{3} + r_{3}(w)^{T} R_{3}^{-1} \left( V^{500} - 1_{500} \hat{\beta}_{3} \right), \quad w \in [0, 1]^{d_{x}}.
\]

Here \( V^{500} \) is a 500 \( \times \) 1 vector drawn from a stationary GP based on a 500 \( \times \) 4 approximate MmLHD, denoted by \( L_{d_{x}} \). The sampled GP has mean \( \beta_{3} = 0 \), variance \( \sigma^{2} = 1 \), and Gaussian correlation function (3), where \( \rho_{g} \) is either 0.5 \( \times \) 1, or 0.75 \( \times \) 1. For numerical stability, a nugget 10\(^{-6}\) was added to the diagonal of the covariance matrix, \( \sigma^{2} R_{g} \). In (15), \( \hat{\beta}_{3} = (1_{500} R_{3}^{-1} 1_{500})^{-1} 1_{500} R_{3}^{-1} V^{500}, r_{3}(w) \) is the 500 \( \times \) 1 vector of correlations with ith component \( R_{3}(\mathbf{x}_{i}, \mathbf{w}) \) for \( \mathbf{x}_{i}^{T} \in L_{d_{x}} \), and \( R_{j} \) is the 500 \( \times \) 500 matrix of correlations having (i, j)th element \( \hat{R}_{j}(\mathbf{x}_{i}, \mathbf{x}_{j}) \) for \( \mathbf{x}_{i}^{T}, \mathbf{x}_{j}^{T} \in L_{d_{x}} \). For each \( \rho_{g} \), 30 surfaces \( \delta_{\text{test}}(w) \) were constructed. The collections of bias surfaces are denoted \( B_{0.5}^{\text{Krig}} \) and \( C_{0.75}^{\text{Krig}} \), where the subscript specifies the common correlation value used to construct the surface. Three representative bias surfaces from \( B_{0.5}^{\text{Krig}} \) can be seen in panel (b) of Figs. 3–5. When no bias is required, \( \delta_{\text{test}}(w) \) is set to zero for \( w \in [0, 1]^{d_{x}} \).
Fig. 5. A surface \( \zeta_{\text{test}}(w) \) when \( d_x = 2, d_t = 1, \) and \( \theta = 0.25 \): (a) a 2-d simulator surface \( S_{0.25}^{\text{MOB}} \) evaluated at \( \theta = 0.25 \), (b) a Kriging discrepancy surface \( B_{0.5}^{\text{Krig}} \), (c) the mean surface \( \zeta_{\text{test}}(w) \) obtained by summing (a) and (b).

Table 5

<table>
<thead>
<tr>
<th>( y_{\text{test}}^d(\mathbf{x}_0, \theta) )</th>
<th>( \delta_{\text{test}}(\mathbf{x}_0) )</th>
<th>( \rho_{\text{Krig}}^{\text{MOB}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{0.25}^{\text{Krig}} )</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>( S_{0.5}^{\text{Krig}} )</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>( S_{0.25}^L )</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>( S_{0.5}^L )</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>( S_{0.25}^{\text{MOB}} )</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>( S_{0.5}^{\text{MOB}} )</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

5.2.3. Mean of the physical system

For the simulation study of Section 5, the physical surface means are constructed using \( \theta = 0.25 \times 1_{d_1} \) for the true value of the calibration parameter. Notice that this true parameter value is different from the “naïve” selection of \( \theta = 0.5 \times 1_{d_1} \) that was used to construct the local IMSPE-optimal designs described in Section 5.1. Alternative values of the calibration parameter used for test-bed generation are investigated in Section 6.

In more detail, each mean physical response surface for Section 5 is the summation of a simulator surface (12), (13) or (14) and a discrepancy surface (15); that is,

\[
\zeta_{\text{test}}(w, \mathbf{x}_0) = y_{\text{test}}^d(w, \mathbf{x}_0) + \delta_{\text{test}}(w, \mathbf{x}_0), \quad w \in [0, 1]^{d x},
\]

or, for no bias, \( \delta_{\text{test}}(w) = 0 \). Each “observable” physical response is the summation of a mean physical response (16) plus i.i.d. observation error:

\[
y_{\text{test}}^d(w, 0.25 \times 1_{d_1}) = \zeta_{\text{test}}(w) + \epsilon, \quad w \in [0, 1]^{d x},
\]

where \( \epsilon \) is Normal(0, \( \sigma^2_{\epsilon} = 1 \)). The 18 settings used to construct the physical response surfaces \( \zeta_{\text{test}}(w) \) are listed in Table 5. For each of these parameter settings, 30 physical surfaces were drawn along with their corresponding simulator surfaces as described above. A different set of 18 \( \times 30 \) surfaces was drawn for each of the (physical, simulator) dimensions \((d_x, d_t + d_1) \in \{(2, 3), (3, 5), (4, 5), (4, 6)\}\).

Three representative \( \zeta_{\text{test}}(w, \mathbf{x}_0) \) in (16) are shown in panels (c) of Figs. 3–5. In Fig. 3, panels (a) and (b) plot a representative simulator surface \( S_{0.25}^{\text{Krig}} \) and discrepancy surface \( B_{0.5}^{\text{Krig}} \) respectively, while \( \zeta_{\text{test}}(w, \mathbf{x}_0) \) in panel (c) is the sum of panels (a) and (b). Similarly panels (a) and (b) of Fig. 4, show representative simulator and discrepancy surfaces \( S_{0.5}^U \) and \( B_{0.5}^{\text{Krig}} \), respectively, that were summed to construct panel (c) of the same figure. Lastly, panels (a) and (b) of Fig. 5, show representative simulator and discrepancy surface \( S_{0.5}^{\text{MOB}} \) and \( B_{0.5}^{\text{Krig}} \), respectively, that were summed to construct panel (c) of the same figure.

5.3. Measuring design prediction accuracy

In order to quantify the prediction accuracy of a combined physical and simulator design \((X^p, X^s)\) across the physical system design space \([0, 1]^{d_x}\), the EMSPE was calculated using an equally-spaced and computationally-feasible grid of \( g \).
points taken from $[0, 1]^d$. Given the mean of a physical test-bed surface $\zeta_{\text{test}}(\mathbf{x}) \equiv \zeta_{\text{test}}(\mathbf{x}, \mathbf{\theta})$ the EMSPE is defined by

$$\text{EMSPE}(\mathbf{X}^p) = \frac{1}{g} \sum_{i=1}^{g} \left( \zeta_{\text{test}}(\mathbf{x}_i) - \zeta_{\text{test}}(\mathbf{x}_i) \right)^2,$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_g$ are the $g$ grid points. In the examples below, $g = 50^2$ for $d_x = 2$, $g = 20^3$ for $d_x = 3$, and $g = 10^4$ for $d_x = 4$. For the comparisons made in Sections 5.4 and 6, the predictor $\widehat{\zeta}_p(\cdot)$ in (7) was calculated using the Markov Chain Monte Carlo posterior distribution $p(\mathbf{\beta}, \mathbf{\Omega} | \mathbf{Y})$ that is constructed using the methodology and GPM/SA software described in Higdon et al. (2008) and Gattiker (2008). Alternatively, one could make predictions with the BLUP $\widehat{\zeta}_{blup}(\cdot)$ in (5) with REML estimates of the model parameters $\mathbf{\Omega}$. No matter the predictor used, designs with small EMSPE values are desirable as they allow for more accurate predictions on average across the control inputs.

5.4. Design comparisons and recommendations

Thirty representative surfaces were drawn from each of the 18 test beds listed in Table 5. Training data were collected from each of the $30 \times 14$ surfaces using each of the 14 designs in Table 3 and for each of eight $(n_p, d_x, n_s, d_s + d_f)$ design sizes. For each of the $30 \times 14 \times 8$ sets of training data, predictions were made using $\widehat{\zeta}_p(\cdot)$ in Eq. (7) at a grid of inputs and the EMSPE in Eq. (18) was calculated. The EMSPE values were standardized within each test bed and design size combination because the surfaces vary in complexity across test beds and the prediction accuracy differs depending on the amount of training data available as specified by the $(n_p, d_x, n_s, d_s + d_f)$ design size. For each test bed and design size combination, the $420 = 30 \times 14$ EMSPE values were standardized by subtracting the mean (taken over the 30 surface realizations and 14 designs) and dividing by the standard deviation. Figs. 6–9 are four comparative plots of the 90th percentiles of the 30 standardized EMSPE values for the $18 \times 14$ test bed by design combinations, for the four design sizes

$$(n_p, d_x, n_s, d_s + d_f) \in \{(10, 2, 15, 3), (20, 4, 25, 5), (30, 3, 50, 5), (40, 4, 50, 5)\}.$$  

Corresponding plots of the 90th percentiles of standardized EMSPE values for the additional 4 design sizes $(15, 3, 25, 5), (20, 2, 30, 3), (20, 4, 30, 6),$ and $(40, 4, 60, 6)$ included in the Supplementary material (see Appendix A), as are tables of the 90th percentiles of the 30 non-standardized EMSPE values for all $8 \times 18 \times 14$ cases.

Starting with the tables in the Supplementary material (see Appendix A), the effect of increasing the sample size from 5 runs per dimension to 10 runs per dimension can be quantified by comparing the entries in the pairs of tables with common $(d_x, d_f)$; there are four such pairs of tables. The 90th percentile of the non-standardized EMSPE values is reduced in 91% of
Fig. 7. \((n_p, d_x, n_s, d_x + d_t) = (20, 4, 25, 5)\): A grayscale heatmap of the 90th percentile of the standardized EMSPE values for the 14 designs listed in Table 3 and the 18 test-bed surface types listed in Table 5. Test bed surfaces 1–3, 4–6, 7–9, 10–12, 13–15 and 16–18 use \(S_{Krig_{0.25}}, S_{Krig_{0.5}}, S_{SL_E}, S_{SL_U}, S_{MXB_{edge}},\) and \(S_{MXB_{mid}},\) respectively, as \(y^*(x, t)\). Within each grouping of three simulator surfaces, the \(\zeta_{test}(x)\) to be estimated is the sum of \(y_{test}^*(x, 0.25)\) and \(\delta_{test}(x)\) which is: \(\equiv 0, B_{0.5}, B_{0.75}\), in order.

Fig. 8. \((n_p, d_x, n_s, d_x + d_t) = (30, 3, 50, 5)\): A grayscale heatmap of the 90th percentile of the standardized EMSPE values for the 14 designs listed in Table 3 and the 18 test-bed surface types listed in Table 5. Test bed surfaces 1–3, 4–6, 7–9, 10–12, 13–15 and 16–18 use \(S_{Krig_{0.25}}, S_{Krig_{0.5}}, S_{SL_E}, S_{SL_U}, S_{MXB_{edge}},\) and \(S_{MXB_{mid}},\) respectively, as \(y^*(x, t)\). Within each grouping of three simulator surfaces, the \(\zeta_{test}(x)\) to be estimated is the sum of \(y_{test}^*(x, 0.25 \times 1_2)\) and \(\delta_{test}(x)\) which is: \(\equiv 0, B_{0.5}, B_{0.75}\), in order.

The 1008 (design \(\times\) test-bed \(\times\) design size pair) combinations when using 10 runs per input compared with 5 runs per input. The 6 test-beds formed using the Sobol’–Levitans simulator in Eq. (13) were most likely to have the largest reduction, with one design \(\times\) test bed combination having a 94% reduction in the 90th percentile of the non-standardized EMSPE values.
The 6 test-beds using the non-stationary modified Xiong/Ba simulator surface in Eq. (14) occasionally predicted worse when using 10 runs per input compared with 5 runs per input; with the largest observed increase in the 90th percentile of the non-standardized EMSPE values being 250%. Similar increases were observed for the simulator-only prediction of modified Xiong/Ba surfaces in Leatherman et al. (2016).

Figs. 6–9 and the corresponding figures in the Supplementary material tell a complicated story about the designs’ prediction accuracy (see Appendix A). First, there is no single design that predicts better than all competing designs for all test-beds and all design sizes, i.e., no (design) row is uniformly “lighter” than all other rows for all of Figs. 6–9. However, some designs are clearly inferior to others for sufficiently many design sizes, such that one should use one of the better alternatives. First, the I-opt + MmLHD should be avoided because it is often inferior to ANLHD. Second, the four designs that add simulator runs to the I-opt physical experiment design, the $D_{opt}$ designs, are each inferior to the corresponding $P_{opt}$ designs. Third, the $D_{0.5,-}$ designs are inferior to the $D_{0.25,-}$ designs. Fourth, the $D_{opt}$ designs are inferior to the $D_{0.25,-}$ designs.

This leaves five designs that are candidates to provide the most accurate predictions: $D_{PS}^{0.25,0.25}$, $D_{PS}^{0.25,0.5}$, $D_{PS}^{0.25,0.25}$, $D_{0.25,0.25}$, and ANLHD. ANLHD is particularly effective for predicting non-stationary surfaces, such as the Xiong/Ba surfaces (#13–#18) with the center displaying different behavior than the edge or vice-versa. The same is true for the Sobol’–Levitan surfaces (#7–#12) which have a large spike in one corner of the input space. In both the Xiong/Ba and the Sobol’–Levitan surfaces, not having a design point near the non-stationary activity leads to large prediction errors.

Among the local IMSPE-optimal designs $D_{PS}^{0.25,-}$ and $D_{PS}^{0.5,-}$, those with $\rho_z = 0.5$ produce more accurate predictions than designs with $\rho_z = 0.25$. Thus the designs producing the most accurate predictions appear robust to the choice of $\rho_z$ but depend heavily on the choice of $\rho_x$. Designs $D_{0.25,-}$ and $D_{0.25,-}$ predict well for the Kriging surfaces (#1–#6), which is not surprising as these are the surfaces for which the designs were constructed. Additionally, the locally optimal design $D_{0.25,0.5}^{0.25}$ has many points near the edges of the design space and therefore predicts well for the test-beds with surfaces that have non-stationary activity near the edges of the input space (#7–#15). Similarly, designs with inputs near the middle of the design space predict well for the modified Xiong/Ba surfaces that have non-stationary activity near the middle of the input space (#16–#18); designs ANLHD, $D_{0.25,0.25}$, and $D_{0.25,0.5}$ all predict well for these surfaces.

Recalling that $D_{opt}$ designs require roughly 10% of the computing time to construct compared to the corresponding $D_{PS}$ designs, combining a MmLHD for the simulator experiment with a design that optimizes the IMSPE criterion for the physical experiment is an attractive design choice. While the prediction criterion for this paper is different than that of Ranjan et al. (2011), our conclusions agree qualitatively with their assessment that designs that are locally IMSPE-optimal in the physical element ($D_{opt}$) predict best, followed by the designs that are combined locally IMSPE-optimal ($D_{PS}^{0.25,-}$), and lastly followed by designs that are locally IMSPE-optimal in the simulator element ($D_{PS}^{0.5,-}$).
6. The dependence of prediction accuracy on the assumed \( \theta \)

All local IMSPE-optimal designs used for the simulation study in Section 5 were constructed with \( \theta = 0.5 \times 1_d \), \( \sigma_f^2/\sigma_{\eta}^2 = 0.1 \), and \( \sigma_f^2/\sigma_{\eta}^2 = 0.01 \). The Kriging test-bed surfaces in the simulation study of Section 5 were used to assess prediction accuracy when the test-bed correlation parameters \( \rho_z \) and \( \rho_d \) are different than their assumed values for design construction. The current section focuses on prediction accuracy in test beds formed from stationary GP draws with \( \theta \neq 0.5 \times 1_d \). Each of the \( D_{25.5}^{PS} \), \( D_{25.5}^{P} \), and \( D_{25.5}^{S} \) designs was used to predict 30 test surfaces \( \xi_{test}(x) = y_{test}(x, \theta) + \delta_{test}(x) \) for a given \( \theta \in \{0.125 \times 1_d, 0.25 \times 1_d, \ldots, 0.875 \times 1_d\} \), where \( y_{test}(x, \theta) \) is a draw from \( \xi_{test}^{Krig} \) and \( \delta_{test}(x) \) is a draw from \( b_{0.5}^{Krig} \).

For the eight design sizes listed in and directly below Eq. (19), the EMSPE in (18) was calculated; because all test surfaces are based on the same stationary GP model, the EMSPE was not standardized. Boxplots of the EMSPE values for the 30 surface realizations are shown in Figs. 10 and 11 for design sizes (10, 2, 15, 3) and (40, 4, 50, 5), respectively, and in the Supplementary material for the remaining six design sizes (see Appendix A). Each individual boxplot corresponds to a test bed of surfaces constructed for a specific value of \( \theta \).

Most panels in Figs. 10 and 11 (and in the related figures in the Supplementary material, Appendix A) provide intuitive results. The EMSPE values are smallest for test-bed surfaces constructed using \( \theta = 0.5 \times 1_d \), the calibration parameter used to construct the design. For many panels (across design types and sizes), the EMSPE increases as the absolute value of the...
distance between $\theta = 0.5 \times 1_d$ used in the design construction and the $\theta$ used to construct the test bed increases. Also the range of the EMSPEs appears to increase as the absolute distance between the design $\theta$ and surface $\theta$ increases. Focusing on Fig. 11, where ($n_p$, $d_x$, $n_s$, $d_x + d_1$) = (40, 4, 50, 5), the middle panel shows that $D^{P}_{25.0.5.}$ is robust to the calibration parameter used to construct the stationary GP. However, the EMSPE values for $D^{P}_{25.0.5.}$ never reach the best EMSPE values achieved by $D^{S}_{25.0.5.}$ when the surface is constructed using $\theta = 0.5$. On the other hand, the EMSPE values for $D^{P}_{25.0.5.}$ also do not reach the worst EMSPE values achieved by $D^{S}_{25.0.5.}$ when the surface is constructed using small or large $\theta$.

7. Summary

This paper compares a variety of criteria to select the initial design of a physical system experiment combined with that of a deterministic simulator of the physical system when the goal is accurate prediction of the mean of the physical system. Design criteria and construction algorithms are described for local IMSPE-optimal designs and maximin augmented nested LHMs (MmANLHMs). A local minimum IMSPE design for one of the experiments, simulator or physical system, is constructed when the other experiment uses an “off-the-shelf” design, either an MmLHD for the simulator experiment or an $l$-optimal design for the physical experiment. These designs are compared with the frequently-used combination of at-l-optimal design with an MmLHD. For a large test bed of stationary and non-stationary surfaces and for each design studied, a Bayesian calibrated predictor is used to estimate the mean of the physical system. EMSPE values are calculated for each test surface at a grid of inputs. This simulation study shows that no single design performs better than all competing designs for all test-beds and all design sizes. However certain designs should be avoided, general recommendations can be made, and particular designs can be recommended for specific surface types. The $l$-optimal physical design + MmLHD for the simulator is inferior to the MmANLHD. The $l$-optimal physical design + local IMSPE-optimal simulator designs are inferior to both the local IMSPE-optimal combined designs and to the MmLHD for the simulator + local IMSPE-optimal physical system experimental designs. Local IMSPE-optimal combined designs or local IMSPE-optimal physical system experimental designs that use $\rho_z = 0.5$ are inferior to those that use $\rho_z = 0.25$.

The five recommended designs, ANLHD, $D^{PS}_{0.25.0.25.}$, $D^{PS}_{25.0.25.0.5.}$, $D^{P}_{25.0.25.0.5.}$, and $D^{P}_{0.25.0.5.}$, have different strengths. ANLHD is least expensive to compute, and $D^{P}_{25.0.25.}$. is the most expensive to compute. Among the local IMSPE-optimal designs, $D^{P}_{0.25.}$—requires the least computational effort. For stationary GP draws, $D^{P}_{0.25.}$—and $D^{P}_{0.25.}$—predict well, with $D^{P}_{0.25.}$—often being superior to the correlation comparable $D^{PS}_{0.25.}$. For non-stationary surfaces, designs with inputs near the non-stationary activity predict well. For example, $D^{PS}_{0.25.0.5.}$ and ANLHD have design points near the spike in the Sobol’–Levitus simulator surface and consistently predict well for this surface. ANLHD predicts well for the non-stationary modified Xiong/Ba simulator surfaces that have activity near the edges of the input space. Finally, many designs, including ANLHD, $D^{P}_{0.25.0.25.}$, and $D^{P}_{0.25.0.5.}$, have points near the middle of the design space and thus predict well for the non-stationary simulator surfaces with activity near the middle of the input space.

A second simulation study shows that local IMSPE-optimal designs perform best for draws from stationary GPs having the same $\theta$ used to construct the design. These studies show both an increase in the median EMSPE and the range of EMSPE values as the physical surface draws have $\theta$ values that are further from the $\theta$ used to construct the design.

Acknowledgments

The authors wish to thank the reviewers and the Editor for suggestions which led to the improvement of this paper. The authors also wish to thank Bradley Jones for helpful discussions. This research was sponsored, in part, by an allocation of computing time from the Ohio Supercomputer Center, and by the National Science Foundation under Agreements DMS-0806134 and DMS-1310294 (awarded to The Ohio State University). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Supplementary material

The designs and construction code discussed in this manuscript can be found on the first author’s website http://stat.wvu.edu/~erl/CombinedDesigns/. Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2016.07.013.

References

Optimal experimental design on the loading frequency for a probabilistic fatigue model for plain and fibre-reinforced concrete

M.J. Rivas-López a, R.C. Yu b,*, J. López-Fidalgo c, G. Ruiz b

a Departamento de Estadística, Universidad de Salamanca, Plaza de los Caídos s/n, 37008 Salamanca, Spain
b ETS de Ingenieros de Caminos, C. y P., Universidad de Castilla-La Mancha, Avenida Camilo José Cela s/n, 13071 Ciudad Real, Spain
c Departamento de Matemáticas, Universidad de Castilla-La Mancha, Avenida Camilo José Cela s/n, 13071 Ciudad Real, Spain

HIGHLIGHTS

• Develop a general procedure for optimal design of fatigue characterisation.
• Derive FIM for chosen fatigue models to improve design efficiency.
• Perform robustness analysis to evaluate design efficiency.

ARTICLE INFO

Article history:
Received 6 March 2016
Received in revised form 18 August 2016
Accepted 19 August 2016
Available online 26 August 2016

Keywords:
Fisher Information Matrix (FIM)
D-optimality
Weibull cumulative distribution function
Loading frequency
Fatigue model

ABSTRACT

The objective is to improve the fatigue characterisation process based on the concept of optimal experimental design. This is carried out through a probabilistic model, previously developed, which takes into account the experimentally observed loading frequency effect on the fatigue life in plain and fibre-reinforced concrete. The Fisher Information Matrix is first obtained for the simplified fatigue model. The optimal design is found to be located at the minimum values allowed for both the maximum stress and stress ratio, whereas the two loading frequencies are the minimum and maximum values in the defined range. Next, the FIM is derived for the extended fatigue model. The previously carried out experimental plan is 65% efficient compared to the optimum. Even though it has been developed for the specific chosen fatigue model, the current procedure can be applied to any other fatigue model to significantly improve the fatigue characterisation process of any material.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Fatigue tests are known to be time consuming, and can sometimes be unachieviable if not properly designed. The procedure on how to determine the number of tests needed to characterise certain materials is an open issue. In the current work, we make use of the Fisher Information Matrix (FIM) to derive the optimal location of tests to characterise fatigue performance of concrete-related materials under given loading conditions. In particular, the fatigue model based on an initial distribution developed by Saucedo et al. (2013) is chosen as an example to carry out the optimal design process. The developed methodology, however, can be applied to any other given fatigue model.

* Corresponding author.
E-mail address: rena@uclm.es (R.C. Yu).

http://dx.doi.org/10.1016/j.csda.2016.08.014
0167-9473/© 2016 Elsevier B.V. All rights reserved.
Development of high-performance concrete for structures undergoing dynamic and cyclic loading has led to experiments conducted to study the influence of different fatigue parameters (Graf and Brenner, 1934; Kesler, 1953; Nordby, 1958; Rusch, 1960; Mudock, 1965; Aas-Jakobsen, 1970; Awad and Hilsdorf, 1971; Sparks and Menzies, 1973; ACI Committee, 1974; Teppers and Kutt, 1979; Teppers, 1979; Hsu, 1981; Furtak, 1984; Oh, 1991; Zhang et al., 1996; Li et al., 2007; Zhao et al., 2007; Medeiros et al., 2015). These parameters are related to either the fatigue test conditions, such as the minimum stress, $\sigma_{\text{min}}$, the maximum stress, $\sigma_{\text{max}}$, the loading frequency, $f$, or the material properties; for example the static material strength, $\sigma_c$, (compressive or tensile). In the case of concrete, it has been detected that the influence of the stress ratio, $R$ ($\sigma_{\text{min}}$ divided by $\sigma_{\text{max}}$), the loading frequency, $f$, and the stress level, $\sigma_{\text{max}}/\sigma_c$, is quite relevant (Aas-Jakobsen, 1970; Oh, 1991; Plekho et al., 2011; Medeiros et al., 2015).

Although the influence of loading frequency has been observed as early as the 1950s (Kesler, 1953), the loading time (Hsu, 1981) or frequency (Furtak, 1984) was not included in the fatigue equation until the 1980s. Moreover, the static strength in concrete exhibits large dispersion and a Weibull distribution has been considered as the best distribution to fit fatigue life in concrete (Oh, 1991; Li et al., 2007). Recently Saucedo et al. (2013) proposed a fatigue model to take into account the fact that, on the one hand, the static strength of concrete is a Weibull distribution; on the other hand, such a distribution can be considered as the limit behaviour when the fatigue life approaches one cycle. Meanwhile, influences of the loading frequency and of the stress ratio have also been incorporated. The application range of the proposed fatigue model is below 10 Hz according to experimental tests by Ruiz et al. (2011), Medeiros et al. (2015). Depending on the effect of the loading frequency on the hardening exponent of the dynamic material strength is ignored or not, correspondingly the fatigue model is viewed as simplified or complete. We aim to explore the procedure to calibrate the model parameters based on the concept of optimal experimental design herein.

Optimal experimental designs are especially useful when experimentation is expensive, time consuming or difficult to carry out. A good design will definitely save time, money and provide a better fitting of the model. An experimental design means selecting experimental conditions in such a way that they result in precise estimates or good predictions with a minimum of sample size, which in other words means adequately choosing the levels of the covariates. The optimality of a design is model-oriented in the sense that a design could be quite good for a particular model and not so good for a rival model. In the case of the existence of rival models a discrimination test may be performed to detect the most appropriate one. Optimal designs can be computed for discrimination purposes (López-Fidalgo et al., 2007). In reliability and survival analysis, the Cox proportional hazards model (Cox, 1972) is quite traditional, although proportionality is not always justified. The model is frequently fitted using partial likelihood. Optimal experimental designs in this context were recently computed by López-Fidalgo and Rivas-López (2014). Even though optimal failure testing models have been increasingly used in recent years, experimental designs have not been studied as much in this field; see Rivas-López et al. (2014). In the current work, optimal experimental design is studied for the aforementioned fatigue model for concrete-like materials; in particular, to best estimate the model parameters under given loading conditions.

The rest of this paper is organised as follows. The concept of optimal experimental design is presented next. Applications to the simplified and complete fatigue model are developed in Sections 3 and 4, respectively. Relevant conclusions are drawn in Section 5.

2. Optimal experimental design

Let $x$ be the vector of covariates, say an experimental condition (loading frequency, maximum stress and stress ratio for a fatigue test), which can be chosen from a compact design space, $\chi$; typically a product of intervals. For a value of $x$, a response variable time-to-event, $t$, (equivalently, the fatigue life, $N$), is observed. This is considered as a random variable from a parametric family of distributions, indexed by $\beta$, a vector of parameters. An exact design of size $n$ is defined by a collection of experimental conditions $x_1, x_2, \ldots, x_n$ in $\chi$, where some of these may be repeated. Thus, a probability measure can be defined with support on the distinct points of the design with weights proportional to the number of repetitions (replicates). This leads to an extension of this definition to any probability measure, $\xi$, the so called approximate design (Kiefer, 1974), which will be used in this work. An exact design of a particular size, say $n$, is what one can only put in practice using just $n$ experiments. Extending the concept to any probability measure (approximate design) means a kind of abstraction of the real world. This is made because the concept allows a mathematical result, which proves very useful to compute optimal designs. As a matter of fact, computing optimal exact designs is rather difficult and frequently convergence of the typical algorithms cannot be proved. The main drawback is that once the optimal approximate design is computed, an exact design has to be obtained using some rounding procedure. For instance, for a discrete measure, $\xi$, the experimenter has to perform “approximately” $n_i \approx n \xi(x_i)$ experiments at $x_i$, in such a way $\sum n_i = n$. Imhof et al. (2001) provided some examples to show that if the sample size is large, any rounding procedure leads to a quite efficient exact design. Otherwise, if $n$ is small, then the impact of the rounding may be quite important. Hereafter $n$ will be the total number of experiments to be performed, while $k$ will be the distinct experimental conditions, some of them replicated. An approximate design with $k$ different points in its support will be frequently called a $k$-point design.

The Fisher information is a way of measuring the amount of information that an observable random variable $x$ contains about an unknown parameter upon which the probability model of $x$ depends. The Fisher Information Matrix (FIM)
where the parameters called General Equivalence Theorem (GET) \((Kiefer, 1974)\) can be used to check whether this is actually the D-optimal design maximised by equal weights. Thus, the best of this type of design may be computed in an easy way. Consequently the so\-of the weightsofthedesignandofafunctiondependingjustonthesupportpoints. Then the product of the weights is maximised by equal weights. Thus, the best of this type of designs may be computed in an easy way. Consequently the so called General Equivalence Theorem (GET) \((Kiefer, 1974)\) can be used to check whether this is actually the D-optimal design or not. The GET for D-optimality states that a design \(\xi_D\) is D-optimal if and only if
\[
\psi(\xi_D, \beta; x) = \ln \left[ M^{-1}(\beta; \xi_D) I(\beta; x) \right] \leq m, \quad x \in \chi.
\]
Moreover, equality is reached at the support points of the design.

If the model is non-linear then the FIM will depend on the parameters. Using nominal values of the parameters leads to the so called locally optimal designs. The way of using the GET to check D-optimality of a particular design \(\xi\) means computing the FIM for this design, \(M(\beta; \xi)\), which, with specific nominal values of the parameters, will be a matrix with numerical entries. Then \(I(\beta; x)\) will be a matrix depending just on \(x\) when the nominal values of the parameters are used. Thus, for a specific design and nominal values of the parameters, \(\psi(\xi, \beta; x)\) is just a function of \(x\). It may be checked that this function is non-negative on the design space \(x \in \chi\), either analytically or numerically, e.g. plotting it, if \(\chi\) is one or two-dimensional.

A detailed introduction to the theory of optimal experimental design is provided, e.g., by Pázman (1986), among others. Next, we search such designs for both the simplified and complete fatigue models.

3. Applications to the simplified fatigue model

In this section, we first summarise the fatigue model developed by Saucedo et al. (2013) for completeness. Then the concept of optimal experimental design is applied for the case when the effect of loading frequency on the hardening exponent of the dynamic material strength is ignored.

On the one hand, the distribution of the static compressive strength, \(\sigma_{f_0}\), for a quasi-brittle material like concrete is shown to be a three-parameter Weibull distribution (Saucedo et al., 2013)
\[
F(\sigma_{f_0}) = 1 - \exp\left\{- \left( \frac{\sigma_{f_0} - \sigma_{\text{min}_0}}{\lambda} \right)^K \right\},
\] (3)
where the parameters \(\sigma_{\text{min}_0}\) (location), \(\lambda\) (scale) and \(K\) (shape) are properties of the material and have to be estimated through static compressive tests. On the other hand, three limit conditions must be fulfilled for a fatigue test,
\[
\lim_{N \to \infty} \sigma_f = \sigma_{\text{min}_0},
\]
\[
\lim_{R \to 1} \sigma_f = \sigma_{f_0}, \quad \text{and}
\]
\[
\lim_{N \to 1} \sigma_f = \sigma_{f_0}.
\]
The first limit condition reflects the fact that when the stress level reaches the fatigue endurance limit, \(\sigma_{\text{min}_0}\), an infinite number of cycles can be resisted without causing any damage to the material. The second limit condition states that, if the stress ratio, \(R\), approaches one, the loading condition is actually static. Consequently, the material fails at its static strength, \(\sigma_{f_0}\). In the same way, the third limit condition emphasises that, if the fatigue life \(N\) is equal to one, the employed maximum stress level is necessarily the material’s static strength, \(\sigma_{f_0}\). If any of these three conditions is not satisfied for a fatigue model, the implication is that the model will not reflect physical reality at asymptotic conditions.

The following expression was proposed to relate the equivalent static strength of the fatigue load with the number of cycles resisted, \(N\),
\[
\sigma_{f_0} = \sigma_{\text{min}_0} + \left( \sigma_{\text{max}_0} - \sigma_{\text{min}_0} \right) N^{-\left(b+c \ln(1+f)/(1-R)\right)},
\] (4)
where the effect of the loading frequency, \(f\), has been taken into consideration.
According to the FIB Code (CEB-FIP, 2008), the dynamic stress $\sigma_{\text{max}}$ can be related to its static counterpart by the equation,

$$\sigma_{\text{max}0} = \sigma_{\text{max}} \left( \frac{\dot{\sigma}_0}{2f \Delta \sigma} \right)^{0.014} \equiv C_{014} \sigma_{\text{max}}, \quad (5)$$

where 0.014 is the dynamic hardening coefficient, $\dot{\sigma}_0$ is the stress rate at which static characterisation tests were carried out, whereas

$$d = \frac{\dot{\sigma}_0}{2f \Delta \sigma} = \frac{\dot{\sigma}_0}{2f(1 - R)\sigma_{\text{max}}}.$$ 

Here $\Delta \sigma$ is the stress range, i.e., $\sigma_{\text{max}} - \sigma_{\text{min}} = (1 - R)\sigma_{\text{max}}$.

By defining the non-dimensional maximum stress, $S_{\text{max}}$, and defining two intermediate variables $A$ and $B$ as

$$A = S_{\text{max}}d^{0.014} - S_{m0}, \quad (7)$$

$$B = [b + c \ln(1 + f)](1 - R), \quad (8)$$

the cumulative distribution function of $N$ can be rewritten as

$$F(N) = 1 - \exp \left\{ -A^K N^{BK} \right\}. \quad (9)$$

Thus, once the parameters of the Weibull distribution in Eq. (3) are estimated, the distribution of $N$ for a given test condition (where $\sigma_{\text{max}}, R$ and $f$ are specified), depends on the two parameters $b$ and $c$, which are to be estimated.

The likelihood and log-likelihood functions for specific values of the parameter set, $(f, \sigma_{\text{max}}, R)$, and fatigue life, $N$, are respectively

$$L(b, c) = \frac{\partial F(N)}{\partial N} = A^K B^N K^{B - 1} \exp \left\{ -A^K N^B \right\},$$

$$LL(b, c) = K \ln(A) + \ln(B) + \ln(K) + (BK - 1) \ln(N) - A^K N^B.$$

Since

$$\frac{\partial B}{\partial b} = (1 - R),$$

$$\frac{\partial B}{\partial c} = (1 - R) \ln(1 + f) = \frac{\partial B}{\partial b} \ln(1 + f),$$

then

$$\frac{\partial^2 LL}{\partial b^2} = -\frac{(1 - R)^2}{B^2} - A^K N^B K^2 (1 - R)^2 [\ln(N)]^2,$$

$$\frac{\partial^2 LL}{\partial b \partial c} = \frac{\partial^2 LL}{\partial b^2} \ln(1 + f),$$

$$\frac{\partial^2 LL}{\partial c^2} = \frac{\partial^2 LL}{\partial b^2} [\ln(1 + f)]^2,$$

and the first element of the FIM, $l_1(b, c; f, \sigma_{\text{max}}, R)$, is

$$l_{11}(b, c; f, \sigma_{\text{max}}, R) = E \left[ -\frac{\partial^2 LL}{\partial b^2} \right] = \int_0^{\infty} \left( -\frac{\partial^2 LL}{\partial b^2} \right) L(N) dN$$

$$= (\pi^2/6) + \left[ 1 - C_{EM} - K \ln(A) \right]^2$$

$$\frac{[b + c \ln(1 + f)]^2}{[b + c \ln(1 + f)]^2},$$

where $E$ stands for the mathematical expectation according to the probability distribution of the data; that is, the integral of the quantity multiplied by the density function (or the likelihood function); and

$$C_{EM} = \lim_{m \to \infty} \sum_{i=1}^m \left( \frac{1}{i} - \ln(m) \right) \simeq 0.577216$$

represents the Euler–Mascheroni constant (Lagarias, 2013). For simplicity, we omit the arguments of the $ij$th component of the FIM, $l_i(b, c; f, \sigma_{\text{max}}, R)$, by denoting it as $l_{ij}$. In addition, by defining $X(f) = \log(1 + f)$, the remaining components are written as follows

$$l_{12} = l_{21} = l_{11}X(f), \quad l_{22} = l_{11}[X(f)]^2.$$
Therefore the FIM

\[ I(b, c; f, \sigma_{\text{max}}, R) = \phi \cdot \phi^T, \]

where

\[ \phi = I_{11} \left( \frac{1}{X(f)} \right), \]

which is the FIM of a linear model with these regressors. If the D-optimal design has two points, then their weights must be equal, as stated in the Introduction. We will consider a 2-point approximate design of this type,

\[ \xi = \left\{ \begin{array}{l} (f_1, \sigma_{\text{max}1}, R_1) \\ 1/2 \\ (f_2, \sigma_{\text{max}2}, R_2) \\ 1/2 \end{array} \right\}. \]

The FIM is

\[ M(b, c; \xi) = \frac{1}{2} \left[ \begin{array}{cc} \sum_{i=1}^{2} I_{11}(f_i, \sigma_{\text{max}i}, R_i) & \sum_{i=1}^{2} I_{11}(f_i, \sigma_{\text{max}i}, R_i)X(f_i) \\ \sum_{i=1}^{2} I_{11}(f_i, \sigma_{\text{max}i}, R_i)X(f_i) & \sum_{i=1}^{2} I_{11}(f_i, \sigma_{\text{max}i}, R_i)X(f_i)^2 \end{array} \right] \]

and its determinant

\[ \det(M) = \frac{1}{4} I_{11}(f_1, \sigma_{\text{max}1}, R_1)I_{11}(f_2, \sigma_{\text{max}2}, R_2)[X(f_1) - X(f_2)]^2. \]  

(10)

Since \( I_{11} \) is non-increasing with respect to both \( \sigma_{\text{max}} \) and \( R \), the best of these designs will be supported at the smallest possible values of them. It remains for the function to be maximised solely for the variable \( f \).

For demonstration purposes, we apply the optimal design procedure to the plain concrete C2 and the reinforced concretes CF1 and CF2. The design spaces \( f \in (0, 10), \sigma_{\text{max}} \in [0.5\lambda, \lambda] \) and \( R \in (0.0001, 1) \), where 0.0001 is a small value chosen to avoid a zero stress ratio.

Table 1: Model parameters for material C2 (plain concrete), CF1 (steel-fibre reinforced concrete) and CF2 (polypropylene-fibre reinforced concrete) and fatigue test conditions in Saucedo et al. (2013) and Medeiro et al. (2015).

<table>
<thead>
<tr>
<th>Mat.</th>
<th>( b )</th>
<th>( c )</th>
<th>( \lambda ) (MPa)</th>
<th>( K )</th>
<th>( \sigma_{\text{min}} ) (MPa/s)</th>
<th>( \sigma_0 ) (MPa)</th>
<th>( \sigma_{\text{max}} ) (MPa)</th>
<th>( R ) (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>0.063</td>
<td>−0.011</td>
<td>76.1</td>
<td>19.8</td>
<td>3.1</td>
<td>0.2</td>
<td>66.4</td>
<td>0.3</td>
</tr>
<tr>
<td>CF1</td>
<td>0.049</td>
<td>0.0066</td>
<td>76.1</td>
<td>31.0</td>
<td>12.0</td>
<td>0.2</td>
<td>75.6</td>
<td>0.3</td>
</tr>
<tr>
<td>CF2</td>
<td>0.052</td>
<td>−0.001</td>
<td>70.2</td>
<td>15.8</td>
<td>4.8</td>
<td>0.2</td>
<td>62.9</td>
<td>0.3</td>
</tr>
</tbody>
</table>

By maximising Eq. (10), the determinant of the FIM for each concrete, the best two-point designs with equal weights are obtained as follows:

\[ \xi^{*}_{\text{C2}} = \left\{ \begin{array}{l} (0.0759, 38.05, 0.0001) \\ 1/2 \\ (10, 38.05, 0.0001) \\ 1/2 \end{array} \right\}. \]  

(11)

\[ \xi^{*}_{\text{CF1}} = \left\{ \begin{array}{l} (0.0339, 38.05, 0.0001) \\ 1/2 \\ (10, 38.05, 0.0001) \\ 1/2 \end{array} \right\}. \]  

(12)

\[ \xi^{*}_{\text{CF2}} = \left\{ \begin{array}{l} (0.0458, 35.10, 0.0001) \\ 1/2 \\ (10, 35.10, 0.0001) \\ 1/2 \end{array} \right\}. \]  

(13)

The sensitivity function Eq. (2) is used to check whether these designs are actually D-optimal or not. In Fig. 1, we plot the sensitivity function and the zoomed view of the top-left corner for C2, CF1 and CF2 respectively. As mentioned before, once the values of \( \sigma_{\text{max}} \) and \( R \) were fixed, the maximum (two for a two-parameter design) is attained at two points of the design; one corresponds to a low-loading frequency, the other, the maximum frequency (10 Hz) allowed.

In order to check dependency on the choice of the nominal parameters, a robustness analysis is carried out. Suppose that the values of the real parameters are \( (b_1, c_1) \). The FIM of the corresponding optimal design, \( \xi^{*}_{\text{n}} \), for them is computed. Consider another design, for example one of those given in Eqs. (11)–(13), which we denote as \( \xi^{*}_{\text{n}} \). The efficiency of the design \( \xi^{*}_{\text{n}} \) is measured by

\[ \text{Eff}_{(b_1, c_1)}(\xi^{*}_{\text{n}}) = \left[ \frac{\det(M(b_1, c_1; \xi^{*}_{\text{n}}))}{\det(M(b_1, c_1; \xi^{*}_{\text{n}}))} \right]^{1/2}. \]

For all values of \( b \) and \( c \) ranging between 50% and 150% of their initial values, the efficiency, in all the three cases, surpasses 99%.
Fig. 1. (For the simplified fatigue model) Sensitivity function for the D-optimal design over the entire frequency range, from 0 to 10 Hz, (left column) and the zoomed view at the top-left corner (right column) for (a) C2, (b) CF1 and (c) CF2, where the maximum is located at 0.0759 Hz, 0.0339 Hz and 0.0458 Hz, respectively.

Since, in the experimental design, both the maximum stress, \( \sigma_{\text{max}} \), and the stress ratio, \( R \), are fixed, as can be seen from Table 1, we apply these values to carry out the optimisation procedure again. The following designs (see Fig. 2) are obtained:

\[
\begin{align*}
\xi^*_{\text{C2}} &= \left\{ \left( 0.1199, 66.4, 0.3 \right), \left( 4, 66.4, 0.3 \right) \right\}, \\
\xi^*_{\text{CF1}} &= \left\{ \left( 0.0879, 75.6, 0.3 \right), \left( 4, 75.6, 0.3 \right) \right\}, \\
\xi^*_{\text{CF2}} &= \left\{ \left( 0.0880, 62.9, 0.3 \right), \left( 4, 62.9, 0.3 \right) \right\}.
\end{align*}
\] (14, 15, 16)

The experimental designs used in practice is equally replicated at four points for each material, that is,

\[
\xi = \left\{ \left( 4^{-2}, \sigma_{\text{max}}, 0.3 \right), \left( 4^{-1}, \sigma_{\text{max}}, 0.3 \right), \left( 4^0, \sigma_{\text{max}}, 0.3 \right), \left( 4^1, \sigma_{\text{max}}, 0.3 \right) \right\},
\] (17)

where \( \sigma_{\text{max}} \) is a constant for each material. The efficiencies of this experimental design, relative to each of the above, are 87.58%, 86.75% and 86.96% for C2, CF1 and CF2, respectively.

4. Applications to the complete fatigue model

Taking into account the loading frequency, Saucedo et al. (2013) proposed a new relationship between dynamic and static compressive strengths; namely

\[
\sigma_{\text{max}0} = \sigma_{\text{max}} \left( \frac{\sigma_0}{2f \Delta \sigma} \right)^{0.014 \exp[\gamma / f]} \equiv \sigma_{\text{max}} d^{0.014 \exp[\gamma / f]},
\] (17)

where the exponent was fitted for frequencies below 10 Hz. The difference between the above equation and Eq. (5) lies in the fact that the hardening coefficient is explicitly dependent on the loading frequency. As a result, a new model parameter,
γ, is added. Plugging Eq. (17) into Eq. (3) to determine the cumulative distribution function, which we term as the complete fatigue model, we obtain

\[ F(N) = 1 - \exp \left\{ -\overline{A}^K N^K B^K \right\}, \]

where

\[
\overline{A} = S_{\text{mx}} d^{0.014 \exp(\gamma f)} - S_{m0}, \\
B = [b + c \ln(1 + f)](1 - R),
\]

and \( S_{\text{mx}}, S_{m0} \) and \( d \) are defined as in Section 3. Thus, once the parameters of the Weibull distribution are estimated, the distribution of \( N \) depends on the variable \( \sigma_{\text{max}} \) and \( R \) in the fatigue test and the frequency, \( f \), as well as the parameters \( b, c \) and \( \gamma \).

In the previous section it was shown that the minimum values of \( \sigma_{\text{max}} \) and \( R \) had to be included in the support of the D-optimal design. For this reason, we expect the same will happen here again. So we will concentrate on the design for the loading frequency, \( f \).

The log-likelihood function for a specific value of \( f \) is

\[ LL(N) = K \ln(\overline{A}) + \ln(B) + \ln(K) + (BK - 1) \ln(N) - \overline{A}^K N^K B^K. \]

Since

\[
\frac{\partial \overline{A}}{\partial \gamma} = S_{\text{mx}} d^{0.014 \exp(\gamma f)} 0.014 \exp(\gamma f) f \ln(d)
\]

\[ = 0.014 \exp(\gamma f) f \ln(d)(\overline{A} + S_{m0}), \]

\[
\frac{\partial B}{\partial b} = (1 - R), \]

\[
\frac{\partial B}{\partial c} = (1 - R) \ln(1 + f) = \frac{\partial B}{\partial b} \ln(1 + f),
\]
after differentiating $LL$ twice, we have

$$\frac{\partial^2 LL}{\partial b^2} = -\frac{(1 - R)^2}{B^2} - \frac{K}{A} N^{BK} K^2 (1 - R)^2 (\ln(N))^2,$$

$$\frac{\partial^2 LL}{\partial bc} = \frac{\partial^2 LL}{\partial b^2} \ln(1 + f),$$

$$\frac{\partial^2 LL}{\partial b\gamma} = -\frac{K}{A} N^{BK} K^2 (1 - R) 0.014 \exp(\gamma f) f \ln(d) \ln(N)(\bar{A} + S_{m0}),$$

$$\frac{\partial^2 LL}{\partial c^2} = \frac{\partial^2 LL}{\partial b^2} [\ln(1 + f)]^2,$$

$$\frac{\partial^2 LL}{\partial c\gamma} = \frac{\partial^2 LL}{\partial b\gamma} \ln(1 + f),$$

$$\frac{\partial^2 LL}{\partial \gamma^2} = -\frac{K0.014 \exp(\gamma f) f^2 \ln(d)(\bar{A} + S_{m0})}{A^2} \left\{ -\bar{A} + 0.014 \exp(\gamma f) \ln(d)S_{m0}N^{2BK} + \bar{A} N^{BK} [\bar{A} - 0.014 \exp(\gamma f) \ln(d)S_{m0} + (\bar{A} + S_{m0})0.014 \exp(\gamma f) \ln(d)K] \right\}.$$

Defining

$$H(f) = \frac{0.42278 - K \ln(\bar{A})}{b + c \ln(1 + f)}, \quad M(f) = \frac{K0.014 \exp(\gamma f) f \ln(d)(\bar{A} + S_{m0})}{\bar{A}}$$

and $X(f) = \log(1 + f)$ as before, the elements of the FIM are the following expected values

$$I_{11}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial b^2} \right] = \frac{(\pi^2/6)}{[b + c \ln(1 + f)]^2} + H(f)^2 \equiv I_{11}(f),$$

$$I_{12}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial bc} \right] = I_{11}(f)X(f) \equiv I_{12}(f),$$

$$I_{13}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial b\gamma} \right] = H(f)M(f) \equiv I_{13}(f),$$

$$I_{22}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial c^2} \right] = I_{11}(f)[X(f)]^2 \equiv I_{22}(f),$$

$$I_{23}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial c\gamma} \right] = I_{13}(f)X(f) \equiv I_{23}(f),$$

$$I_{33}(b, c, \gamma; f) = E \left[ -\frac{\partial^2 LL}{\partial \gamma^2} \right] = [M(f)]^2 \frac{\bar{A} + S_{m0})K - S_{m0} + 2S_{m0}\bar{A}^{-2K}}{(\bar{A} + S_{m0})K} \equiv I_{33}(f).$$

Again $E$ stands for the mathematical expectation; that is the integral of the quantity multiplied by the density function, or likewise the likelihood function. Typical integration methods were used to find these expressions. Consequently, the FIM at a particular value of $f$ can be expressed as

$$I(f) = \begin{pmatrix} I_{11}(f) & I_{12}(f)X(f) & I_{13}(f) \\ I_{12}(f)X(f) & I_{11}(f)[X(f)]^2 & I_{13}(f)X(f) \\ I_{13}(f) & I_{13}(f)X(f) & I_{33}(f) \end{pmatrix}.$$

For an approximate design with $k$ points in its support,

$$\xi = \begin{pmatrix} f_1 & f_2 & \cdots & f_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix},$$
then the FIM is

\[
M(b, c, \gamma; \xi) = \begin{pmatrix}
\sum_{i=1}^{k} I_{11}(f_i)p_i & \sum_{i=1}^{k} I_{11}(f_i)X(f_i)p_i & \sum_{i=1}^{k} I_{13}(f_i)p_i \\
\sum_{i=1}^{k} I_{11}(f_i)X(f_i)p_i & \sum_{i=1}^{k} I_{11}(f_i)X(f_i)^2 p_i & \sum_{i=1}^{k} I_{13}(f_i)X(f_i)p_i \\
\sum_{i=1}^{k} I_{13}(f_i)p_i & \sum_{i=1}^{k} I_{13}(f_i)X(f_i)p_i & \sum_{i=1}^{k} I_{33}(f_i)p_i
\end{pmatrix}
\]

and its determinant

\[
\det(M) = \sum_{i=1}^{k} I_{33}(f_i)p_i \sum_{i=1}^{k} I_{11}(f_i)p_i \sum_{j=1}^{k} I_{11}(f_j)p_j [X(f_j) - X(f_i)]^2 \\
- \sum_{i=1}^{k} I_{11}(f_i)p_i \left( \sum_{j \neq i}^{k} I_{13}(f_j)p_j [X(f_j) - X(f_i)] \right)^2.
\]

As mentioned before, we concentrate on designing for loading frequency. In order to compare with the experimental design presented in Saucedo et al. (2013), the same values for \( \sigma_{\text{max}} \) and \( R \) are adopted as minimal values in our design. The nominal values assumed for the fatigue tests are listed in Table 2.

After some computations, the designs maximising the determinants of the FIMs for C2, CF1 and CF2 when \( f \in (0, 10] \) are as follows

\[
\xi^{*}_{C2} = \begin{pmatrix}
0.1087 \\
1/3
\end{pmatrix},
\]

\[
\xi^{*}_{CF1} = \begin{pmatrix}
0.1229 \\
1/3
\end{pmatrix},
\]

\[
\xi^{*}_{CF2} = \begin{pmatrix}
0.1181 \\
1/3
\end{pmatrix}.
\]

Table 2

Summary of the model parameters for the materials considered and fatigue test conditions in Saucedo et al. (2013) and Medeiros et al. (2015).

<table>
<thead>
<tr>
<th>Mat.</th>
<th>( b )</th>
<th>( c )</th>
<th>( \gamma )</th>
<th>( \lambda ) (MPa)</th>
<th>( K )</th>
<th>( \sigma_{\text{min}} ) (MPa/s)</th>
<th>( \phi_0 ) (MPa)</th>
<th>( \sigma_{\text{max}} ) (MPa)</th>
<th>( R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>0.061</td>
<td>0.0105</td>
<td>0.24</td>
<td>76.1</td>
<td>19.8</td>
<td>3.1</td>
<td>0.2</td>
<td>66.4</td>
<td>0.3</td>
</tr>
<tr>
<td>CF1</td>
<td>0.049</td>
<td>0.0066</td>
<td>0</td>
<td>76.1</td>
<td>31.0</td>
<td>12.0</td>
<td>0.2</td>
<td>75.6</td>
<td>0.3</td>
</tr>
<tr>
<td>CF2</td>
<td>0.052</td>
<td>0.0035</td>
<td>0.086</td>
<td>70.2</td>
<td>15.8</td>
<td>4.8</td>
<td>0.2</td>
<td>62.9</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The sensitivity function is used to check whether these designs are actually D-optimal or not. Fig. 3 shows that this function attains the maximum at the two points of the designs for the three concretes, C2, CF1 and CF2. As mentioned above, since the model is nonlinear, the optimal design will depend on the nominal values of the parameters. A sensitivity analysis versus an alternative choice of the parameters is now performed. For this, different possible true values of the parameters, \( \theta_t = (b_t, c_t, \gamma_t) \), around the nominal values will be checked. As a first step, the optimal design for those possible true values is computed, say \( \xi^* \). Then the efficiency of the designs (18)–(20), obtained for the nominal values, will be computed using the FIM built with \( \theta_t \); namely

\[
\text{Eff}_{\theta_t}(\xi^*) = \left[ \frac{\det(M(\theta_t, \xi^*))}{\det(M(\theta_t, \xi^*_{\text{opt}}))} \right]^{1/2}.
\]

This efficiency is computed for values of \( b_t \) and \( c_t \) between 0.5 and 1.5 times the nominal values and \( \gamma_t \) from 0.75 to 1.25 times the nominal value. In all cases the efficiency is over 99%. Since the term \( \log(\bar{A}) \) appears in the FIM, the values of \( \gamma_t \) are chosen to preserve the positiveness of \( \bar{A} \). This is the reason why less variation is considered for the parameter \( \gamma_t \). The experimental design used in Saucedo et al. (2013) to estimate the parameters was

\[
\xi_e = \begin{pmatrix}
4^{-2} & 4^{-1} & 4^0 & 4^1 \\
1/4 & 1/4 & 1/4 & 1/4
\end{pmatrix}
\]

and the efficiency of this design for the different concretes is \( \text{Eff}(\xi_e, \xi^{*}_{C2}) = 0.64, \text{Eff}(\xi_e, \xi^{*}_{CF1}) = 0.6465 \) and \( \text{Eff}(\xi_e, \xi^{*}_{CF2}) = 0.65. \) This means that using the D-optimal design, 35%–36% of experiments can be saved while still obtaining the same statistical results in terms of power of hypotheses tests on the parameters.
In the following design, we fix $\sigma_{\text{max}}$ and $R$ as given in Table 2. The optimisation procedure is carried out for the loading frequency $f$ solely in the range of $[0, 4]$ Hz. The following results (see Fig. 4) are obtained

$$\xi_{C2}^* = \begin{bmatrix} 0.0760 & 4 \\ 1/3 & 2/3 \end{bmatrix},$$

$$\xi_{CF1}^* = \begin{bmatrix} 0.0879 & 4 \\ 1/3 & 2/3 \end{bmatrix},$$

$$\xi_{CF2}^* = \begin{bmatrix} 0.0799 & 4 \\ 1/3 & 2/3 \end{bmatrix}.$$ (21, 22, 23)

Consequently, the efficiencies of the experimental design carried out in Saucedo et al. (2013)

$$\xi = \left\{ \begin{array}{cccc}
(4^{-1/2}, \sigma_{\text{max}}, 0.3) & (4^{-1}, \sigma_{\text{max}}, 0.3) & (4^{0}, \sigma_{\text{max}}, 0.3) & (4^{1}, \sigma_{\text{max}}, 0.3) \\
1/4 & 1/4 & 1/4 & 1/4 
\end{array} \right\}$$

are 64.01%, 65.27% and 64.82% for C2, CF1 and CF2, respectively.

5. Summary and conclusions

We have developed a general procedure for obtaining the optimal design in a fatigue characterisation process and applied the concept of optimal experimental design to a fatigue model based on the initial Weibull distribution of the static strength of concrete-like materials. The Fisher Information Matrix (FIM) is obtained both for the simplified and complete fatigue model. In particular, the optimisation process is carried out for given maximum stress, stress ratio and loading frequency ranges and a D-optimal design is obtained. The previously carried out experimental designs are shown to be 87% and 65% efficient for the simplified and complete fatigue model, respectively. In other words, 13% or 35% of the experiments could have been saved if the attained optimal experimental designs had been considered. In addition, a robustness analysis was carried out to check the dependence of the choice of the nominal model parameters; the efficiency surpasses 99% when $b$ and $c$ vary between 50% and 150%, and $\gamma$ changes between 75% and 125% of their respective nominal values.
Fig. 4. (Complete fatigue model) Sensitivity function for the D-optimal design over the entire frequency range from 0 to 4 Hz (left column) and the zoomed view (right column) for (a) C2, (b) CF1, and (c) CF2, where the maximum is located at 0.0760 Hz, 0.0879 Hz and 0.0799 Hz, respectively.

Even though the procedure has been demonstrated for a chosen fatigue model, it is applicable to any fatigue model for any other material.

Acknowledgements

The authors acknowledge the financial support from the Ministerio de Economía y Competitividad and funds FEDER BIA2015-68678-C2-1-R, MTM2013-47879-C2-1-P, MTM2013-47879-C2-2-P, Junta de Castilla-La Mancha (PEII-2014-016-P) and Junta de Castilla y León (SA130U14).

Appendix

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\text{max}}$</td>
<td>Maximum stress under fatigue loading</td>
</tr>
<tr>
<td>$\sigma_{\text{min}}$</td>
<td>Minimum stress under fatigue loading</td>
</tr>
<tr>
<td>$\sigma_{\text{max}}^0$</td>
<td>Equivalent static value of the dynamic stress, $\sigma_{\text{max}}$</td>
</tr>
<tr>
<td>$\Delta \sigma$</td>
<td>Stress range $\sigma_{\text{max}} - \sigma_{\text{min}}$</td>
</tr>
<tr>
<td>$R$</td>
<td>Stress ratio defined as $\sigma_{\text{min}}/\sigma_{\text{max}}$</td>
</tr>
<tr>
<td>$\sigma_f$</td>
<td>Failure stress</td>
</tr>
<tr>
<td>$\sigma_{\text{min}}^0$</td>
<td>Fatigue endurance limit</td>
</tr>
<tr>
<td>$\dot{\sigma}_0$</td>
<td>Loading rate of the quasi-static test</td>
</tr>
<tr>
<td>$\sigma_c$</td>
<td>Critical stress, can be compressive $(f_c)$, tensile $(f_t)$</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of cycles to failure (fatigue life)</td>
</tr>
<tr>
<td>$\lambda, K$</td>
<td>Scale and shape parameter of the Weibull distribution</td>
</tr>
<tr>
<td>$S_{\text{mx}}$</td>
<td>Normalised maximum stress $\sigma_{\text{max}}/\lambda$</td>
</tr>
<tr>
<td>$S_{\text{mn}}$</td>
<td>Normalised minimum stress $\sigma_{\text{min}}/\lambda$</td>
</tr>
<tr>
<td>$PF$</td>
<td>Probability of failure in any point of the domain $\sigma_f - \ln N$</td>
</tr>
<tr>
<td>$f$</td>
<td>Loading frequency employed in a fatigue test</td>
</tr>
</tbody>
</table>

(continued on next page)
Parameter that adjusts the relation between \( \ln N \) and \( f, R \)

Coefficient that takes into consideration of loading frequency for material strength under dynamic loading

Approximate design as a probability measure

The vector of covariates defining a particular experiment

Vector of model parameters

Mathematical expectation

The log-likelihood function of the observation \( t \) at \( x \)

Fisher Information Matrix at point \( x \)

Fisher Information Matrix at design \( \xi \)

References


Graf, O., Brenner, E., 1934. Experiments for investigating the resistance of concrete under often repeated compression loads (in German), Bulletin No. 76 and 83. Deutscher Ausschuss fr Eisenbeton.


Convex relaxation for IMSE optimal design in random-field models

B. Gauthier\textsuperscript{a,b}, L. Pronzato\textsuperscript{b,*}

\textsuperscript{a} ESAT-STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium
\textsuperscript{b} Université Côte d’Azur, CNRS, I3S, France

HIGHLIGHTS

- A method is proposed to construct IMSE optimal designs for random field models.
- The model is interpreted as an (heteroscedastic) approximate Bayesian linear model.
- The design problem is transformed into a convex optimisation problem.
- The presence of a parametric trend is considered in detail through kernel reduction.

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form 19 October 2016
Accepted 20 October 2016
Available online 27 October 2016

Keywords:
Random-field model
Bayesian linear model
IMSE
Optimal design of experiments
Integral operator
Kernel reduction

ABSTRACT

The construction of optimal designs for random-field interpolation models via convex design theory is considered. The definition of an Integrated Mean-Squared Error (IMSE) criterion yields a particular Karhunen–Loève expansion of the underlying random field. After spectral truncation, the model can be interpreted as a Bayesian (or regularised) linear model based on eigenfunctions of this Karhunen–Loève expansion, and can be further approximated by a linear model involving orthogonal observation errors. Using the continuous relaxation of approximate design theory, the search of an IMSE optimal design can then be turned into a Bayesian $A$-optimal design problem, which can be efficiently solved by convex optimisation. A careful analysis of this approach is presented, also including the situation where the model contains a linear parametric trend, which requires specific treatments. Several approaches are proposed, one of them enforcing orthogonality between the trend functions and the complementary random field. Convex optimisation, based on a quadrature approximation of the IMSE criterion and a discretisation of the design space, yields an optimal design in the form of a probability measure with finite support. A greedy extraction procedure of the exchange type is proposed for the selection of observation locations within this support, the size of the extracted design being controlled by the level of spectral truncation. The performance of the approach is investigated on a series of examples indicating that designs with high IMSE efficiency are easily obtained.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

This work addresses the problem of computing designs of experiments for second-order Random Field (RF) interpolation models with known covariance (structure and parameters), that are optimal in terms of Integrated Mean-Squared Error

\* Corresponding author.
E-mail addresses: bgauthie@esat.kuleuven.be (B. Gauthier), pronzato@cnrs.fr (L. Pronzato).

http://dx.doi.org/10.1016/j.csda.2016.10.018
0167-9473/© 2016 Elsevier B.V. All rights reserved.
(IMSE), see, e.g., Sacks et al. (1989) and Rasmussen and Williams (2006). The direct computation of IMSE-optimal designs for kernel-based models is often considered as a numerically challenging problem, see, e.g., Fang et al. (2010, Chap. 2) and Santner et al. (2003, Chap. 6), in particular due to the presence of local minima, and we shall make an excursion through convex design theory to avoid such difficulties.

The definition of an IMSE criterion for the interpolation of a RF yields a particular Karhunen–Loève (KL) expansion of the considered field. We follow the original approach of Fedorov (1996) (see also Spöck and Pilz, 2010) and, for a given truncation level, we interpret the initial RF model as a Bayesian (or regularised) Linear Model (BLM) based on a subset of the eigenfunctions of the (IMSE-induced) KL expansion. This exact BLM involves correlated errors, and in order to apply the classical machinery of approximate design theory we need to introduce an approximate BLM with uncorrelated errors. Using continuous relaxation with design measures, the construction of an IMSE-optimal design is then turned into a standard (convex) Bayesian A-optimal design problem, as considered by Pilz (1983); see also Chaloner (1984) and Pukelsheim (1993, Chap. 11). Many convex optimisation algorithms are available to solve this problem, ensuring fast and guaranteed convergence to an optimal design measure with a finite support, see, e.g., Pronzato and Pázman (2013, Chap. 9).

The contributions of the paper are twofold. First, it presents a careful analysis of the approach described above, including the situation where the model contains a linear parametric trend. Second, since repeating observations at the same location is pointless in an interpolation context, an optimal design measure must be approximated by an exact design without repetition. We present a greedy extraction procedure that selects observations sites within the support of an optimal measure, the size of the extracted design being controlled by the truncation level in the KL expansion.

The paper is organised as follows. For the sake of simplicity, we first assume (Sections 2–4) that the mean structure of the RF is known (and equal to zero without any loss of generality); the case of RF models involving an unknown linear parametric trend is considered in Section 5. Section 2 introduces the main notions and notation used in this work. Section 3 defines the exact and approximate BLMs induced by the initial RF model and the IMSE criterion. In particular, we show the equivalence between the IMSE for the exact BLM and the truncated-IMSE in Gauthier and Pronzato (2014), and consider two different choices for the variance structure of the observation noise in the approximate BLM: one is homoscedastic, the other is heteroscedastic and has the same variance as in the exact BLM. Section 4 presents the Bayesian A-optimal design problem. Section 5 describes how the convex relaxation approach presented in previous sections can be applied to RF models with a linear parametric trend. We first recall the direct approach of Spöck and Pilz (2010). We then propose two alternatives for the construction of the exact BLM, depending whether prior information on the first two moments of the trend parameters is available or not. When an informative prior is available, we consider the initial RF model with unknown trend as a RF model with known trend and augmented covariance kernel. If no informative prior is available, we construct a reduction of the initial kernel that brings orthogonality between the trend and the complementary centred RF—without any modification of the predictive properties of the model, see Theorem 2. The numerical implementation of the approach is described in Section 6, assuming that a pointwise quadrature is used to approximate the integral of the Mean-Squared Error (MSE) and that the design space is restricted to quadrature points. A greedy extraction procedure is proposed in the same section to construct an exact design without repetition from a (finitely supported) optimal design measure. Finally, some numerical experiments are carried out in Section 7, and Section 8 concludes.

2. General framework and notations

2.1. Random fields and related Hilbert structures

We consider a real RF \((Z_x)_{x \in \mathcal{X}}\) indexed by \(\mathcal{X}\), where \(\mathcal{X}\) can be any general set (even if most applications concern compact subsets of \(\mathbb{R}^d, d \geq 1\)). In what follows \(Z\) will refer to the RF \((Z_x)_{x \in \mathcal{X}}\). We assume that \(Z\) is centred, second-order, and defined on a probability space \((\Omega, \mathcal{F}, P)\). We denote by \(L^2(\Omega, P)\) the Hilbert space of second-order real random variables (r.v.) on \((\Omega, \mathcal{F}, P)\), where we identify r.v. that are equal \(P\)-almost surely. The inner product between two r.v. \(U\) and \(V\) of \(L^2(\Omega, P)\) is denoted by \(E(UV)\).

Let \(K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}\) be the covariance kernel of \(Z\), i.e., for all \(x, y \in \mathcal{X}\), \(E(Z_x Z_y) = K(x, y)\). Also, let \(\mathbb{H}\) be the Hilbert space (sometimes called Gaussian Hilbert space) associated with \(Z\), i.e., the closed linear subspace of \(L^2(\Omega, P)\) spanned by the r.v. \(Z_x, x \in \mathcal{X}\), endowed with the Hilbert structure induced by \(L^2(\Omega, P)\). We assume that \(\mathbb{H}\) is separable.

We denote by \(\mathcal{H}\) the Reproducing Kernel Hilbert Space (RKHS) of real valued functions on \(\mathcal{X}\) defined by the kernel \(K(\cdot, \cdot)\). The two Hilbert spaces \(\mathcal{H}\) and \(\mathbb{H}\) are isometric owing to the relation, for all \(x, y \in \mathcal{X}\), \((K_\mathcal{H}(K_x, K_y)) = K(x, y) = E(Z_x Z_y)\), where \(\langle \cdot, \cdot \rangle_{\mathcal{H}}\) is the inner product of \(\mathcal{H}\). We denote by \(I : \mathcal{H} \rightarrow \mathbb{H}\) the isometry defined by \(I(K_x) = Z_x\), where \(K_x\) stands for the function \(t \mapsto K(x,t), t \in \mathcal{X}\).

2.2. Hilbert space embedding and integral operator

We suppose that \(\mathcal{X}\) is a measurable space; we denote by \(\mathcal{A}\) the associated \(\sigma\)-algebra and consider a \(\sigma\)-finite measure \(\mu\) on \(\mathcal{X}\) (used to define the IMSE criterion, see Section 2.4). We denote by \(L^2(\mathcal{X}, \mu)\) the Hilbert space of real-valued functions on \(\mathcal{X}\) that are square integrable with respect to \(\mu\). Notice that elements of \(L^2(\mathcal{X}, \mu)\) are in fact equivalent classes of functions
that coincide $\mu$-almost everywhere; however, we shall identify elements of $L^2(\mathcal{X}, \mu)$ with functions on $\mathcal{X}$ when it will not be source of confusion.

We assume that the kernel $K(\cdot, \cdot)$ is measurable on $\mathcal{X} \times \mathcal{X}$ endowed with the product $\sigma$-algebra, see for instance Steinwart and Christmann (2008, Chap. 4). Notice that since the RHKS $\mathcal{H}$ is separable, this assumption is equivalent to assuming that for all $x$ in $\mathcal{X}$, the function $K(\cdot, x)$ is measurable. In particular, $\mathcal{H}$ then consists of measurable functions on $\mathcal{X}$.

We also suppose that the diagonal of $K(\cdot, \cdot)$ is measurable, i.e., that the function $x \mapsto K(x, x)$ is measurable on $(\mathcal{X}, \mathcal{A})$.

Finally, we assume that $\mathcal{H}$ is continuously embedded into $L^2(\mathcal{X}, \mu)$ and

$$\|h\|_{L^2} \leq \tau \|h\|_{\mathcal{H}}, \quad \text{with} \quad \tau = \int_{\mathcal{X}} K(x, x) \, d\mu(x) < +\infty;$$

one may refer for instance to Gauthier and Pronzato (2014) for more precisions (notice that the integral is well defined since $K(x, x) \geq 0$). We denote by $\mathcal{H}_0$ the closed linear subspace of $\mathcal{H}$ defined by $\mathcal{H}_0 = \{h_0 \in \mathcal{H} \mid \|h_0\|_{L^2}^2 = 0\}$ and by $\mathcal{H}_\mu$ the orthogonal of $\mathcal{H}_0$ in $\mathcal{H}$ (i.e., $\mathcal{H}_\mu = \mathcal{H}_0^\perp$).

We introduce the following linear operator $T_\mu$ on $L^2(\mathcal{X}, \mu)$,

$$\forall f \in L^2(\mathcal{X}, \mu), \quad \forall x \in \mathcal{X}, \quad T_\mu[f](x) = (K_\mu[f])_T = \int_{\mathcal{X}} f(t)K(x, t) \, d\mu(t).$$

The operator $T_\mu$ is compact, positive semidefinite and self-adjoint, and $T_\mu[f] \in \mathcal{H}_\mu$ for all $f \in L^2(\mathcal{X}, \mu)$. Let \{$\lambda_k \mid k \in \mathbb{I}_+\}$ be the set (at most countable) of all strictly positive eigenvalues of $T_\mu$. We denote by $\tilde{\varphi}_k \in L^2(\mathcal{X}, \mu)$ their associated eigenfunctions, i.e., in $L^2(\mathcal{X}, \mu)$,

$$\forall k \in \mathbb{I}_+, \quad T_\mu[\tilde{\varphi}_k] = \lambda_k \tilde{\varphi}_k, \quad \text{with} \quad \lambda_k > 0,$$

chosen to be orthonormal in $L^2(\mathcal{X}, \mu)$. We shall also consider their canonical extensions $\varphi_k \in \mathcal{H}$,

$$\forall x \in \mathcal{X}, \quad \varphi_k(x) = \frac{1}{\lambda_k} T_\mu[\tilde{\varphi}_k](x), \quad k \in \mathbb{I}_+, \tag{1}$$

so that \{$\sqrt{\lambda_k} \varphi_k \mid k \in \mathbb{I}_+$\} forms an orthonormal basis of $\mathcal{H}_\mu$ for the Hilbert structure of $\mathcal{H}$, see Gauthier and Pronzato (2014, Prop. 3.1).

### 2.3. Hilbert space decomposition

To the orthogonal decomposition $\mathcal{H} = \mathcal{H}_\mu \perp \mathcal{H}_0$ of Section 2.2 corresponds the orthogonal decomposition

$$\mathbb{H} = \mathbb{H}_\mu \perp \mathbb{H}_0, \tag{2}$$

via the isometry $I$. We denote by $P_{\mathbb{H}_\mu}$ and $P_{\mathbb{H}_0}$ the orthogonal projections of $\mathbb{H}$ onto $\mathbb{H}_\mu$ and $\mathbb{H}_0$, respectively. The covariance kernel $K_\mu(\cdot, \cdot)$ of the RF $\{P_{\mathbb{H}_\mu}[Z_k]\}_{k \in \mathbb{I}_+}$ is given by

$$\forall x, y \in \mathcal{X}, \quad K_\mu(x, y) = \sum_{k \in \mathbb{I}_+} \lambda_k \varphi_k(x) \varphi_k(y),$$

and $K_0(\cdot, \cdot) = K(\cdot, \cdot) - K_\mu(\cdot, \cdot)$ is the covariance of $\{P_{\mathbb{H}_0}[Z_k]\}_{k \in \mathbb{I}_+}$.

For all $k \in \mathbb{I}_+$, we introduce the r.v. $\xi_k = I(\sqrt{\lambda_k} \varphi_k)$, so that by construction $\{\xi_k \mid k \in \mathbb{I}_+\}$ is an orthonormal basis of $\mathbb{H}_\mu$ (see also Remark 1). We then have the following decomposition (or Karhunen–Loève expansion) in $L^2(\Omega, \mathcal{F})$:

$$\forall x \in \mathcal{X}, \quad P_{\mathbb{H}_\mu}[Z_k] = \sum_{k \in \mathbb{I}_+} \sqrt{\lambda_k} \xi_k \varphi_k(x). \tag{3}$$

**Remark 1.** If we assume that the realisations of the RF $Z$ belong to $L^2(\mathcal{X}, \mu)$ with $\mathcal{P}$-probability one, then the r.v. $\xi_k$ have the representation $\xi_k = (1/\sqrt{\lambda_k}) \int_{\mathcal{X}} Z_k \varphi_k(x) \, d\mu(x)$. However, this assumption is stronger than those used in Section 2.2 and is not essential for our study.

### 2.4. IMSE and truncated-IMSE

Let $\mathbb{H}_D$ be a closed linear subspace of $\mathbb{H}$ and denote by $P_{\mathbb{H}_D}$ the orthogonal projection of $\mathbb{H}$ onto $\mathbb{H}_D$. For $x \in \mathcal{X}$, the r.v. $P_{\mathbb{H}_D}[Z_k]$ is the best linear predictor (unbiased and in terms of the MSE) of the r.v. $Z_k$ with regard to $\mathbb{H}_D$. When $Z$ is Gaussian, this predictor is optimal and corresponds to the conditional mean of $Z_k$ relative to $\mathbb{H}_D$. The IMSE associated with $\mathbb{H}_D$ is given by

$$\text{IMSE}(\mathbb{H}_D) = \int_{\mathcal{X}} \mathbb{E}\left[\left(Z_k - P_{\mathbb{H}_D}[Z_k]\right)^2\right] \, d\mu(x) = \int_{\mathcal{X}} \mathbb{E}\left[\left(Z_k - (P_{\mathbb{H}_D}[Z_k])^2\right)\right] \, d\mu(x).$$
For \( \mathcal{I}_{\text{trc}} \) a subset of \( \mathbb{I}^+ \) (truncation subset), we introduce \( \mathbb{H}_{\text{trc}} = \operatorname{span} \{ \xi_k | k \in \mathcal{I}_{\text{trc}} \} \subseteq L^2(\Omega, \mathbb{P}) \), the closure in \( L^2(\Omega, \mathbb{P}) \) of the linear space spanned by the r.v. \( \xi_k, k \in \mathcal{I}_{\text{trc}} \). The truncated-IMSE, with truncation subset \( \mathcal{I}_{\text{trc}} \), is defined by

\[
\text{IMSE}_{\text{trc}}(\mathbb{H}_D) = \int_{\mathcal{X}} \mathbb{E} \left[ (P_{\text{trc}}[Z])^2 - (P_{\text{trc}}[P_{\text{trc}}[Z]])^2 \right] d\mu(x).
\]

(4)

More details concerning the truncated-IMSE can be found in Gauthier and Pronzato (2014); see also Harari and Steinberg (2014). We have in particular:

\[
\text{IMSE}_{\text{trc}}(\mathbb{H}_D) \leq \text{IMSE}(\mathbb{H}_D) \leq \text{IMSE}_{\text{trc}}(\mathbb{H}_D) + \sum_{k \notin \mathcal{I}_{\text{trc}}} \lambda_k,
\]

(5)

where \( k \notin \mathcal{I}_{\text{trc}} \) stands for \( k \in \mathbb{I}^+ \setminus \mathcal{I}_{\text{trc}} \).

An \( n \)-point design \( D_n = \{x_1, \ldots, x_n\} \) (with \( n \in \mathbb{N}^+ \) and \( x_i \in \mathcal{X} \)) is canonically associated with the subspace \( \mathbb{H}_{D_n} = \operatorname{span}(Z_1, \ldots, Z_n) \) of \( \mathbb{H} \). We shall use the notation \( \text{IMSE}(D_n) = \text{IMSE}(\mathbb{H}_{D_n}) \) to refer to the IMSE associated with the design \( D_n \).

3. Spectral truncation and Bayesian linear models

3.1. Interpretation of the random-field model as a Bayesian linear model

Consider a (finite) truncation subset \( \mathcal{I}_{\text{trc}} \) of \( \mathbb{I}^+ \) (usually corresponding to the \( n_{\text{trc}} \) largest eigenvalues of \( T_\mu \)). From (2) and (3), we have, in \( L^2(\Omega, \mathbb{P}) \),

\[
\forall x \in \mathcal{X}, \quad Z_x = \sum_{k \in \mathcal{I}_{\text{trc}}} \beta_k \psi_k(x) + E_x,
\]

(6)

with \( \beta_k = \sqrt{\lambda_k} \xi_k \) and \( E_x = \sum_{k \notin \mathcal{I}_{\text{trc}}} \sqrt{\lambda_k} \xi_k \psi_k(x) + P_{\text{trc}}[Z_x] \). The \( \beta_k, k \in \mathcal{I}_{\text{trc}} \), are therefore mutually orthogonal centred r.v. in \( L^2(\Omega, \mathbb{P}) \), with variance \( \lambda_k \). Also, \( (E_x)_{x \in \mathcal{X}} \) is a centred RF with covariance given by

\[
\forall x, y \in \mathcal{X}, \quad K_{\text{err}}(x, y) = K(x, y) - K_{\text{trc}}(x, y),
\]

(7)

where \( K_{\text{trc}}(x, y) = \sum_{k \in \mathcal{I}_{\text{trc}}} \lambda_k \psi_k(x) \psi_k(y) \). In addition, for all \( x \in \mathcal{X} \), the \( \beta_k \) are orthogonal to \( E_x \).

According to (6), we can thus interpret the RF model \( Z \) as a Bayesian Linear Model (BLM) with regression functions the \( \psi_k, k \in \mathcal{I}_{\text{trc}}, \) observation errors \( E_x \) and a given prior on the model parameters \( \beta_k \). We shall refer to (6) as the exact BLM induced by \( Z \) and the truncation subset \( \mathcal{I}_{\text{trc}} \). Note that the \( \beta_k \) are Gaussian when \( Z \) is Gaussian.

With vector–matrix notation, we shall denote by \( \beta \) the (column) random vector with components \( \beta_k \) and by \( \phi_{\text{trc}}(x) \) the (column) vector with components \( \psi_k(x), k \in \mathcal{I}_{\text{trc}}, x \in \mathcal{X} \), so that (6) becomes

\[
Z_x = \phi_{\text{trc}}^T(x) \beta + E_x.
\]

We shall also denote by \( \Lambda_{\text{trc}} = \operatorname{diag}(\lambda_k | k \in \mathcal{I}_{\text{trc}}) \) the covariance matrix of the random vector \( \beta \); we thus have in particular \( K_{\text{trc}}(x, y) = \phi_{\text{trc}}^T(x) \Lambda_{\text{trc}} \phi_{\text{trc}}(y) \).

3.2. IMSE for the exact Bayesian linear model

Consider the exact BLM (6) for an \( n \)-point design \( D_n = \{x_1, \ldots, x_n\} \). Define the design matrix \( \Phi_{\text{trc}} \), with \( i, k \) entry \( \psi_k(x_i) \), \( 1 \leq i \leq n \) and \( k \in \mathcal{I}_{\text{trc}} \), and the covariance matrix \( \Lambda_{\text{trc}} \) of the observation errors \( (E_{x_1}, \ldots, E_{x_n}) \), with \( i, j \) entry \( \Lambda_{\text{trc}}(x_i, x_j) \), see (7). The covariance matrix \( \mathbf{K} \) of the vector of observations \( \mathbf{z} = (Z_{x_1}, \ldots, Z_{x_n})^T \) is then given by \( \mathbf{K} = \Phi_{\text{trc}} \Lambda_{\text{trc}} \Phi_{\text{trc}}^T + \mathbf{K}_{\text{err}} \). For the sake of simplicity, we assume that the design \( D_n \) is such that \( \mathbf{K}_{\text{err}} \) (and thus \( \mathbf{K} \)) is invertible, but extension to singular matrices is possible through the use of generalised inverses.

We consider the following estimator of \( \beta \):

\[
\hat{\beta} = \left( \Phi_{\text{trc}}^T \Lambda_{\text{trc}}^{-1} \Phi_{\text{trc}} + \Lambda_{\text{trc}}^{-1} \right)^{-1} \Phi_{\text{trc}}^T \mathbf{K}_{\text{err}}^{-1} \mathbf{z},
\]

which is solution of the regularised least-squares problem defined by the minimisation of

\[
\mathcal{L}^2(\beta) = (\mathbf{z} - \Phi_{\text{trc}} \beta)^T \Lambda_{\text{trc}}^{-1} (\mathbf{z} - \Phi_{\text{trc}} \beta) + (\beta - \beta_0)^T \Lambda_{\text{trc}}^{-1} (\beta - \beta_0),
\]

(8)

with \( \beta_0 = \mathbb{E}(\beta) = 0 \). Note that \( \hat{\beta} \) is simply the posterior mean of \( \beta \) when \( Z \) is Gaussian. The Mean-Squared prediction Error (MSE) at \( x \in \mathcal{X} \) for the exact BLM is

\[
\text{MSE}_{\text{trc}}^\mathbf{K}(x; D_n) = \mathbb{E} \left[ \left( \phi_{\text{trc}}^T(x) (\beta - \hat{\beta}) \right)^2 \right] = \phi_{\text{trc}}^T(x) (\Phi_{\text{trc}}^T \Phi_{\text{trc}} + \Lambda_{\text{trc}})^{-1} \phi_{\text{trc}}(x).
\]

(9)
Remark 2. From the Sherman–Morrison–Woodbury matrix identity, we have

\[
(\Phi^T_{\text{trc}} K^{-1}_{\text{err}} \Phi_{\text{trc}} + \Lambda^{-1}_{\text{trc}})^{-1} = A_{\text{trc}} - A_{\text{trc}} \Phi^T_{\text{trc}} (\Phi_{\text{trc}} A_{\text{trc}} \Phi^T_{\text{trc}} + K_{\text{err}})^{-1} \Phi_{\text{trc}} A_{\text{trc}},
\]

\[
= A_{\text{trc}} - A_{\text{trc}} \Phi^T_{\text{trc}} K^{-1} \Phi_{\text{trc}} A_{\text{trc}}.
\]

For \( x \in \mathcal{X} \), the predictor \( \hat{\beta}(x) \) can thus be written as

\[
\hat{\beta}(x) = \Phi^T_{\text{trc}}(x) \Lambda_{\text{trc}} \Phi^T_{\text{trc}} (\Phi_{\text{trc}} A_{\text{trc}} \Phi^T_{\text{trc}} + K_{\text{err}})^{-1} \Phi_{\text{trc}} A_{\text{trc}}.
\]

It corresponds to the usual kriging predictor for a (centred) RF model with covariance kernel \( K_{\text{err}}(\cdot, \cdot) \) combined with centred observation errors with covariance \( K_{\text{err}}(\cdot, \cdot) \). In the same way, we obtain for the MSE

\[
\text{MSE}_{\text{err}}(x; D_n) = K_{\text{err}}(x, x) - \Phi^T_{\text{trc}}(x) A_{\text{trc}} \Phi^T_{\text{trc}} (\Phi_{\text{trc}} A_{\text{trc}} \Phi^T_{\text{trc}} + K_{\text{err}})^{-1} \Phi_{\text{trc}} A_{\text{trc}} \Phi_{\text{trc}}(x),
\]

which is, as expected, the usual (simple) kriging variance for a RF model with observation errors.

Integrating (9) with respect to \( \mu \) and applying Fubini’s theorem, we obtain

\[
\text{IMSE}_{\text{err}}(D_n) = \int_{\mathcal{X}} \text{MSE}_{\text{err}}(x; D_n) \, d\mu(x),
\]

\[
= \text{trace}\left\{ (\Phi^T_{\text{trc}} K^{-1}_{\text{err}} \Phi_{\text{trc}} + \Lambda^{-1}_{\text{trc}})^{-1} \int_{\mathcal{X}} \Phi^T_{\text{trc}}(x) \Phi^T_{\text{trc}}(x) \, d\mu(x) \right\},
\]

\[
= \text{trace}\left\{ (\Phi^T_{\text{trc}} K^{-1}_{\text{err}} \Phi_{\text{trc}} + \Lambda^{-1}_{\text{trc}})^{-1} \right\},
\]

(10)

where we have used the property that the eigenfunctions \( \varphi_k \) are orthonormal in \( L^2(\mathcal{X}, \mu) \). Using (10), we obtain that \( \text{IMSE}_{\text{err}}(D_n) \) given by (11) coincides with the truncated-IMSE (4) for the design \( D_n \), i.e., \( \text{IMSE}_{\text{err}}(D_n) = \text{IMSE}_{\text{trc}}(D_n) \), which can be written as

\[
\text{IMSE}_{\text{trc}}(D_n) = \text{trace}\left( A_{\text{trc}} - A_{\text{trc}} \Phi^T_{\text{trc}} K^{-1}_{\text{err}} \Phi_{\text{trc}} A_{\text{trc}} \right) = \text{trace}(A_{\text{trc}}) - \text{trace}(K^{-1}_{\text{trc}} A_{\text{trc}}^2 \Phi^T_{\text{trc}}),
\]

see Gauthier and Pronzato (2014). Notice that although the model (6) is exact, it only gives access to the truncated version of \( \text{IMSE}(D_n) \).

3.3. Approximate Bayesian linear model

The main motivation for interpreting a RF model as a BLM is to be in a position to use the classical machinery of approximate-design theory for linear models. However, this cannot be applied directly to the exact model (6) where the observation errors are correlated. We therefore introduce an approximate linear model with uncorrelated errors, and consider

\[
\forall x \in \mathcal{X}, \quad \tilde{Z}_x = \sum_{k \in \mathcal{I}_{\text{err}}} \beta_k \varphi_k(x) + \varepsilon_x,
\]

(12)

where the \( \beta_k \) and \( \varphi_k \) are defined as in (6) and where the errors \( \varepsilon_x \in L^2(\Omega, \mathcal{F}) \) are centred, orthogonal to the \( \beta_k \), \( k \in \mathcal{I}_{\text{err}} \), and such that, for \( x, y \in \mathcal{X} \),

\[
E(\varepsilon_x \varepsilon_y) = \Sigma(x, y) = \begin{cases} \sigma^2(x) & \text{if } x = y, \\ 0 & \text{otherwise}. \end{cases}
\]

In order to ensure that the two models (6) and (12) have the same integrated variance with respect to \( \mu \), we impose that the variance \( \sigma^2(x) \) satisfies

\[
\tau_{\text{err}} = \int_{\mathcal{X}} K_{\text{err}}(x, x) \, d\mu(x) = \sum_{k \not\in \mathcal{I}_{\text{err}}} \lambda_k = \tau - \tau_{\text{trc}} = \int_{\mathcal{X}} \sigma^2(x) \, d\mu(x),
\]

(13)

with \( \tau_{\text{trc}} = \sum_{k \in \mathcal{I}_{\text{err}}} \lambda_k \).

For an \( n \)-point design \( D_n = \{ x_1, \ldots, x_n \} \), we denote by \( \Sigma \) the (diagonal) covariance matrix of the observation errors, with \( i, j \) entry \( \Sigma(x_i, x_j) \). The covariance matrix \( K \) of the vector of observations \( \tilde{Z} = (\tilde{Z}_{x_1}, \ldots, \tilde{Z}_{x_n})^T \) is thus \( K = \Phi_{\text{trc}} A_{\text{trc}} \Phi^T_{\text{trc}} + \Sigma \). As in Section 3.2, we assume for the sake of simplicity that \( \Sigma \) (and thus \( K \)) is invertible. The IMSE calculated with the approximate model (12) is then given by (see (11))

\[
\text{IMSE}_{\text{trc}}(D_n) = \text{trace}\left[ (\Phi^T_{\text{trc}} \Sigma^{-1} \Phi_{\text{trc}} + \Lambda^{-1}_{\text{trc}})^{-1} \right].
\]

(14)
Homoscedastic errors with $\sigma^2(x) = \sigma^2$ constant are considered in Spöck and Pilz (2010), which gives (assuming that the measure $\mu$ is finite) $\sigma^2 = \tau_{err}/\mu(\mathcal{X})$ for any $x \in \mathcal{X}$. However, even in the case when $K(x, x)$ is constant, the function $x \in \mathcal{X} \mapsto K_{err}(x, x)$ is generally strongly oscillating due to the form of the eigenfunctions that enter $K_{err}(x, x)$, see for instance Figs. 1 and 2 in Section 7.1. Choosing a heteroscedastic model with $\sigma_x \in \mathcal{X}$ and the measure $\Psi$ and the convexity and differentiability of the information matrix $\Psi$ are illustrated in Section 7. One can readily check that for a given truncation subset $\mathcal{S}$, which is convex in $\nu$ and $M$, has full rank, the function $\Phi_{\text{trc}}(\nu)$ is measurable on $\mathcal{X}$. These assumptions ensure that the information matrix

$$
\mathbf{M}_\nu = \int_{\mathcal{X}} \frac{1}{\sigma^2(x)} \Phi_{\text{trc}}(x) \phi_{\text{trc}}^T(x) \, d\nu(x),
$$

with $k$-entry $\int_{\mathcal{X}} [1/\sigma^2(x)] \phi(x) \phi(x) \, d\nu(x)$, is well defined for any measure $\nu$ in the set $\mathcal{P}$ of probability measures on $(\mathcal{X}, \mathcal{A})$. $\mathbf{M}_\nu$ is bounded, symmetric and non-negative definite.

4. Truncated-IMSE and continuous-design relaxation

We consider a linear model of the form (12) with orthogonal observation errors having variance $\sigma^2(\cdot)$. We assume that $x \mapsto \sigma^2(x)$ is measurable on $(\mathcal{X}, \mathcal{A})$, with $\sigma^2(x) \geq \tau > 0$ for all $x \in \mathcal{X}$ (see Remark 3), and that $K(x, x)$ is bounded on $\mathcal{X}$ by some constant $C$. This implies that

$$
\forall x \in \mathcal{X}, \quad |\phi_{\text{trc}}(x)| = |(\phi_{\text{trc}}|_{\mathcal{X}})| \leq \|\phi_{\text{trc}}\|_{\mathcal{X}} = \sqrt{K(x, x)/\lambda_k} \leq \sqrt{C/\lambda_k},
$$

and $\phi_{\text{trc}}(x)$ is bounded on $\mathcal{X}$. These assumptions ensure that the information matrix $\Phi_{\text{trc}}(x)\phi_{\text{trc}}^T(x)$ is measurable on $\mathcal{X}$. The truncation subset $\mathcal{S}$ and parameter $\alpha$ have a strong impact on the optimal measures that minimise $\psi_a(\cdot)$, as illustrated in Section 7. One can readily check that for a given truncation subset $\mathcal{S}$ and for a fixed $\nu \in \mathcal{P}$ such that $\mathbf{M}_\nu$ has full rank, the function $\alpha \in \mathbb{R}^+ \mapsto \psi_a(\nu)$ is positive, decreasing, and satisfies

$$
\psi_a(\nu) = \sum_{k \in \mathcal{S}} \lambda_k \quad \text{and} \quad \lim_{\alpha \to +\infty} \psi_a(\nu) = 0.
$$

For example, for the particular measure $d\nu(x) = [\sigma^2(x)/\tau_{err}] \, d\mu(x)$, with $\tau_{err} = \tau - \tau_{err}$ and $\sigma^2(\cdot)$ satisfying (13), we obtain $\mathbf{M}_\nu = \mathbf{I}d/\tau_{err}$ and $\psi_a(\nu) = \sum_{k \in \mathcal{S}} (\alpha \tau_{err} + 1/\lambda_k)^{-1}$.

The function $\psi_a(\cdot)$ defined by (16) (with $\alpha > 0$) corresponds to a Bayesian $A$-optimality criterion, see Pilz (1983), Chaloner (1984) and Pukelsheim (1993, Chap. 11), which is convex in $\nu$ and non-increasing for the Loewner ordering; that is, $\psi_a(\nu') \leq \psi_a(\nu)$ when $\mathbf{M}_{\nu'} \preceq \mathbf{M}_\nu$ is non-negative definite. In order to be able to address the minimisation of $\psi_a(\cdot)$ over $\mathcal{P}$, we assume additionally that the set $\{\Phi_{\text{trc}}(x)\phi_{\text{trc}}^T(x)/\sigma^2(x)\} \subseteq \mathcal{X}$ is compact. This is not too restrictive, and is satisfied in particular when $\mathcal{X}$ is finite, or is compact with $\sigma(\cdot)$ and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ continuous (so that the $\phi_{\text{trc}}(\cdot)$ are continuous on $\mathcal{X}$). Then, from Caratheodory’s theorem (see, e.g., Fedorov, 1972, p. 66), there exists at least one probability measure $\nu^* \in \mathcal{P}$ which minimises $\psi_a(\cdot)$ and has $m \leq 1 + \mathcal{S} + 1/2$ support points in $\mathcal{X}$; moreover, the optimal matrix $\mathbf{M}_{\nu^*}$ is unique (from the strict convexity of $\mathbf{M} \mapsto \text{trace}[(\alpha \mathbf{M} + \Lambda_{\mathcal{A}}^{-1})^{-1}]$).

In fact this result holds more generally (without any continuity assumption) whenever one considers the minimisation of $\psi_a(\cdot)$ over a (convex) simplex $\mathcal{P} \subseteq \mathcal{P}$, the extreme points of which being Dirac measures on $\mathcal{X}$ such that the set $\{\Phi_{\text{trc}}(x)\phi_{\text{trc}}^T(x)/\sigma^2(x)\} \subseteq \mathcal{X}$ is compact. This is the case in particular when $\mathcal{P}$ is the set of all probability measures supported by a given finite subset of $\mathcal{X}$, as considered in Section 6.

The directional derivative of $\psi_a(\cdot)$ at $\nu$ in the direction $\eta - \nu$ is given by

$$
F_{\psi_a}(\nu, \eta) = \lim_{\gamma \to 0^+} \frac{\psi_a[(1 - \gamma)\nu + \gamma \eta] - \psi_a(\nu)}{\gamma},
$$

$$
= -\alpha \text{trace} \left\{ (\alpha \mathbf{M}_\nu + \Lambda_{\mathcal{A}}^{-1})(\mathbf{M}_\eta - \mathbf{M}_\nu)(\alpha \mathbf{M}_\nu + \Lambda_{\mathcal{A}}^{-1})^{-1} \right\},
$$

and the convexity and differentiability of $\psi_a(\cdot)$ imply that a measure $\nu^* \in \mathcal{P}$ is optimal if and only if $F_{\psi_a}(\nu^*, \eta) \geq 0$ for any $\eta \in \mathcal{P}$. This corresponds to the Equivalence Theorem for Bayesian $A$-optimal design stated below, see Pilz (1983).
Assume that there exists a construction of a sequence of measures \( \nu \) (Chap. 9).

Theorem 1. For any given \( \alpha > 0 \), the measure \( \nu^* \in \mathcal{P} \) minimises \( \Psi_\alpha (\cdot) \) if and only if, for all \( \delta_\alpha \in \mathcal{P} \),

\[
\phi_{trc}^T (x) (\alpha M_{\nu^*} + \Lambda_{trc}^{-1})^{-2} \phi_{trc} (x) \leq \sigma^2 (x) \text{trace} \left[ M_{\nu^*} (\alpha M_{\nu^*} + \Lambda_{trc}^{-1})^{-2} \right].
\] (18)

Equality is achieved in (18) on the support of \( \nu^* \). The convexity and differentiability of \( \Psi_\alpha (\cdot) \) imply that, for any \( v \in \mathcal{P} \),

\[
\Psi_\alpha (v) \leq \Psi_\alpha (\nu^*) - \min_{\delta_\alpha \in \mathcal{P}} F_{\phi_{trc}} (v, \delta_\alpha),
\] (19)

which can be used to check distance from optimality. Many efficient convex optimisation algorithms are available for the construction of a sequence of measures \( \nu^{(k)} \) such that \( \Psi_\alpha (\nu^{(k)}) \) converges to \( \Psi_\alpha (\nu^*) \), see, e.g., Pronzato and Pázman (2013, Chap. 9).

Remark 3. Assume that there exists \( x_0 \in \mathcal{X} \) such that \( \sigma^2 (x_0) = 0 \); we can distinguish two cases.

1. There exists \( k \in \mathcal{I}_{trc} \) such that \( \varphi_k (x_0) \neq 0 \) (and therefore, \( K_{trc} (x_0, x_0) \neq 0 \)). One should then include \( x_0 \) in the design since it allows observation of the exact value of \( \phi_{trc}^* (x_0) \).

2. \( K_{trc} (x_0, x_0) = 0 \). In that case, we can roughly say that there is nothing to learn in \( x_0 \), and we may exclude \( x_0 \) from the search space. This can be achieved for instance by considering the pseudo-inverse \((\sigma^2 (x))^T\) of \( \sigma^2 (x) \) in (15), with \( r^T = 1/r \) if \( r \neq 0 \), and \( r = 0 \) if \( r = 0 \).

5. Random-field models with unknown linear parametric trend

The convex relaxation described in previous sections can be extended to RF models that include an unknown linear parametric trend, as described in Spöck and Pilz (2010). Their direct approach is detailed in Sections 5.1 and 5.2. When a prior on the trend parameters is available, we show in Section 5.3 how a rather straightforward kernel augmentation allows us to recast the problem as one involving a RF with known trend. In absence of informative prior on the trend parameters, a kernel reduction is proposed in Section 5.4 that permits to avoid confusion between the trend and RF behaviours.

5.1. IMSE for a random field with unknown linear parametric trend

We still consider the framework and notation of Section 2, but now the RF is \( (Y_x)_{x \in \mathcal{X}} \) such that, for all \( x \in \mathcal{X} \),

\[
Y_x = g^T (x) \theta + Z_x,
\] (20)

where \( g(x) = (g_1(x), \ldots, g_p(x))^T \) is a column vector of (known) real-valued functions on \( \mathcal{X} \) and where \( \theta = (\theta_1, \ldots, \theta_p)^T \in \mathbb{R}^p \) is an unknown vector, \( p \in \mathbb{N}^* \). We denote by \( T \) the linear subspace spanned by the trend functions \( \{g_1, \ldots, g_p\} \). We suppose that \( g_j \in L^2 (\mathcal{X}, \mu) \) for all \( j \in \{1, \ldots, p\} \) and, in view of Section 4, we also assume that the set \( \{g(x)g^T (x) / \sigma^2 (x) | x \in \mathcal{X} \} \) is compact.

Following Spöck and Pilz (2010), we consider a prior on the first two moments of \( \theta \) and assume that \( \theta_i \in L^2 (\Omega, \mathcal{F}) \) for all \( i \in \{1, \ldots, p\} \) with

\[
E (\theta) = \theta_0 \quad \text{and} \quad \text{Cov} (\theta) = A,
\] (21)

where \( \theta_0 \in \mathbb{R}^p \) and \( A \) is a \( p \times p \) positive-definite matrix. We also assume that the \( \theta_i \) are orthogonal to \( \mathbb{H} \). The case where no prior information on \( \theta \) is available corresponds to replacing \( A^{-1} \) by the null matrix and will be detailed in Section 5.4.

For the RF model (20) and design \( D_n = \{x_1, \ldots, x_n\} \), the MSE at \( x \in \mathcal{X} \) is given by (assuming that all matrix inverses are well-defined)

\[
\text{MSE} (x; D_n) = K(x, x) - k^T (x) K^{-1} k(x) + [g(x) - G^T K^{-1} k(x)]^T (G^T K^{-1} G + A^{-1})^{-1} [g(x) - G^T K^{-1} k(x)],
\] (22)

where \( G \) is the \( n \times p \) design matrix with \( i, j \) entry \( G_{ij} = g_i (x_j) \), and where

\[
k(x) = (K_{x_1} (x), \ldots, K_{x_n} (x))^T.
\]

This is the usual expression obtained in Bayesian kriging, see Omre and Halvorsen (1989) and Santner et al. (2003, Chap. 4). The IMSE criterion is then given by

\[
\text{IMSE} (D_n) = \int_{\mathcal{X}} \text{MSE} (x; D_n) \, d\mu (x).
\] (23)
5.2. Direct spectral approximation and induced exact BLM

Consider the decomposition (6) for a truncation subset $\mathbb{1}_{\text{trc}}$. It yields the following exact model (with equality in $L^2(\Omega, \mathbb{P})$),

$$Y_k = \tilde{f}(x) \gamma + E_k,$$

where $\tilde{f}(x) = (\phi_{\text{trc}}(x))^T \beta(x)$ and $\gamma = (\beta^T \beta)^T$, with

$$E(\gamma) = \gamma_0 = \begin{pmatrix} \theta_0 \\ 0 \end{pmatrix} \quad \text{and} \quad \text{Cov}(\gamma) = \Gamma = \begin{pmatrix} A & 0 \\ 0 & A_{\text{trc}} \end{pmatrix}.$$  

(25)

For a design $D_n = \{x_1, \ldots, x_n\}$, denote by $\mathbf{y}$ the vector of observations $(Y_{x_1}, \ldots, Y_{x_n})^T$ and let $\mathbf{F} = (\mathbf{G}, \Phi_{\text{trc}})$. The underlying regularised least-squares problem yields the estimator

$$\hat{\gamma} = (\mathbf{G}^T \mathbf{K}_{\text{err}}^{-1} \mathbf{F} + \Gamma^{-1})^{-1} (\mathbf{G}^T \mathbf{K}_{\text{err}}^{-1} \mathbf{y} + \Gamma^{-1} \gamma_0).$$

The MSE at $x \in \mathcal{X}$ is then

$$\text{MSE}_{\text{trc}}^K(x; D_n) = \mathbb{E} \left[ (\mathbf{F}(x) (\gamma - \hat{\gamma})^2 \right] = \mathbf{F}(x) (\mathbf{F}^T \mathbf{K}_{\text{err}}^{-1} \mathbf{F} + \Gamma^{-1})^{-1} \mathbf{f}(x).$$

(26)

By expanding (26), we also obtain

$$\text{MSE}_{\text{trc}}^K(x; D_n) = \mathbf{K}_{\text{trc}}(x, x) - \mathbf{k}_{\text{trc}}^T(x) \mathbf{K}_{\text{trc}}^{-1} \mathbf{k}_{\text{trc}}(x)$$

$$+ \left[ \mathbf{g}(x) - \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{g}(x) \right]^T \left[ \mathbf{G}^T \mathbf{G} + \mathbf{A}^{-1} \right] \left[ \mathbf{g}(x) - \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{g}(x) \right],$$

(27)

where $\mathbf{k}_{\text{trc}}(x) = \Phi_{\text{trc}} A_{\text{trc}} \phi_{\text{trc}}(x)$. Similarly to Section 3.2, $\text{MSE}_{\text{trc}}^K(x; D_n)$ corresponds to the MSE obtained from (22) through spectral truncation.

Denote by $\mathbf{M}_g$ the Gram matrix of the trend functions $g_1, \ldots, g_p$ in $L^2(\mathcal{X}, \mu)$, that is, in matrix notation,

$$\mathbf{M}_g = \int_{\mathcal{X}} \mathbf{g}(x) \mathbf{g}(x)^T d\mu(x).$$

We assume that $\mathbf{M}_g$ is invertible. The truncated-IMSE is obtained by integrating (26) with respect to $\mu$, and we obtain

$$\text{IMSE}_{\text{trc}}^K(D_n) = \text{trace} \left[ (\mathbf{F}^T \mathbf{K}_{\text{err}}^{-1} \mathbf{F} + \Gamma^{-1})^{-1} \right],$$

(28)

where $\mathbf{U}$ is the Gram matrix of $f(\cdot)$ in $L^2(\mathcal{X}, \mu)$, i.e.,

$$\mathbf{U} = \int_{\mathcal{X}} f(x) f(x)^T d\mu(x) = \left( \begin{array}{c} \mathbf{M}_g \\ \left( \phi_{\text{trc}} \mathbf{g} \right)_2 \\ \mathbf{I}_{\text{trc}} \end{array} \right) \left( \begin{array}{c} \mathbf{g} \\ \phi_{\text{trc}} \mathbf{g} \\ \mathbf{I}_{\text{trc}} \end{array} \right).$$

(29)

Substituting a diagonal matrix for $\mathbf{K}_{\text{trc}}$, we obtain an approximated truncated-IMSE that can be used for convex design optimisation, in exactly the same way as in Sections 3.3 and 4, defining a Bayesian $A$-optimality criterion $\Psi_a(\cdot)$ for the model (24).

It is instructive to compare the IMSE (23) with the truncated-IMSE (28). For the initial RF model (20), after recombination we obtain $\text{IMSE}(D_n) = \tau + B - C - 2D$, where $\tau = \int_{\mathcal{X}} K(x, x) d\mu(x) = \sum_{k \in \mathcal{I}_+} \lambda_k$ and where we have set

$$B = \int_{\mathcal{X}} \mathbf{g}(x) (\mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} + \mathbf{A}^{-1})^{-1} \mathbf{g}(x) d\mu(x),$$

$$C = \int_{\mathcal{X}} \mathbf{k}_{\text{trc}}^T(x) (\mathbf{K}_{\text{trc}}^{-1} - \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} + \mathbf{A}^{-1})^{-1} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{k}_{\text{trc}}(x) d\mu(x),$$

and

$$D = \int_{\mathcal{X}} \mathbf{g}(x) (\mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} + \mathbf{A}^{-1})^{-1} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{k}_{\text{trc}}(x) d\mu(x).$$

In the same way, considering (27), we get $\text{IMSE}_{\text{trc}}^K(D_n) = \tau_{\text{trc}} + B - C_{\text{trc}} - 2D_{\text{trc}}$, with

$$C_{\text{trc}} = \int_{\mathcal{X}} \mathbf{k}_{\text{trc}}^T(x) (\mathbf{K}_{\text{trc}}^{-1} - \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} + \mathbf{A}^{-1})^{-1} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{k}_{\text{trc}}(x) d\mu(x),$$

and

$$D_{\text{trc}} = \int_{\mathcal{X}} \mathbf{g}(x) (\mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{G} + \mathbf{A}^{-1})^{-1} \mathbf{G}^T \mathbf{K}_{\text{trc}}^{-1} \mathbf{k}_{\text{trc}}(x) d\mu(x).$$

We then obtain the following (the proof is given in the Appendix).

**Proposition 1.** For any truncation subset $\mathbb{1}_{\text{trc}}$, we have $0 \leq C - C_{\text{trc}} \leq \tau_{\text{err}} = \tau - \sum_{k \in \mathcal{I}_{\text{trc}}} \lambda_k$.

On the other hand, one may note that $D$ and $D_{\text{trc}}$ respectively involve terms of the form $\int_{\mathcal{X}} K(x, x) g_i(x) d\mu(x) = T_{\mu} [g_i] (x_i) = \sum_{k \in \mathcal{I}_+} \lambda_k (\phi_k (g_i)_2 \phi_k (x_i))$ and $\int_{\mathcal{X}} K_{\text{trc}}(x, x) g_i(x) d\mu(x) = \sum_{k \in \mathcal{I}_{\text{trc}}} \lambda_k (\phi_k (g_i)_2 \phi_k (x_i))$, for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$. The derivation of general error bounds on the difference $D - D_{\text{trc}}$ seems therefore more complicated than for $C - C_{\text{trc}}$, which partly motivates the following two sections.
5.3. Presence of a prior: equivalence with a model with known trend

In presence of an informative prior on $\theta$, the model defined by (20) and (21) can be interpreted as a RF model with known trend. Indeed, for $x$ and $y$ in $\mathcal{X}$ we have

$$E(Y_x) = g^T(x)\theta_0 \quad \text{and} \quad \text{Cov}(Y_x, Y_y) = g^T(x)Ag^T(y) + K(x, y) = K^{\text{full}}(x, y).$$ \hspace{1cm} (30)

We shall refer to the kernel $K^{\text{full}}(\cdot, \cdot)$ as the augmented kernel. Applying the Sherman–Morrison–Woodbury identity (assuming, for the sake of simplicity, that $D_0$ is such that $K + GAG^T$ is invertible), we can easily check that the two RF models (20) and (30) yield the same predictors.

We can then consider the integral operator

$$T^\mu_{\text{full}}[f](x) = \int_{\mathcal{X}} f(t)K^{\text{full}}(x, t)\,d\mu(t),$$

with $f \in L^2(\mathcal{X}, \mu)$ and $x \in \mathcal{X}$, and apply the same approach as in Sections 3 and 4 for models without trend. In particular, bounds similar to (5) are available and straightforward calculation shows that

$$T^\mu_{\text{full}} = \int_{\mathcal{X}} K^{\text{full}}(x, x)\,d\mu(x) = \tau + \text{trace}(AM_g) \geq \tau.$$

5.4. Absence of informative prior: IMSE-adapted kernel reduction

Here we can take advantage of the non-uniqueness of the kernels associated with a given semi-Hilbert structure. Denote by $p$ the orthogonal projection of $L^2(\mathcal{X}, \mu)$ onto $\mathcal{T}$ and define $q = \text{id}_{L^2} - p$. For $f \in L^2(\mathcal{X}, \mu)$, we obtain, in matrix notation, $pf = g^T M_g^{-1} (g(f))_{L^2}$.

Assume, for the sake of simplicity, that the realisations of $(Z_x)_{x \in \mathcal{X}}$ belong to $L^2(\mathcal{X}, \mu)$ with $P$-probability 1 (this assumption is not necessary, however, for the construction of the kernel $K^q(\cdot, \cdot)$ given in (32) and for Theorem 2). For $x \in \mathcal{X}$, we can then define

$$pZ_x = g^T(x)M_g^{-1}\int_{\mathcal{X}} g(t)Z_t\,d\mu(t),$$

so that $pZ_x \in L^2(\Omega, \mathcal{P})$, $E(pZ_x) = 0$ and, for $y \in \mathcal{X}$,

$$E((pZ_x)(pZ_y)) = g^T(x)M_g^{-1}(T_{\mu}(g)||g^T\|)^{T^2}M_g^{-1}g^T(x),$$

and $E((pZ_x)Z_y) = g^T(x)M_g^{-1}T_{\mu}(g|y)$. We can now write the model (20) as

$$Y_x = g^T(x)\theta + pZ_x + qZ_x = g^T(x)\theta^i + qZ_x,$$ \hspace{1cm} (31)

with $\theta^i = \theta + M_g^{-1}\int_{\mathcal{X}} g(x)Z_t\,d\mu(x)$. Since no informative prior on $\theta$ is available, the prior on $\theta^i$ is non-informative too (see Remark 6). The covariance kernel of $(qZ_x)_{x \in \mathcal{X}}$ in (31) is given by

$$K^q(x, y) = E((qZ_x)(qZ_y)) = K(x, y) + g^T(x)Sg(y) - b^T(x)g(y) - g^T(x)b(y),$$ \hspace{1cm} (32)

with $S = M_g^{-1}(T_{\mu}(g||g^T)^{T^2}M_g^{-1}$ and $b(x) = M_g^{-1}T_{\mu}(g|x)$. Such a kernel $K^q(\cdot, \cdot)$ is sometimes called a reduction of the kernel $K(\cdot, \cdot)$, see Schaback [1999].

Our motivation for introducing the model (31) is that we now have orthogonality in $L^2(\mathcal{X}, \mu)$ between the realisations of $(qZ_x)_{x \in \mathcal{X}}$ and the trend subspace $\mathcal{T}$. The property below shows that predictions are not modified when using (31) instead of (20), i.e., when considering the kernel $K^q(\cdot, \cdot)$ instead of the kernel $K(\cdot, \cdot)$.

**Theorem 2.** Assume that the design $D_n$ is such that the design matrix $G$ has full rank $p$ (such a design is said to be $\mathcal{T}$-unisolvent). Then the two RF models (20) and (31) yield the same predictors and mean-squared prediction errors.

This result is a direct consequence of the non-uniqueness of the kernels associated with a given semi-Hilbert space, and of the uniqueness of the optimal prediction in semi-Hilbert spaces, see Duchon [1977] and Gauthier [2011]. A proof is given in the Appendix. Notice that if we denote by $\widehat{Y}_x$ the resulting optimal linear predictor, and by $\widehat{\theta}^i$ the underlying estimator of $\theta^i$, then, by construction, the following orthogonality holds:

$$\int_{\mathcal{X}} (\widehat{Y}_x - g^T(x)\widehat{\theta}^i)\,g(x)\,d\mu(x) = 0.$$
Remark 4. The substitution of the kernel \( K^q(\cdot, \cdot) \) for \( K(\cdot, \cdot) \) can be related to the interpretation of the initial model (20) as an intrinsic random model, see Matheron (1971, 1973). For a real-valued function \( f \) on \( \mathcal{X} \), let \( \delta_x \) be the evaluation functional at \( x \in \mathcal{X} \), that is, \( \delta_x[f] = f(x) \). Using a notation similar to Schaback (1999, Sect. 5), this kernel substitution amounts to replacing the evaluation functional \( \delta_x \) by the functional \( \delta(x) \) defined by

\[
\forall x \in \mathcal{X}, f \in L^2(\mathcal{X}, \mu), \ \ \delta(x)[f] = af(x).
\]

More precisely, for \( g \in L^2(\mathcal{X}, \mu) \), let \( I_{\mathcal{X}, \mu} \) denote the functional defined by \( I_{\mathcal{X}, \mu}[f] = (f, g)_{L^2} \). Using vector–matrix notation, we get, for \( x \in \mathcal{X} \), \( \delta_x = g^T \mathbf{M}_g^{-1} I_{\mathcal{X}, \mu} \), and we can write \( \delta(x)[f] = \delta_x[qf] = (Iq\delta_g)[f] \). Notice that this kernel substitution is defined whenever \( \mathcal{X} \subset L^2(\mathcal{X}, \mu) \) and the assumptions of Section 2.2 are verified. In particular, it covers the case of a general linear trend \( g(x) \theta \) in (20), whereas the theory of intrinsic random functions concerns translation-invariant kernels and only addresses the case of polynomial regression, see Matheron (1971).

The integral operator associated with \( K^q(\cdot, \cdot) \) is

\[
T^q_\mu[f](x) = \int_{\mathcal{X}} f(t)K^q(x, t) \, d\mu(t),
\]

with \( f \in L^2(\mathcal{X}, \mu) \) and \( x \in \mathcal{X} \). By construction, it satisfies \( T^q_\mu[g] = 0 \) for all \( j \in \{1, \ldots, p\} \). Denote by \( \{\lambda^q_k[k \in \mathbb{N}_+^\delta] \} \) the set of all strictly positive eigenvalues of \( T^q_\mu \) and let \( \phi^q_k \) be their (canonically extended) associated eigenfunctions. We have

\[
\tau^q = \int_{\mathcal{X}} K^q(x, x) \, d\mu(x) = \tau - \text{trace} (\mathbf{M}_g^{-1}(T^q_\mu | g | g^T))_{L^2} \leq \tau.
\]  

For a truncation subset \( \mathcal{X}_{\text{tr}} \), using the same notations as in Section 5.2, the terms \( D^\theta \) and \( D^\theta_{\text{tr}} \) now equal 0, and inequalities similar to (5) are available, with \( \tau^\theta_{\text{tr}} = \sum_k k^2 \lambda^q_k \) quantifying the error due to spectral truncation.

Remark 5. Comparing with (29), we now have \( (\phi^q_{\text{tr}} | g^T)_{L^2} = 0 \). In order to further reduce the computational cost when using the reduced kernel \( K^q(\cdot, \cdot) \), one may consider trend functions \( g \) that form an orthonormal basis of \( \mathcal{S} \) for \( L^2(\mathcal{X}, \mu) \), making \( \mathbf{U}^\theta \) equal to the identity matrix.

Remark 6. Starting from the prior (21), we obtain

\[
E(\theta^\delta) = \theta_0 \quad \text{and} \quad \text{Cov}(\theta^\delta) = \mathbf{A}^\delta = \mathbf{A} + \mathbf{S}.
\]

When the prior is non-informative, that is, roughly speaking, when \( \mathbf{A}^{-1} = 0 \), then the same holds for \( \mathbf{A}^\delta \) and the prior on \( \theta^\delta \) remains non-informative.

The situation would be different in the presence of an informative prior: in that case, orthogonality in \( L^2(\mathcal{X}, \mu) \) between the components of \( \theta^\delta \) and the r.v. \( qz \), \( x \in \mathcal{X} \), is lost when using the model (31), and \( E(\theta^\delta(qz)) = b(x) - Sg(x) = \mathbf{M}_g^{-1}(T^q_{\mu} | g | g^T)(x) \). The consequence on the IMSE calculation is that the matrix \( \mathbf{F}^\delta \) corresponding to \( \mathbf{F} \) in (25) is no longer block diagonal, with the two off-diagonal blocks being not trivial to evaluate.

6. Numerical approach and relaxation-based design strategy

In Section 6.1 we discuss the implementation of the convex-relaxation method presented in Sections 3 and 4, assuming that a pointwise quadrature is used to approximate the integral of the MSE and restricting the design space to quadrature points (a similar approach could be used for any design region formed by a finite set of points). We then present (Section 6.2) a heuristic for the extraction of an exact design from an optimal measure obtained by minimising the criterion \( \psi_\delta(\cdot) \) defined in (16).

6.1. Quadrature approximation and quadrature-restricted continuous design

Assume that the measure \( \mu \) has the following form (quadrature approximation)

\[
\mu = \sum_{j=1}^{N_q} \omega_j \delta_{s_j},
\]  

with \( N_q \) quadrature points \( s_j \in \mathcal{X} \) receiving weights \( \omega_j > 0 \). We introduce the two \( N_q \times N_q \) matrices \( \mathbf{W} = \text{diag}(\omega_1, \ldots, \omega_{N_q}) \) and \( \mathbf{Q} \) with \( i, j \) term \( Q_{ij} = K(s_i, s_j) \), for \( 1 \leq i, j \leq N_q \); \( \mathbf{W} \) is thus the diagonal matrix of quadrature weights and \( \mathbf{Q} \) is the covariance matrix for quadrature points.

Consider the spectral decomposition of the matrix \( \mathbf{Q} \mathbf{W} \) in the Hilbert space \( \mathbb{R}^{N_q} \) endowed with the inner product \( \langle \cdot, \cdot \rangle_w \). with, for \( \mathbf{x} \) and \( \mathbf{y} \in \mathbb{R}^{N_q} \), \( \langle \mathbf{x}, \mathbf{y} \rangle_w = \mathbf{x}^T \mathbf{W} \mathbf{y} \), see Gauthier and Pronzato (2014, 2016a) for more details. We denote by
λ_1 ≥ λ_2 ≥ ⋯ ≥ λ_{N_q} ≥ 0 the eigenvalues of the matrix QW and by v_1, . . ., v_{N_q} the corresponding eigenvectors, i.e., QW = PΛP^{-1} with Λ = diag(λ_1, . . ., λ_{N_q}) and P = (v_1| ⋯ |v_{N_q}). Then, \{v_1, . . ., v_{N_q}\} forms an orthonormal basis of \mathbb{R}^{N_q} for the inner product (\cdot|\cdot)_{W}, so that P^TWP = I_{N_q}, the \(N_q\)-dimensional identity matrix.

Let P \subseteq \mathcal{P} denote the subset of all probability measures (dominated by \(\mu\)) of the form
\[
v = \sum_{j=1}^{N_q} p_j \delta_j,
\]
with \(p_j \geq 0\) and \(\sum_{j=1}^{N_q} p_j = 1\). For a truncation subset \(I_{trc}\), the matrix \(M_v\) given by (15) and associated with a measure \(v \in \mathcal{P}\) is then given by
\[
M_v = (P_{\cdot|I_{trc}})^T \Omega (P_{\cdot,|I_{trc}}),
\]
where \(\Omega = diag(p_1/\sigma^2(s_1), . . ., p_{N_q}/\sigma^2(s_{N_q}))\) and \(P_{\cdot,|I_{trc}}\) stands for the \(N_q \times n_{trc}\) matrix consisting of the columns of \(P\) with index in \(I_{trc}\). The various matrices used in Section 5 can be computed in a similar way. We shall only consider truncation subsets \(I_{trc}\) that correspond to the \(n_{trc}\) largest eigenvalues \(\lambda_k\) (eigenvectors associated with multiple eigenvalues being ordered randomly).

6.2. Exact-design extraction

Suppose that an optimal design measure \(\nu^*\) minimising \(\Psi_{\alpha}(\cdot)\) has been determined. We must still define a procedure for extracting an exact design from \(\nu^*\). This issue differs significantly from the usual rounding problem in approximate design theory (see, e.g., Fedorov, 1972, p. 157; Pukelsheim and Reider, 1992), since here we request exactly one observation per support point, with the key consequence that the size \(n\) of a design \(D_n\) that can be “naturally” extracted from a given optimal measure \(\nu^*\) is severely constrained.

The usual direct approach that consists in setting \(n\) and then constructing an optimal exact design of size \(n\) minimising IMSE\((D_n)\) faces the difficulties mentioned in introduction. On the other hand, using the branch-and-bound method of Welch (1982) to minimise \(\Psi_{\alpha}(\cdot)\) under the constraints that all \(p_j\) equal 0 or 1/\(n\) in (35) seems computationally cumbersome. This is why we suggest not to specify \(n\) precisely, but to use the scalar \(\sigma\) and truncation level \(n_{trc}\) as tuning parameters for the generation of optimal measures adapted to the extraction of exact designs with size close to the desired value. A greedy procedure for merging support points of an optimal measure \(\nu^*\) associated with given values of \(\alpha\) and \(n_{trc}\) is presented in Section 6.2.1; Section 6.2.2 proposes a heuristic fixed-point method for setting the value of \(\alpha\), the size of the exact design being finally controlled by the truncation level \(n_{trc}\) only.

6.2.1. Merging support points

Let \(\nu^* \in \mathcal{P}\) be a design measure minimising \(\Psi_{\alpha}(\cdot)\) for given values of \(\alpha\) and \(n_{trc}\), with \(m\) support points. As it will be illustrated in Section 7, the support of \(\nu^*\) is generally formed by (a) isolated points with significative weight, (b) isolated clusters of points having altogether significative weight, and (c) isolated residual points (or clusters) with negligible weight. The merging strategy presented below (Algorithm 1) iteratively reduces the size of the support of \(\nu^*\) in a greedy way, while trying to keep the value of \(\Psi_{\alpha}\) as low as possible; the objective is to preserve isolated points (a), merge each cluster (b) into a single point, and remove isolated points (c). It is rather straightforward to implement, has low complexity (of order \(\mathcal{O}(m^3)\)) and proved rather efficient on the numerical tests that we performed. A few examples are given in Section 7.

If \(v = \sum_{j=1}^{n_{trc}} p_j \delta_{i_j}\) is the current design measure, supported by \(n_{trc}\) quadrature points \(s_{i_j}\), with \(i_j \in \{1, . . ., N_q\}\) and \(p_j > 0\) for all \(j\) with \(\sum_{j=1}^{n_{trc}} p_j = 1\), the algorithm considers all \(n_{trc}(n_{trc}-1)/2\) possible measures \(v_{(a\rightarrow b)}\) obtained by transferring the weight \(p_a\) from \(s_{i_a}\) to \(s_{i_b}\), with \(a\) and \(b\) \(\in\{1, . . ., n_{trc}\}\) and \(a \neq b\). Note that \(v_{(a\rightarrow b)}\) is supported by \(n_{trc} - 1\) points, \(s_{i_b}\) being removed from the support of \(v\). The measure \(v_{(a\rightarrow b)}\) with smallest value of \(\Psi_{\alpha}\) is then carried forward to the next iteration. The measure \(v^{(k)}\) obtained after \(k\) iterations initialised at \(\nu^*\) has \(m - k\) support points.

When the model used contains an unknown parametric trend without informative prior, we need to keep \(q = p\) support points at least; otherwise (model without trend, or approach of Section 5.3), we only need to keep \(q = 1\) point. We shall thus perform no more than \(m - q\) merging iterations, since \(v^{(m-q)}\) has \(q\) support points only. We denote \(\psi_{k} = \Psi_{\alpha}(v^{(k)})\), for \(k = 0, . . ., m-q\), and \(k = \psi_k - \psi_{k-1}\) (with \(\psi_0 = 0\)).

The sequence \(\{\psi_{k}\}\) is generally non-decreasing and the measure \(v^{(k)}\), with \(m - k\) support points, obtained after \(k\) iterations satisfies \(\Psi_{\alpha}(v^{(k)}) \leq \Psi_{\alpha}(v^{(k+1)})\) (due to the fact that \(v^{(k)}\) is optimal for \(\Psi_{\alpha}(\cdot)\)). Usually, \(\psi_k\) remains almost constant for a few iterations, and then abruptly increases, see Figs. 4–6 in Section 7. The value \(n_{ext}^{[\alpha]}\) returned by Algorithm 1, with the superscript \(ext\) standing for \(extracted\), is the number of support points just before the first significant “jump” of \(\psi_k\), detected with the threshold parameter \(T\). The value \(T = (\psi_{m-\bar{q}} - \psi_{m-q})/10(m-q)\) gave satisfactory results for all examples in Section 7. Algorithm 1 also returns the measure \(v^{[\alpha]} = v^{(m-n_{ext}^{[\alpha]})}\), and we shall denote by \(D_{\alpha}^{ext}\) the exact design given by the \(n_{ext}^{[\alpha]}\) support points of \(v^{[\alpha]}\).
Remark 7. Alternative methods could also be considered for merging support points of a measure \( \nu \). For instance, Gauthier and Pronzato (2016b) aggregate support points via minimum-spanning-tree clustering, with the metric \( \Delta (x, x') = K(x, x') + K(x', x') - 2K(x, x') \) induced by \( K \) (the design points are given by the weighed centroids of the clusters, with weights given by the \( p_k \)). Notice that this approach requires the computation of canonical extensions (1), i.e., of summations over the \( N_q \) quadrature points, since in general the cluster centroids are not quadrature-points.

6.2.2. Design size related to a truncation level

For a given truncation level \( n_{\text{trc}} \), our objective consists in tuning \( \alpha \) such that \( n_{\text{ext}}^{[\alpha]} = \alpha \), in order to have \( \Psi_{\alpha}(D_{\text{ext}}) = \text{IMSE}_{n_{\text{trc}}}^\Sigma(D_{\text{ext}}^{[\alpha]}) \), see Section 4. We shall denote by \( \alpha_{n_{\text{trc}}} \) such a value of \( \alpha \); \( \alpha_{n_{\text{trc}}} \) can be considered as a suitable design size for the truncation level \( n_{\text{trc}} \) and can be obtained by a fixed-point algorithm (Algorithm 2).

Algorithm 2 Fixed-point iterations to compute \( \alpha_{n_{\text{trc}}} \)

Require: truncation level \( n_{\text{trc}} \) and initial value \( \alpha > 0 \)

1: Compute an optimal measure \( \nu^* \) for \( \Psi_{\alpha} \) and use Algorithm 1 to obtain \( n_{\text{ext}}^{[\alpha]} \) and \( \nu_{\text{ext}}^{[\alpha]} \).
2: while \( \alpha \neq n_{\text{ext}}^{[\alpha]} \) do
3: \( \alpha \leftarrow n_{\text{ext}}^{[\alpha]} \).
4: Compute an optimal measure \( \nu^* \) for \( \Psi_{\alpha} \) and use Algorithm 1 to obtain \( n_{\text{ext}}^{[\alpha]} \) and \( \nu_{\text{ext}}^{[\alpha]} \).
5: end while
6: return \( \alpha_{n_{\text{trc}}} = n_{\text{ext}}^{[\alpha]} \) and \( \nu_{\text{ext}}^{[\alpha]} \).

In all the examples we have considered, the algorithm always converged after a very small number of iterations, see, e.g., Table 1 in Section 7.1. When it is initialised at \( \alpha \) equal to the number of trend functions involved in the BLM (i.e., \( n_{\text{trc}} \) or \( n_{\text{trc}} + p \), depending on the model considered), it generally converges after two iterations only, see for instance Table 2 in the same section. Notice that the optimal measure obtained for a certain \( \alpha \) may be used as warm start for the next optimisation.

For a given \( n_{\text{trc}} \), let \( \nu_{\text{ext}}^{[\alpha_{n_{\text{trc}}}]} \) denote an optimal measure minimising \( \Psi_{\alpha_{n_{\text{trc}}}}(\cdot) \), with \( \alpha_{n_{\text{trc}}} \) the value returned by Algorithm 2. We can expect the design \( D_{\alpha_{n_{\text{trc}}}}^{\text{ext}} \) to have a high truncated-IMSE efficiency, which can be evaluated by computing the extraction efficiency

\[
\rho_{n_{\text{trc}}} = \frac{\Psi_{\alpha_{n_{\text{trc}}}}(\nu_{\text{ext}}^{[\alpha_{n_{\text{trc}}}])}}{\text{IMSE}_{n_{\text{trc}}}^\Sigma(D_{\alpha_{n_{\text{trc}}}}^{\text{ext}})}.
\]

It satisfies \( 0 \leq \rho_{n_{\text{trc}}} \leq 1 \), with high values of \( \rho_{n_{\text{trc}}} \) indicating that \( D_{\alpha_{n_{\text{trc}}}}^{\text{ext}} \) is nearly IMSE-optimal for the approximate BLM.

Remark 8. The extracted design \( D_{\alpha_{n_{\text{trc}}}}^{\text{ext}} \) can also be used to initialise a discrete local minimisation of the IMSE over the quadrature points, see the algorithm in Gauthier and Pronzato (2016a); this generally yields designs with very high IMSE efficiency, as illustrated by the examples in Section 7.

When \( \mathcal{X} \) is a compact subset of \( \mathbb{R}^d \), a continuous local minimisation of the IMSE with respect to \( D_{\psi} \in \mathbb{R}^{n \times d} \) (i.e., with respect to \( n \times d \) variables), initialised at \( D_{\alpha_{n_{\text{trc}}}}^{\text{ext}} \), can also be performed at reasonable computational cost, using a standard optimisation algorithm (the canonical extensions (1) must be used, similarly to Remark 7). For most covariance kernels the optimal design points lie in the convex hull of \( \mathcal{X} \), and no constraint need to be taken into account if \( \mathcal{X} \) is convex. However, the decrease of IMSE compared to \( \nu_{\text{ext}}^{[\alpha_{n_{\text{trc}}}]} \) is usually marginal when the set of quadrature points is dense enough, and we shall not consider this procedure any further in the paper. \( \Delta \)
7. Numerical experiments

We shall consider RF models on $\mathcal{X} = [0, 1]^d$, with kernel $K(x, y) = \prod_{i=1}^d K_i(x_i, y_i)$, where $x = (x_1, \ldots, x_d)$ and $K_i(x_i, y_i) = (1 + \sqrt{3}|x_i - y_i|/\ell) \exp(-\sqrt{3}|x_i - y_i|/\ell)$, $\ell > 0$ (Matérn 3/2). We use $d = 2$ in Sections 7.1, 7.2, 7.4 and 7.5, and $d = 4$ in Section 7.3.

In all of Section 7, the measures $\mu$ considered correspond to pointwise quadrature approximations of the uniform probability on $\mathcal{X} = [0, 1]^d$, and we apply the methodology described in Section 6. Optimal design measures $\nu^*$ on the quadrature points are obtained by a vertex-exchange algorithm, see Böhning (1985, 1986). The minimisation of $\Psi_\alpha(\cdot)$ can be accelerated by removing points $s_i$ that cannot be support points of an optimal measure from the search space, see Pronzato (2013). The iterations are stopped when the directional derivative (17) for the current approximated solution $\hat{\nu}$ satisfies

$$
\min_{j \in \{1, \ldots, N_q\}} F_{\psi_\alpha}(\hat{\nu}, \delta_j) + \epsilon \geq 0.
$$

which, by convexity, ensures that $\psi_\alpha(\hat{\nu}) \leq \psi_\alpha(\nu^*) + \epsilon$, see (19). With a slight abuse of terminology, we refer to the obtained measures as the “optimal measures” (and we denote them by $\nu^*$).

The extraction efficiency is expressed through the ratio $\delta_{n_{rec}}$ given by (37), and the IMSE efficiency of a design $D_n$ is measured by the ratio $\text{IMSE}(D_n^{ext})/\text{IMSE}(D_n)$, where $D_n^{ext}$ is the best $n$-point quadrature-design (i.e., a design only composed of quadrature-points) that we were able to obtain using the simulated-annealing algorithm presented in Gauthier and Pronzato (2016a). Notice that, although being rather efficient, this global optimisation is much more time consuming than the convex-relaxation considered here. All computations have been performed with the programming language R, interfaced with programs in language C for the convex optimisation algorithm and Algorithm 1.

The model is without trend in Sections 7.1, 7.2, 7.3 and 7.5; the presence of a linear parametric trend is considered in Section 7.4.

7.1. Regular grid approximation

Here $\mu$ is the discrete probability measure on $\mathcal{X} = [0, 1]^2$ defined by a $33 \times 33$ regular square grid (midpoint rule), each grid-point receiving the same weight $1/N_q$, with $N_q = 33^2 = 1089$. We take $\ell = 0.15$ in $K_i(\cdot, \cdot)$.

Fig. 1 shows the optimal measures for the heteroscedastic and homoscedastic models, respectively with $\sigma^2(x) = K_{err}(x, x)$ (left) and $\sigma^2 = \tau_{err}/\mu(\mathcal{X})$ (right), for $\alpha = n_{rec} = 7$ (so that $\tau_{rec} \approx 0.4484$). The optimal measure for the heteroscedastic model is supported by 11 points, but 97.71% of the mass is supported by 7 points only. The design $D_{7,\text{het}}^{\text{ext}}$ obtained by Algorithm 1 corresponds to those 7 points, with an extraction efficiency $\delta_{n_{rec}} = 99.94\%$ (note that $\alpha_{n_{rec}} = n_{rec} = 7$ for this particular case). In addition, the IMSE efficiency of $D_{7,\text{het}}^{\text{ext}}$ equals 99.84%, and a grid-restricted local descent starting from $D_{7,\text{het}}^{\text{ext}}$ yields $D_7^\ast$.

The optimal measure for the homoscedastic model is supported by 15 points, 7 of which carrying 79.39% of the total mass; the design $D_{15,\text{hom}}^{\text{ext}}$ has an extraction efficiency of 98.90%, its IMSE efficiency equals 99.59%; a local descent initialised at $D_{7,\text{hom}}^{\text{ext}}$ yields $D_7^\ast$. On the other hand, Algorithm 1 suggests the extraction of a design with $n_{\text{ext}}^{[\alpha]} = 9$ points (which carry 90.67% of the mass); the extraction efficiency for $\alpha = 9$ is 98.67%.

As illustrated by this first example, the optimal measures obtained with the homoscedastic model tend to have more support points than those obtained with the heteroscedastic model, or present clusters of points that are more spread apart, which often complicates the extraction of an exact design. Only the heteroscedastic model is used in the following.

Fig. 2 shows the optimal measures $\nu^*$ obtained for $n_{\text{rec}} = 15$ and different values of $\alpha$: for a fixed truncation level, we generally observe that the number of support points (or clusters of support points) of $\nu^*$ tends to increase when $\alpha$ increases (the larger $\alpha$, the more scattered the support of $\nu^*$). For $\alpha = 0.001$ (not shown), $\nu^*$ is reduced to a Dirac measure at the

![Fig. 1. Contour-plot of the variance $x \rightarrow \sigma^2(x) = K_{err}(x, x)$ ($\alpha = n_{rec} = 7$) and optimal measure $\nu^*$ for $\Psi_{\alpha}(\cdot)$ (disks with surface proportional to the weights $p_k, \epsilon = 10^{-7}$ in (38)) for the heteroscedastic (left) and homoscedastic (right) models. The quadrature points are indicated by grey crosses, the IMSE-optimal 7-point quadrature-design is indicated by triangles.](image-url)
Fig. 2. Optimal measures ($\varepsilon = 10^{-7}$) for the heteroscedastic model with $n_{trc} = 15$ and $\alpha = 0.08, 1, 15$ and 1000. A contour-plot of the variance $x \mapsto \sigma^2(x) = K_{trc}(x,x)$ is also given (top-left).

Table 1
Algorithm 2 (evolution of $\alpha$) for the example discussed in Fig. 2 ($n_{trc} = 15$, different initialisations).

<table>
<thead>
<tr>
<th>Initialisation</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Iteration 3</th>
<th>Iteration 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.001$</td>
<td>1</td>
<td>13</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>$\alpha = 0.08$</td>
<td>5</td>
<td>13</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td>13</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>$\alpha = n_{trc} = 15$</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>$\alpha = 1000$</td>
<td>15</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

centre of the grid; for $\alpha = 0.08$, $\nu^*$ is supported by 5 points. Remarkably, for $\alpha = 1$ the support of $\nu^*$ coincides with the 13-point optimal quadrature-design $D^*_1$. The optimal measure for $\alpha = n_{trc} = 15$ is supported by 29 points, and Algorithm 1 then suggests the extraction of a 21-point design $D^*_{21}$ (these 21 points carrying 96.75% of the mass of $\nu^*$); the optimal measure for $\alpha = 1000$ is supported by 37 points. Algorithm 2 yields $\alpha_{n_{trc}} = 21$, with $D^*_{n_{trc}} = D^*_{21}$. Although this design is quite different from $D^*_{21}$, see Fig. 2-bottom-left, its IMSE efficiency is 96.47% and a grid-restricted local descent initialised at $D^*_{21}$ yields a design with IMSE efficiency 99.62%.

Table 1 shows the number of iterations required for the convergence of Algorithm 2 when initialised at the values $\alpha$ considered above ($\alpha = 0.001$ and those in Fig. 2): the algorithm converges to $\alpha_{n_{trc}} = 21$ in a few iterations only, whatever the initial $\alpha$.

Fig. 3 presents the optimal measures obtained for $\alpha = 24$ and two different values of $n_{trc}$. For $n_{trc} = 22$ ($\tau_{trc} \approx 0.7648$), $\nu^*$ is supported by 52 points and is presented in Fig. 3-left. The behaviour of the design-extraction procedure (Algorithm 1) is illustrated in Fig. 4-left. The evolution of $k_0$ suggests the extraction of a design $D^*_{24}$ with size $n = 24$, see Fig. 4-right; its IMSE efficiency is 99.37% and a quadrature-restricted local descent initialised at $D^*_{24}$ yields $D^*_{24}$. Algorithm 2 yields $\alpha_{n_{trc}} = 24$ in two iterations, the extraction efficiency is 99.45%. The optimal measure for $\alpha = 24$ and $n_{trc} = 100$ ($\tau_{trc} \approx 0.9711$) is supported by 276 points, see Fig. 3-right, and its exploitation for the extraction of a design with size 24 seems difficult. This motivates the recommendation to control the size of the extracted design through the truncation level.

Table 2 gives an overview of the design extractions obtained for values of $n_{trc}$ between 1 and 24. We can first remark that when $\alpha = n_{trc}$, for all values of $n_{trc}$ considered Algorithm 1 yields a value $n_{\text{ext}}[n_{trc}]$ that coincides with $\alpha_{n_{trc}}$, indicating that Algorithm 2 converges in 2 iterations at most (it even converges in one iteration for $n_{trc} = 1, 2, 4, 7$ and 12). We can also observe that Algorithm 2 yields designs with high extraction efficiency. Finally, the truncation level $n_{trc}$ permits to conveniently control the size $n$ of the exact design generated, even if not all design sizes can be obtained via Algorithm 2 (in the range $1 \leq n \leq 24$, the values $n = 3, 5, 6, 11, 13, 14, 15, 17, 20, 22$ and 23 are missing).
Fig. 3. Optimal measures (\(\epsilon = 10^{-7}\)) for the heteroscedastic model with \(\alpha = 24\) and \(n_{trc} = 22\) (left), \(n_{trc} = 100\) (right).

Fig. 4. Merging procedure for \(\nu^*\) obtained with \(n_{trc} = 22\) and \(\alpha = 24\) (heteroscedastic model, \(\nu^*\) is shown on Fig. 3-left): sequence \(\{I_k\}_k\) (left) and measure \(\nu^*\) obtained after 28 iterations of Algorithm 1 (right); the 24-point IMSE-optimal quadrature-design \(D^*_24\) is also presented.

Table 2

<table>
<thead>
<tr>
<th>(n_{trc})</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>20</td>
<td>12</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>(m_{opt})</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>(\alpha_{n_{trc}})</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>(\alpha_{opt})</td>
<td>100</td>
<td>100</td>
<td>99.99</td>
<td>100</td>
<td>99.86</td>
<td>98.82</td>
<td>99.94</td>
<td>99.71</td>
<td>99.61</td>
<td>97.80</td>
<td>99.58</td>
<td>99.83</td>
</tr>
</tbody>
</table>

7.2. Quasi-Monte-Carlo quadrature

In order to illustrate the impact of the regularity of the quadrature on optimal measures, \(\mu\) corresponds now to a quadrature consisting of the \(N_q = 1089\) first points of a low-discrepancy Halton sequence in \([0, 1]^2\), all points receiving identical weights \(1/N_q\).

For the truncation level \(n_{trc} = 22\), Algorithm 2 gives \(\alpha_{n_{trc}} = 24\). Fig. 5-left shows the optimal measure \(\nu^*\) obtained for the heteroscedastic model with \(\alpha = 24\). There are 49 support points and \(\nu^*\) presents some similarities with the measure obtained with a regular square-grid for the same parameters, see Fig. 3-left. However, \(\nu^*\) is now more irregular, due to the uneven distribution of quadrature points (grey crosses) in the Halton sequence (note for instance the presence of neighbouring points with non-negligible weights in the rectangular box on the top of the figure), which might potentially confuse the design extraction (and in particular the estimation of the size of the design to extract). Nevertheless, the sequence \(\{I_k\}_k\) obtained with Algorithm 1, see Fig. 5-centre, suggests the extraction of a design with size \(n = 24\), see
Fig. 5. Optimal measure ($\epsilon = 10^{-7}$) for the heteroscedastic model with $n_{trc} = 22$ and $\alpha = 24$ for a quadrature based on a low-discrepancy Halton sequence (left); sequence $\{l_k\}_{k}$ (middle); measure $\nu^{(25)}$ obtained after 25 iterations of Algorithm 1, with $n_{ext} = \alpha = 24$ support points (right). The 24-point IMSE-optimal quadrature-design $D_{24}^*$ is also presented.

Fig. 6. Projection on $\{x_1, x_2\}$ of the optimal measure ($\epsilon = 10^{-7}$) for the heteroscedastic model with $n_{trc} = 31$ and $\alpha = 40$ (left), sequence $\{l_k\}_{k}$ (middle) and projection on $\{x_1, x_2\}$ of the measure $\nu^{(16)}$ (supported by 40 points) obtained after 16 iterations of Algorithm 1 (right).

Fig. 5-right, with $\epsilon_{n_{trc}} = 99.01\%$ and IMSE$(D_{24}^*)$/IMSE$(D_{24}^{ext}) = 98.40\%$; a quadrature-restricted local descent initialised at $D_{24}^{ext}$ yields $D_{24}^*$.

7.3. Example in dimension $d = 4$

We take $d = 4$, with $\ell = 0.35$ in $K_{\ell}(\cdot, \cdot)$. The measure $\mu$ corresponds to a $9 \times 9 \times 9 \times 9$ square grid (midpoint rule), all points receiving the same weight $1/N_q$, with $N_q = 9^d = 6,561$; the heteroscedastic model is used.

For $n_{trc} = 31$ ($t_{trc} \approx 0.6658$), Algorithm 2 suggests the extraction of a 40-point design, and Fig. 6 illustrates the results obtained with Algorithm 1 for $\alpha = 40$ (i.e., $\alpha_{n_{trc}}$). The optimal measure $\nu^*$ has 56 support points, 40 of which carry 95.88% of the mass. The sequence $\{l_k\}_{k}$ suggests the extraction of a design with size $n = 40$. The extraction efficiency $\epsilon_{n_{trc}}$ is 99.90% and, remarkably enough, $D_{40}^{supp} = D_{40}^*$.

7.4. Models with unknown parametric trend

We consider the same framework as Section 7.1 but now assume the presence of a parametric trend, with $g(x) = (g_1(x), g_2(x), g_3(x))^T = (1, x_1, x_2)^T$ for $x = (x_1, x_2) \in [0, 1]^2$ ($p = 3$). We take $\text{Cov}(\theta) = A = \text{Id}_3$ when an informative prior on the trend parameters is needed.

Fig. 7-left illustrates the strong linear relationship that exists, in $L^2(\mathcal{X}, \mu)$, between the three trend functions $g_j$ and the eigenfunctions of the IMSE integral operator $T_\mu$. This relationship can be interpreted as a redundancy of the trend functions in the model (20) or (24).
which indicates that the three extracted designs have relatively close IMSE scores. For the design being slightly shrunk for \( \ell \) kernel parameters on the optimal design. We consider two values for \( \ell \) kernel. Algorithm 1 with \( \alpha \) function only. Also, kernel requires 25 regression functions to carry as much information as the BLM induced by the reduced kernel with 22

where \([n_{\text{trc}}] \) and \([n_{\text{trc}} + p] \) indicate the number of eigenfunctions considered (we recall that \( \tau_{\text{err}} = \int g \cdot K_{\text{err}}(x, x) \text{d}\mu(x) \)).

Fig. 8 gives a further illustration that, for the same number of regression functions considered, the BLM induced by the reduced kernel contains more information than the BLM induced by the initial kernel. The optimal measures \( \nu^* \) for the heteroscedastic models are presented, on the left for the initial kernel with \( n_{\text{trc}} = 22 \) and \( \alpha = 24 \) (and therefore 25 regression functions), and on the right for the reduced kernel with \( n_{\text{trc}} = 19 \) and \( \alpha = 24 \) (22 functions, no prior on \( \theta \) is used). The two measures look rather similar, and the contour plots of the variances \( x \mapsto K_{\text{err}}(x, x) \) and \( x \mapsto K_{\text{err}}^* (x, x) \) also present strong similitude (with \( \tau_{\text{err}} = 0.2352 \) and \( \tau_{\text{err}}^* = 0.2370 \)). We thus observe that the BLM induced by the initial kernel requires 25 regression functions to carry as much information as the BLM induced by the reduced kernel with 22 functions only. Also, \( \nu^* \) is supported by 56 points for the initial kernel, and has only 48 support points for the reduced kernel. Algorithm 1 with \( \alpha = 24 \) yields the same 24-point design \( D_{24}^{\text{ext}} \) in both cases, with \( \epsilon_{\text{trc}} = 99.25\% \) for the initial kernel and \( \epsilon_{\text{trc}} = 98.96\% \) for the reduced kernel; a local descent initialised at \( D_{24}^{\text{ext}} \) yields \( D_{24}^{\text{ext}} \).

For a given truncation level \( n_{\text{trc}} \), the total number of regression functions \( n_{\reg} \) in the BLM induced by the initial and reduced kernels is \( n_{\reg} = n_{\text{trc}} + p \), this corresponds to models of type (24). We have \( n_{\reg} = n_{\text{trc}} \) for the model, of type (6), induced by the augmented kernel. The middle and right parts of Fig. 7 aim at comparing the integrated variances of the error terms for the BLMs defined by the initial and modified kernels: initial kernel versus reduced kernel (middle), and initial kernel versus augmented kernel (right). We observe that for the same number of regression functions \( n_{\reg} \), the reduced and augmented kernels yield BLMs that are more accurate than the BLM induced by the initial kernel, in the sense that we have, for this particular example,

\[
\tau_{\text{err}}[n_{\text{trc}}] \leq \tau_{\text{err}}[n_{\text{trc}} + p] \leq \tau_{\text{err}}[n_{\text{trc}}],
\]

where \([n_{\text{trc}}] \) and \([n_{\text{trc}} + p] \) indicate the number of eigenfunctions considered (we recall that \( \tau_{\text{err}} = \int g \cdot K_{\text{err}}(x, x) \text{d}\mu(x) \)).

7.5. Impact of the kernel parameters

We consider again the framework of Section 7.1, with the heteroscedastic model, and illustrate now the impact of the kernel parameters on the optimal design. We consider two values for \( \ell \) in \( K(x, y) \), \( \ell = 0.075 \) and \( \ell = 0.3 \). We take \( n_{\text{trc}} = 7 \), and Algorithm 2 yields \( \alpha_{\text{trc}} = 7 \) for both value of \( \ell \). Fig. 9 shows the optimal measures \( \nu^* \) obtained with \( n_{\text{trc}} = \alpha = 7 \). For \( \ell = 0.075 \), we have \( \tau_{\text{trc}} = 0.17 \) and the extraction efficiency \( \epsilon_{\text{trc}} \) is 99.95%; for \( \ell = 0.3 \), we have \( \tau_{\text{trc}} = 0.7796 \) and \( \epsilon_{\text{trc}} = 97.76\% \). The extracted 7-point designs are much similar to the one obtained for \( \ell = 0.15 \) (also shown in the figure), the design being slightly shrunk for \( \ell = 0.075 \) and slightly stretched for \( \ell = 0.3 \). This observation is confirmed by Table 3, which indicates that the three extracted designs have relatively close IMSE scores.
Fig. 9. Optimal measures for the heteroscedastic models \((\epsilon = 10^{-7})\) with \(n_{\text{trc}} = \alpha = 7\) for \(\ell = 0.075\) (left) and \(\ell = 0.3\) (right); the design \(D^*\) extracted for \(\ell = 0.15\) is presented for comparison.

### Table 3

<table>
<thead>
<tr>
<th>(D^*) for (\ell = 0.075)</th>
<th>(D^*) for (\ell = 0.15)</th>
<th>(D^*) for (\ell = 0.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMSE for (\ell = 0.075)</td>
<td>0.91786</td>
<td>0.91794</td>
</tr>
<tr>
<td>IMSE for (\ell = 0.15)</td>
<td>0.69907</td>
<td>0.69763</td>
</tr>
<tr>
<td>IMSE for (\ell = 0.3)</td>
<td>0.32474</td>
<td>0.31136</td>
</tr>
</tbody>
</table>

8. Concluding discussion

We have shown that the convex relaxation method of Fedorov (1996) and Spöck and Pilz (2010) can be efficiently applied when considering the IMSE-induced Karhunen–Loève expansion of the RF.

We have proposed two extensions to the direct approach of Spöck and Pilz (2010) for RF models that include a linear parametric trend. One is based on kernel augmentation, and applies to the case where an informative prior on the two first moments of the trend parameters \(\theta\) is available. The other amounts to a kernel reduction defined from a linear continuous projection onto the trend space, and concerns the case when no such prior information on \(\theta\) is available. They both permit to bound the error induced by considering the truncated-IMSE instead of the true IMSE, whereas the derivation of such error bounds seems much more difficult with the initial kernel. The numerical experiments carried out in Section 7.4 also point out that, for an equivalent number of regression functions, the modified kernels generally lead to BLMs having smaller errors than the BLMs induced by the initial kernel. Theorem 2 shows the equivalence between the predictors for the initial model and those for a model based on the reduced kernel in absence of informative prior on \(\theta\). This is a very general result (valid in fact for any linear continuous projection onto the trend space), with potential consequences in other contexts involving RF models.

A numerical implementation of the approach has been proposed, based on a quadrature approximation of the IMSE and a discretisation of the design region. Efficient convex-programming algorithms can then be used to construct optimal design measures, with guaranteed convergence to the optimum. A greedy exchange algorithm (Algorithm 1) has been presented for the extraction of an exact design from an optimal measure \(\nu^*\). When combined with a fixed-point algorithm (Algorithm 2), it permits to control the size of the extracted exact design by tuning the truncation level of the KL expansion. We have observed that the heteroscedastic approximate BLM, with \(\sigma^2(x) = K_{\text{err}}(x, x)\), often yields optimal measures that are easier to exploit than those obtained with the homoscedastic model for the extraction of an exact design: in some sense, for the heteroscedastic model the presence of regions with low variance helps to identify important sampling areas, and thereby favours the aggregation of support points.

To summarise, the proposed methodology, based on convex optimisation, permits to generate designs of almost arbitrary size and with good IMSE performance, at low computational cost. These designs can straightforwardly be used to initialise a local minimisation of the IMSE, over the finite set of design points used to compute optimal design measures, or over a compact set using continuous optimisation (see Remark 8). The designs that are obtained depend on the kernel \(K(\cdot, \cdot)\) that is used, and on the kernel parameters \(\ell\) when \(K(\cdot, \cdot) = K_{\ell}(\cdot, \cdot)\) is parameterised. This situation is rather standard for optimal design in nonlinear situations (nonlinear regression models for instance), but some peculiarities are worth being mentioned. First, one is generally not interested in the value of \(\ell\), which constitutes a kind of nuisance parameter; second, the choice of the kernel is often arbitrary in computer experiments, \(K(\cdot, \cdot)\) being mainly a requisite tool introduced to construct a space-filling design; third, and more importantly, as illustrated in Section 7.5, the designs that are constructed are not very sensitive with respect to the choice of the kernel parameters \(\ell\) (which can always be checked numerically by constructing designs corresponding to different values of \(\ell\)).
When a quadrature involving $N_q$ points is considered, the relaxation method relies on the (partial) diagonalisation of a $N_q \times N_q$ matrix. The computation of the IMSE-induced BLM may therefore become too expensive for large values of $N_q$ (that should quickly increase with $d$), which appears as the main numerical bottleneck of the approach (note, however, that values of $N_q$ up to several thousands can easily be handled, see for instance the example in Section 7.3). In the special case where $\mathcal{X}$ is the hypercube $[0, 1]^d$, one may use tensorised kernels, having the property that the eigenfunctions are the tensor products of one-dimensional eigenfunctions (with associated eigenvalues equal to the product of the corresponding one-dimensional eigenvalues). More generally, the combination of the convex-relaxation approach with the construction of sparse quadratures allowing an accurate approximation of the eigendecomposition of the IMSE-related integral operator (see, e.g., Gauthier and Suykens, 2016) forms a promising perspective.

Finally, only Bayesian $A$-optimality has been considered, due to its direct connection with the truncated-IMSE for the initial RF model. Other choices are possible and deserve further investigations, see Fedorov (1996) and Spöck and Pilz (2010) for Bayesian $D$-optimality which may be related to the maximum value of the MSE over $\mathcal{X}$.

Acknowledgements

This work was partially supported by the ANR project DESIRE (DESIgn for spatial Random fiElds), nb. 2011-ISO1-001-01, jointly with the Statistics Department of the Johannes Kepler Universität, Linz (Austria) grant number FWF I 1833-N18. B. Gauthier also wishes to thank the MRI Department of EdF lab, Chatou, France.

We thank the three referees for their encouraging comments that helped to significantly improve the paper.

Appendix. Proofs

Proof of Proposition 1. Consider the Cholesky decomposition $K = CC^T$. For $k \in \mathbb{I}_+$, let $\phi_k = (\varphi_k(x_1), \ldots, \varphi_k(x_n))^T$. Using developments similar to those used to obtain (11), we have

$$C - C_{trc} = \sum_{k \notin \text{trc}} \lambda_k (\sqrt{\lambda_k} C^{-1} \phi_k)^T (\text{Id} - C^{-1} G (\text{Id} - C^{-1} G + A^{-1})^{-1} G^T C^{-T}) (\sqrt{\lambda_k} C^{-1} \phi_k).$$

From Gauthier and Pronzato (2014), we know that, for all $k \in \mathbb{I}_{trc}$,

$$0 \leq (\sqrt{\lambda_k} C^{-1} \phi_k)^T (\sqrt{\lambda_k} C^{-1} \phi_k) \leq 1.$$ 

(39)

Consider the matrix $Q = C^{-1} G (\text{Id} - C^{-1} G + A^{-1})^{-1} G^T C^{-T}$; it satisfies $Q = Q^T$ and

$$Q^2 = Q - C^{-1} G (\text{Id} - C^{-1} G + A^{-1})^{-1} A^{-1} (G^T K^{-1} G + A^{-1})^{-1} G^T C^{-T}.$$ 

(40)

The second matrix on the right-hand side of (40) is symmetric and non-negative definite. Therefore, $Q^2 \preceq Q$ (Loewner ordering), and similarly $(\text{Id} - Q)^2 \preceq (\text{Id} - Q).$ The real matrix $(\text{Id} - Q)$ is therefore a positive contraction, which, combined with (39), concludes the proof. □

Proof of Theorem 2. For $x \in \mathcal{X}$, let $c^T y$ be a linear predictor of $Y_x$ for the model (20) with no informative prior on $\theta$. The MSE associated with this predictor is given by

$$s^2(x) = \mathbb{E}((Y_x - c^T y)^2) = \left((g(x) - G^T c_x)^T \theta \right)^2 + K(x, x) + c_x^T K c_x - 2 c_x^T K x(x).$$

The no-bias condition implies $G^T c_x = g(x)$. The stationary condition for the Lagrangian (a necessary and sufficient condition for the minimisation of $s^2(x)$ with respect to $c_x$ under the no-bias constraint) is

$$
\begin{pmatrix}
0 & G^T \\
G & K
\end{pmatrix}
\begin{pmatrix}
\lambda \\
c_x
\end{pmatrix}
=
\begin{pmatrix}
g(x) \\
k(x)
\end{pmatrix},
$$

(41)

and the fact that $G$ has full rank ensures the existence and uniqueness of the best linear unbiased predictor, see, e.g., Stein (1999, p. 8).

If we consider the model (31) and a predictor of the form $(c^T x)^T y$, we obtain an equation of the same type as (41), where $c_x, \lambda, K$ and $K(x)$ are replaced by $c^T x, \lambda^2, K^2$ and $K^2(x)$ respectively. Applying the no-bias condition, we finally obtain

$$
\begin{pmatrix}
0 & G^T \\
G & K - GB^T
\end{pmatrix}
\begin{pmatrix}
\lambda^2 \\
c^T x
\end{pmatrix}
=
\begin{pmatrix}
g(x) \\
k(x) - GB(x)
\end{pmatrix},
$$

(42)

with $B^T = (b(x_1), \ldots, b(x_n))$. Therefore, if $(\lambda, c_x)$ satisfies (41), then $c^T x = c_x$ and $\lambda^2 = \lambda + B^T c_x - b(x)$ are solutions of (42). The two optimal linear predictors $(c_x)^T y$ and $(c^T x)^T y$ thus coincide and one can check that $s^2_{c^T x}(x) - s^2(x) = 0$. □
References