Efficient algorithms for calculating optimal designs in pharmacokinetics and dose finding studies

Tim Holland-Letz

Ruhr-Universität Bochum

Medizinische Fakultät

44780 Bochum, Germany

email: tim.holland-letz@rub.de

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Didier Renard

Novartis Pharma AG

Modeling & Simulation

4002 Basel, Switzerland

e-mail: didier.renard@novartis.com

October 20, 2010

Abstract

Random effects models are widely used in population pharmacokinetics and dose finding studies. In such models the presence of correlated observations (due to shared random effects and possibly residual serial correlation) usually makes the explicit determination of optimal designs difficult. In this paper we develop a class of multiplicative algorithms for the numerical calculation of optimal experimental designs in such situations. In particular we demonstrate its application in a concrete example of a cross-over dose finding trial. Additionally, we show that the methodology can be modified to determine optimal designs where there exist some requirements regarding the minimal number of treatments for several (in some cases all) experimental conditions.

Keyword and Phrases: correlated observations, heteroscedastic regression, dose-finding studies, pharmacokinetic models, random effects, locally optimal design, multiplicative algorithms

AMS Subject Classification: 62K05

1 Introduction

The work presented in this paper is motivated by several practical problems encountered in finding optimal designs for random effect models appearing both in population pharmakokinetic trials [Sheiner et al. (1977)] and in dose finding trials [Ting (2006)]. For illustration we consider a real case study which was recently analyzed by one of the authors. This was a dose finding clinical trial with the objective of characterizing the dose response of an experimental drug given under two different regimens, either once or twice daily. The study was therefore designed to cover a range of once- and twice-daily doses. The selected dosing groups were: placebo (control), 12.5 once daily, 12.5 twice daily, 25 once daily, 25 twice daily, 50 once daily, 50 twice daily, and 100 once daily. The details of the trial are blinded and numbers are given unitless serving for illustration purposes only. To improve efficiency a cross-over design was considered whereby each patient was to receive one of the 8 treatments in each of two periods. Treatment was given for four weeks in each period and efficacy measurements obtained at weekly intervals. A washout period of 4 weeks was scheduled between the two treatment periods. While the actual analysis integrated data from multiple visits within each treatment period, the simpler modeling approach described in this paper focuses on the efficacy measurements obtained at endpoint, i.e. after four weeks of treatment.

Previous data were available on the once-daily administration and suggested that the dose-response curve could be appropriately described with an Emax function. Although it is not uncommon for dose finding trials to be analyzed using standard statistical tools such as analysis of covariance, it is more efficient to rely on model-based approaches to estimate dose response [Bornkamp et al. (2007)]. While typically dose finding studies include a single mode of administration, the Emax model is easily extended to account for dosing regimen and provide a unified description of dose response. Two assumptions were made at this stage: 1. the maximal effect (Emax) is identical for once- and twice-daily regimens; 2. the regimen affects drug potency (ED50). The first assumption is reasonable based on pharmacological principles although it may not appear to hold over the observed dose range. The second assumption effectively imposes a shift in dose response for the twice-daily regimen relative to once-daily administration. Based on those assumptions a reasonable model can be written as follows:

$$(1.1) \qquad Y_{ij} = E_0 + b_{i0} + \frac{E_{max}x_{ij1}}{\frac{ED_{50}}{\theta^{x_{ij2}}} + x_{ij1}} + \varepsilon_{ij} \quad i = 1, \ldots, N, \quad j = 1, 2,$$

where the index i denotes subject and j treatment period, (x_{ij1} is the total daily dose received by subject i in period j (with values in $\{0, 12.5, 25, 50, 100\}$), and x_{ij2} is the corresponding indicator specifying whether the treatment was given once or twice daily, with values 0=once daily and 1=twice daily). The model was parameterized in terms of E_0 (intercept, corresponding

to placebo effect), E_{max} (maximal drug effect above placebo), ED_{50} (drug potency), and θ (potency modifier due to regimen). Further assumptions are that b_{i0} and ε_{ij} are random terms with normal distributions. A common assumption in such models is that conditionally on random effects, ε_{i1} and ε_{i2} are independent, although this assumption can be relaxed with the approach presented here.

Situations of this type are not unusual in dose finding studies and create challenging problems in statistics. In particular the fact that the study under consideration includes more than one observation per patient and the application of a random effect model requires advanced design methodology for correlated and heteroscedastic data. Moreover, in many cases there exist experimental constraints such as fixed dose levels (due to available dose strengths), requirements for certain dose levels or total sample size in the study, which must be taken into account in design optimization. A challenging question in the present application was to determine optimal treatment sequences and the number of patients that should be randomized to each of those sequences. As pointed out by Pukelsheim (2006), Section 11.8, design problems for these situations result in the determination of optimal designs which have to satisfy several constraints for the number of patients allocated to each experimental condition.

The present paper is devoted to the numerical construction of optimal designs in such situations. Because optimal designs for these kinds of problems are rarely available analytically we concentrate on algorithmic approaches. The most common techniques are the Fedorov-Wynn and the Simplex algorithm [see Fedorov (1972),Wynn (1972) or Mandal and Torsney (2006) and Harman and Pronzato (2007) for some more recent references]. However, these algorithms either change only single design points at a time, or are generic optimization methods which make no use of the special properties of the optimization problems appearing in design theory.

As an alternative we discuss the application of multiplicative algorithms in the present context which are particularly attractive for the determination of optimal designs on finite design spaces. These methods were introduced by Titterington (1976) for the D-optimality criterion [see also Silvey et al. (1978)] and are based on equivalence theorems, which characterize the optimum designs by means of an inequality for the directional derivative [Kiefer and Wolfowitz (1960), Pukelsheim (2006)]. In the recent literature multiplicative algorithms have found considerable interest [Dette et al. (2008), Harman and Trnovska (2009), Yu (2010)]. However - to our best knowledge - all authors only discuss very generic cases without consideration of the special properties (random effects and heteroscedasticity) encountered in the dose finding experiment discussed in the previous paragraph. In this paper we will demonstrate that this methodology can also be modified to solve these challenging optimal design problems. Moreover, we illustrate the methodology in several examples and determine optimal designs for the dose finding study

example. In particular, we show that the initial designs considered for the study described at the beginning of this section could be improved by the methodology developed in this paper. Additionally we extend the methodology to determine optimal designs which guarantee a minimal percentage of the observations at prespecified experimental conditions.

2   Optimal design problems and multiplicative algorithms

We consider a generic nonlinear random effects model, where m different observations are available on a number of N different patients each, and a vector of parameters is to be estimated. This allows us to consider models in population pharmacokinetics, where commonly more than one observation is available per patient, and other types of models such as the cross-over dose-response model discussed in the introduction (with $m = 2$). To be precise, we define the model under consideration as follows

$$(2.1) \qquad\qquad Y_{ij} = \eta(x_{ij}, \boldsymbol{b}_i) + \varepsilon_{ij} \quad i = 1, \dots, N, j = 1, \dots, m.$$

Observations can be taken at measurement conditions $x_{ij}, \quad i = 1, \dots, N, j = 1, \dots, m$, selected from a finite set $\mathcal{X}$ consisting of s different measurement settings. The parameters are assumed to be random, which means that $\boldsymbol{b}_i = (b_{i1}, \dots, b_{ik})^T$ represents a random effect of individual $i, \quad i = 1, \dots, N$. We assume that $\boldsymbol{b}_1, \dots, \boldsymbol{b}_N$ are independent, identically, k-dimensional normally distributed with expectation $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$ and covariance matrix $\Omega \in \mathbb{R}^{k \times k}$, such that the random effects $\boldsymbol{b}_1, \dots, \boldsymbol{b}_N$ are independent of the errors $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N$. The parameter θ (or a function of it) is the object of interest of the study. The function $\eta : \mathcal{X} \times \Theta \to \mathbb{R}$ is known, twice continuously differentiable with respect to $\boldsymbol{\theta}$ and the errors $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})^T$ for each individual patient are assumed to be normally distributed with expectation 0 and variance $W_i \in \mathbb{R}^{m \times m}$, $i = 1, \dots, N$.

In our further considerations we will treat the set of all m observations on a single patient as a single block. In other words, the effective measurement space is given by $\mathcal{X}^m$ and contains $r = s^m$ measurement settings. In this case $\boldsymbol{x}_i = (x_{i1}, \dots, x_{im}) \in \mathcal{X}^m$ denotes the set of experimental conditions for the treatment of the i-th patient. In order to be specific we consider the case of Maximum-Likelihood (ML)-estimation and start by constructing an information matrix for the set $\boldsymbol{x}_i = (x_{i1}, \dots, x_{im}) \in \mathcal{X}^m$ of all measurement settings for the individual patient $i, \quad i = 1, \dots, N$. The information matrix in this case is asymptotically given by the Fisher information matrix [Lehmann and Casella (1998)]. For a vector of normally distributed observations $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{im})^T$ with expectation $\boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{b}_i) = (\eta(x_{i1}, \boldsymbol{b}_i), \dots, \eta(x_{im}, \boldsymbol{b}_i))^T$ and variance V_i

this quantity is given by

$$(2.2) \quad A(\boldsymbol{x}_i) = \frac{\partial \boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{b}_i)}{\partial \boldsymbol{\theta}}^T V_i^{-1} \frac{\partial \boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{b}_i)}{\partial \boldsymbol{\theta}} + \frac{1}{2} \left\{ \begin{array}{ccc} tr(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \cdots & tr(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}) \\ \vdots & & \vdots \\ tr(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \cdots & tr(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}) \end{array} \right\}.$$

[see e.g. Retout and Mentré (2003)] and simplifies to

$$(2.3) \qquad A(\boldsymbol{x}_i) = \frac{\partial \boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{b}_i)}{\partial \boldsymbol{\theta}}^T V_i^{-1} \frac{\partial \boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{b}_i)}{\partial \boldsymbol{\theta}}.$$

in the case where the variance V_i does not depend on the parameter $\boldsymbol{\theta}$, for example under the assumption of homoscedasticity. Thus, the total information matrix for all N patients is given by

$$(2.4) \qquad M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_{i=1}^{N} A(\boldsymbol{x}_i).$$

In order to assess the effect of the random effects on the variance V_i of the observations on patient i we use a first order Taylor approximation of the function $\eta(x_{ij}, \boldsymbol{b_i})$, i.e.

$$(2.5) \qquad Y_{ij} \approx \eta(x_{ij}, \boldsymbol{\theta}) + \frac{\partial \eta(x_{ij}, \boldsymbol{\theta})}{\partial \theta}(\boldsymbol{b}_i - \boldsymbol{\theta}) + \varepsilon_{ij} \quad i = 1, \ldots, N, j = 1, \ldots, m,$$

and obtain

$$(2.6) \qquad V_i = Var(\boldsymbol{Y}_i) \approx \frac{\partial \boldsymbol{\eta}_m(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Omega \frac{\partial \boldsymbol{\eta}_m(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T + W_i.$$

Throughout this paper we will use the concept of approximate designs and consider designs as probability measures on the (finite) design space, which is given by $\mathcal{X}^m = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\}$ [see Kiefer (1974)]. Consequently, a design can be specified by a vector of weights $\mathbf{w} = (w_1, \ldots, w_r)$, where the j-th component w_j represents the relative proportion of total observations to be taken at the point $\boldsymbol{x}_j$ $(j = 1, \ldots, r)$. In practice a rounding procedure is applied to obtain integers $n_i \approx w_i N$ with $\sum_{j=1}^{n} n_i = N$, and the experimenter takes approximately n_i observations at each $\boldsymbol{x}_i$ $(i = 1, \ldots, r)$ [see Pukelsheim and Rieder (1992) for more details and some references]. For an approximate design $\mathbf{w} = (w_1, \ldots w_r)$ the information matrix is given by

$$(2.7) \qquad M(\boldsymbol{w}) = \sum_{i=1}^{r} w_i A(x_i),$$

and we will denote by $\mathcal{M}$ the space of all information matrices $M(\boldsymbol{w})$ of the form (2.7). Note that the information matrix in this situation, and thus any optimal design based on these

5

matrices, will usually depend on the unknown model parameters. Optimal designs, which maximize an appropriate function of the information matrix, will thus be called *locally optimal* designs [see Chernoff (1953)] and require an initial guess of the unknown parameters of the model. These designs might be sensitive with respect to misspecification of the parameters. However, they form of the basis of many more sophisticated design strategies such as sequential designs [see Chang and Ying (2009)], Bayesian designs [Chaloner and Larntz (1989) or Haines (1995)] or standardized maximin optimal designs [see Dette (1997) or Müller and Pázman (1998) among others]. Moreover, the main application of locally optimal designs consists in their use as benchmarks for commonly used designs. Locally optimum designs for nonlinear models with fixed effects have been discussed by numerous authors [see Ford et al. (1992), Biedermann et al. (2006), López-Fidalgo and Wong (2002) among many others]. In contrast to the fixed effect case locally optimal designs for nonlinear random effect models are rarely available in closed form. For some numerical results we refer to the work of Mentré et al. (1997), Retout et al. (2002), Atkinson (2008) or Dette and Holland-Letz (2009).

A (locally) optimal design for estimating the parameter $\boldsymbol{\theta}$ maximizes an appropriate function of the information matrix $M(\boldsymbol{w})$ and numerous criteria have been proposed for this purpose in the literature [see Silvey (1980), Pukelsheim (2006) or Randall et al. (2007)]. In the concrete application described in the introduction the following criteria were suggested by the biostatistics team. A main interest of the experiment consisted in a precise estimation of all parameters in the model, and for that purpose the use of the D-optimality criterion was proposed, which determines a design $\boldsymbol{w}$ maximizing the determinant of the information matrix

$$(2.8) \qquad \phi_D(\boldsymbol{w}) = |M(\boldsymbol{w})|^{1/k}.$$

On the other hand, in the dose-finding study example we have $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_4) = (E_0, E_{\max}, ED_{50}, \theta)$ one might want to focus on the parameters θ_3 and θ_4, as this puts emphasis on potency estimation which is key to the comparison of the two dosing regimens. For such cases the D_s criterion provides an interesting alternative to D-optimality, which determines a design maximizing

$$(2.9) \qquad \phi_{Ds}(\boldsymbol{w}) = |M_{III}^{-1}(\boldsymbol{w})|^{-1/s},$$

where $M_{III}^{-1}(\boldsymbol{w})$ is the lower right $s \times s$ sub-matrix of $M^{-1}(\boldsymbol{w})$ (in our example we have $s = 2$). Finally, if specific functionals of the parameter, such that the area under the curve (AUC), the minimum effective dose (MED) or the maximum of the curve are of interest, the c-optimality criterion is used, which determines the design $\boldsymbol{w}$ such that the functional

$$(2.10) \qquad \phi_c(\xi) = (c^T M^-(\boldsymbol{w}) c)^{-1}$$

6

is maximal. Here $c \in \mathbb{R}^k$ denotes a given vector reflecting the specific interest (AUC, MED, maximum of the curve etc.). Moreover, we assume that the vector c is estimable by the design $\boldsymbol{w}$, that is Range $(c) \subset$ Range $(M(\boldsymbol{w}))$ and for a $k \times k$ matrix A the matrix A^- denotes a generalized inverse of A.

In order to assess whether any given design is optimal for a given criterion ϕ we can use the equivalence theory [see Pukelsheim (2006)] which provides precise conditions to check the optimality of a given design. However, for a numerical construction it is usually more important to define an appropriate stopping criterion, such that the algorithm terminates with a design with a guaranteed efficiency

$$(2.11) \qquad \operatorname{eff}(\boldsymbol{w}) = \frac{\phi(\boldsymbol{w})}{\phi(\mathbf{w}^*)},$$

where $\mathbf{w}^*$ denotes the optimal design maximizing the function ϕ.

Note that the efficiency cannot be calculated without knowledge of the optimal design. However, we demonstrate in the following result that it is possible to present a lower bound for the efficiency defined in (2.11). For the sake of transparency we present here the result for designs with non-singular information matrices and differentiable information functions [see Pukelsheim (2006)]. For the general case we refer to Dette (1996) who discussed a more sophisticated bound for singular information matrices and general (not necessarily differentiable) criteria . The proof of the following result can be found in the appendix.

Theorem 2.1 *Consider a design $\boldsymbol{w}$ with non-singular information matrix $M(\boldsymbol{w})$ and assume that the optimality criterion $\phi : \mathcal{M} \to \mathbb{R}$ is differentiable at the point $M(\boldsymbol{w})$ with gradient $\nabla\phi(M(\boldsymbol{w}))$. Then the efficiency of the design $\boldsymbol{w}$ can be bounded from below by*

$$(2.12) \qquad \operatorname{eff}(\boldsymbol{w}) \geq \left(\max_{i=1}^{r} \frac{\operatorname{tr}(\nabla\phi(M(\boldsymbol{w}))A(\boldsymbol{x}_i))}{\operatorname{tr}(\nabla\phi(M(\boldsymbol{w}))M(\boldsymbol{w}))} \right)^{-1}.$$

In the case of $\operatorname{tr}(\nabla\phi(M(\boldsymbol{w}))M(\boldsymbol{w})) = 0$, we define this statement as $\operatorname{eff}(\boldsymbol{w}) \geq 0$. In particular, the design $\boldsymbol{w}$ is ϕ-optimal if and only if the right hand side of the inequality (2.12) is equal to 1 and in this case there is equality for all x_i with $w_i > 0$.

For the three optimality criteria ϕ_D, ϕ_{Ds} and ϕ_c defined in (2.8) - (2.10) the derivatives appearing in Theorem 2.1 can easily be calculated [see for example Pázman (1986)]. In particular we have

$$(2.13) \qquad \nabla\phi_D(M(\boldsymbol{w})) = \frac{1}{k}|M(\boldsymbol{w})|^{1/k}M^{-1}(\boldsymbol{w})$$

$$(2.14) \qquad \nabla\phi_{D_s}(M(\boldsymbol{w})) = \frac{1}{s}|M_{III}^{-1}(\boldsymbol{w})|^{-1/s}(M^{-1}(\boldsymbol{w}) - \tilde{M}^{-1}(\boldsymbol{w}))$$

$$(2.15) \qquad \nabla\phi_c(M(\boldsymbol{w})) = \frac{M^{-1}(\boldsymbol{w})cc^TM^{-1}(\boldsymbol{w})}{(c^TM^{-1}(\boldsymbol{w})c)^2}$$

where

$$\tilde{M}^{-1}(\boldsymbol{w}) = \begin{pmatrix} M_I^{-1}(\boldsymbol{w}) & 0 \\ 0 & 0 \end{pmatrix}$$

and $M_I(\boldsymbol{w})$ is the upper left $(k-s) \times (k-s)$ sub-matrix of the matrix $M(\boldsymbol{w})$. In the following discussion these quantities will be used in the construction of multiplicative algorithms which represent a very flexible tool to find optimal designs.

The first variant of this class of methods has been introduced by Titterington (1976) for the D-optimality criterion. Dette et al. (2008) extended Titterington's idea and proposed a class of multiplicative algorithms for the determination of D-optimal designs. In the same paper it is indicated that the methodology may be also applicable to other optimality criteria. A very far reaching generalization has been given by Yu (2010), which allows the application of the algorithm to a broad class of optimality criteria. To be precise, recall that the design space is given by $\mathcal{X}^m = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\}$, with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im}), i = 1, \ldots, r$, and that the design is characterized by a vector $\boldsymbol{w}$ from the r-dimensional simplex $\mathcal{W}$ defined in (5.1). Let $\mathbf{w^0}$ be an element of the set $\mathcal{W}$ which will be used as starting point for the algorithm. Furthermore, for a vector $\boldsymbol{w} = (w_1, \ldots, w_r) \in \mathcal{W}$ let $M(\boldsymbol{w})$ be the corresponding information matrix for the parameter $\boldsymbol{\theta}$ in model (2.1), $A(\boldsymbol{x}_i)$ the information matrix of an individual measurement at the point $\boldsymbol{x}_i$, $i = 1, \ldots, r$ and ϕ a differentiable information function to be maximized. As $w_i > 0$ for all $i = 1, \ldots, r$, $M(\boldsymbol{w})$ will be nonsingular for all $\boldsymbol{w} \in \mathcal{W}$.

We follow Yu (2010) and starting with the vector $\boldsymbol{w}^0$ we define for $q = 0, 1, \ldots$ a vector of "new" weights $\mathbf{w}^{q+1} = (w_1^{q+1}, \ldots, w_r^{q+1}) \in \mathcal{W}$ by

(2.16)
$$w_i^{q+1} = w_i^q \frac{d_i^\lambda}{\sum_{j=1}^r w_j^q d_j^\lambda} \qquad i = 1, \ldots, r,$$

where $\lambda \in (0, 1]$ denotes a calibration parameter and

(2.17)
$$d_i = d_i(\mathbf{w}^q) = \text{tr}(\nabla \phi(M(\mathbf{w}^q)) A(\boldsymbol{x}_i)) \quad i = 1, \ldots, r.$$

Yu (2010) proved the monotonicity of this algorithm for all $\lambda \in (0, 1]$, that is

$$\phi(M(\mathbf{w}^{q+1})) \geq \phi(M(\mathbf{w}^q)); \ q = 0, 1, \ldots$$

and in the case $0 < \lambda < 1$ he proved that the sequences $(\mathbf{w}^q)_{q \in \mathbb{N}}$ converges to the optimal design $\boldsymbol{w}^*$ maximizing the function $\phi(M(\boldsymbol{w}))$ in the set $\mathcal{W}$. Observing the representations of the gradients in equations (2.13) - (2.15) and the Fisher information matrix given in (2.2) we can apply these general principles to the problems discussed in the introduction.

3 Examples

3.1 A nonlinear population model

We start with an example from pharmakokinetics where we try to estimate the area under a concentration curve (AUC) for the blood concentration of some medication at specified time points after an intravenous injection. Situations like this are usually modeled by a random effects population model [see Sheiner et al. (1977)]. We thus assume that the regression in model (2.1) is given as an exponential elimination function, that is

$$(3.1) \qquad \eta(x, b_i) = b_{i1} e^{-b_{i2} x},$$

where the gradient is calculated as

$$\frac{\partial \eta(x, b_i)}{\partial \theta} = (e^{-b_{i2} x}, -x b_{i1} e^{-b_{i2} x})^T.$$

We further assume that two observations are available for each patient, i.e. this is a sparse sampling situation. Because of the random effects model, the variance in this situation will be dependent on the parameters. Consequently, the individual information matrix $A(x_i)$ requires the full expression stated in (2.2).

We obtain for the element in the position (j, l) of the variance matrix V_i of observations on patient i, $\quad i = 1, \ldots, N$:

$$(3.2) \qquad [V_i]_{j,l=1}^2 \approx \frac{\partial \eta_2(x_i, \theta)^T}{\partial \theta_j} \Omega \frac{\partial \eta_2(x_i, \theta)}{\partial \theta_l} + \sigma^2 \delta_{jl}.$$

The derivatives of this expression are given by $(q = 1, 2)$

$$(3.3) \qquad \frac{\partial [V_i]_{j,l=1}^2}{\partial \theta_q} \approx \frac{\partial^2 \eta_2(x_i, \theta)^T}{\partial \theta_j \partial \theta_q} \Omega \frac{\partial \eta_2(x_i, \theta)}{\partial \theta_l} + \frac{\partial \eta_2(x_i, \theta)^T}{\partial \theta_j} \Omega \frac{\partial^2 \eta_2(x_i, \theta)}{\partial \theta_l \partial \theta_q}$$

We thus need the first and second derivative of the concentration function, which are easily calculated.

The design space is given by $\mathcal{X} = (0, 0.1, 0.2, \ldots, 3.0)$, with $s = 31$ potential measurement settings. As both measurements on a single patients will be treated as a single experimental unit, the effective design space is $\mathcal{X}^2$. Removing one of the pairs (x_{i1}, x_{i2}) or (x_{i2}, x_{i1}) (because they yield same experimental condition for the ith patient) as well as duplicate measurements at an identical time point (which are not viable in practice), we get the modified design space $\tilde{\mathcal{X}}^2 = \{(0, 0.1), (0, 0.2), \ldots, (0, 3), (0.1, 0.2), \ldots, (2.9, 3.0)\}$ containing $r = s * (s - 1)/2 = 465$

elements.

For illustrative purposes we set $\theta_1 = 46, \theta_2 = 1.7$, assume that $\epsilon_1, \ldots, \epsilon_n$ are independent identically distributed random variables and make the additional assumption that the population variances are $\Omega = diag(1, 0.1)$. Note that in this case the correlation between different observations on the same patient is caused only by the common random effect.

We are trying to estimate the area under the concentration curve (AUC), which for the function considered here is given by $AUC = \frac{\theta_1}{\theta_2}$. This corresponds to a c-optimal design problem, where the vector c given by the gradient of the quantity to be estimated, i.e. $c = (\frac{1}{\theta_2}, -\frac{\theta_1}{\theta_2^2})$. Consequently, the coefficient d_i in the recursion (2.16) is given by

$$(3.4) \qquad d_i = \text{tr}\Big(-\frac{M^{-1}(\boldsymbol{w}^{(i)})cc^T M^{-1}(\boldsymbol{w}^{(i)})}{(c^T M^{-1}(\boldsymbol{w}^{(i)})c)^2} A(\boldsymbol{x}_i)\Big), \quad i = 1, \ldots, r.$$

We begin with an equally spaced starting design $\boldsymbol{w}^0$ with $\boldsymbol{w}_i^0 = \frac{1}{465}$ for all $i = 1, \ldots, 465$ and do repeated iterations of algorithm (2.16) setting $\lambda = 0.99$. The weights obtained after 1000 iterations are shown in Figure 1 (left part). The right part of Figure 1 shows the lower bound of the efficiency (see Theorem 2.1) of the design for each iteration, and we observe that good efficiencies will be reached very quickly.
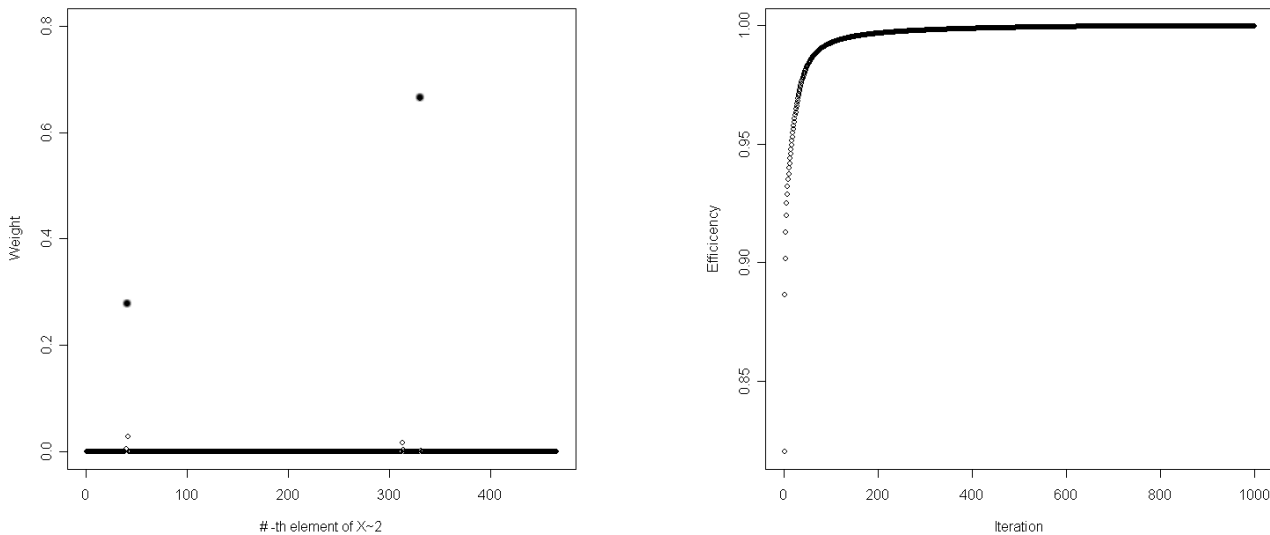

Figure 1: *Left panel: Distribution of weights after 1000 iterations of the algorithm* (2.16) *in the population model. Right panel: lower bound* (2.12) *for the efficiency of the design after each iteration*

10

For the practical implementation of the design we approximate this design with the strict two point design with weights 0.312 and 0.688 at the points $(0.10, 1.20)$ and $(1.40, 1.50)$, respectively, which yields still an efficiency of at least 99.9%. This means that about a 31.8% of the patients should be observed at the time points 0.1 and 1.2, and the rest at 1.4 and 1.5.

3.2 Optimal designs for a dose finding study

In our second example we construct optimal designs for the dose finding study which was described in the introduction. The dose response curve for each dosing regimen is represented by an Emax model, with two observations taken on each patient in a cross-over setup:

$$(3.5) \qquad Y_{ij} = b_{i1} + \frac{b_2 x_{ij1}}{\frac{b_3}{b_4^{x_{ij2}}} + x_{ij1}} + \varepsilon_{ij} \quad i = 1, \ldots, N, \quad j = 1, 2$$

The parameter b_{i1} is modeled as a random effect with expectation θ and variance ω, while the other parameters are treated as fixed.

The experimental conditions x_{ij} consist of total daily dose (x_{ij1}) with values in $\{0, 12.5, 25, 50, 100\}$, and an indicator specifying whether the treatment was given once $(x_{ij2} = 0)$ or twice daily $(x_{ij2} = 1)$. Since regimen should not impact the placebo effect, the design space is given by

$$\mathcal{X} = \{(0,0), (12.5, 0), (25, 1), (25, 0), (50, 1), (50, 0), (100, 1), (100, 0)\}$$

for one measurement, and $\mathcal{X}^2$ for both measurements on one patient. ¿From an optimal design viewpoint it does not matter so much whether a patient receives dose 1 before dose 2, or the opposite sequence and so the two sequences were considered interchangeably [Note: from a practical viewpoint it might be desirable to randomize patients to both sequences, e.g. to test for a period effect, but this issue is not considered here]. As a result the design space effectively includes $r = 36$ combinations. A D-optimal design for ML-estimation was investigated, as well as a D_s-optimal design for the parameters θ_3 and θ_4. Three candidate designs were initially under consideration. The first one (design 1) allocates 25% of the observations at $\{(0,0), (12.5, 0)\}$, $\{(25, 1), (25, 0)\}$, $\{(50, 1), (50, 0)\}$, and $\{(100, 1), (100, 0)\}$, i.e. it favors direct regimen comparisons for different total daily doses. The second (design 2) focused on 8 basic sequences: $\{(0, 0), (50, 0)\}$, $\{(12.5, 0), (50, 1)\}$, $\{(0, 0), (100, 1)\}$, $\{(12.5, 0), (100, 0)\}$, $\{(25, 1), (50, 1)\}$, $\{(25, 0), (50, 0)\}$, $\{(25, 1), (100, 0)\}$, $\{(25, 0), (100, 1)\}$; i.e. it emphasized balanced allocation of low and high doses within patients which was thought to improve dose-response estimation from basic optimal principles (recall that the simple Emax model has optimal doses at 0, ED_{50} and the highest possible dose). Finally the third design (design 3) was an equally spaced design on all 28 possible combinations.

Based on previous data it was possible to assign reasonable values to the parameters. First, the question centered around the comparison to placebo, hence b_1 was set to 0. Estimates of Emax ($b_2 = 0.15$) and ED50 ($b_3 = 15$) were obtained for once-daily dosing. The potency modifier (b_4) could not be estimated from previous data but it was expected to be greater than 1 based on pharmacological principles. A value of 1.2 was used for illustration purposes in this paper. For the variability parameter, it was known that the ratio of within- to between-subject variability was approximately 2 to 1. Therefore, assuming total variance is 0.225^2, the values $var\{\epsilon_{ij}\} = \sigma^2 = 0.225^2 * 2/3$ and $\omega = 0.225^2 * 1/3$ were chosen.

We start with the D-optimal design, for which the gradient of the criterion is given by (2.13) yielding

$$d_i = \frac{1}{k}\text{tr}(M(\boldsymbol{w})^{-1}A(\boldsymbol{x}_i))|M(\boldsymbol{w})|^{1/k}.$$

The formula for the information matrix $A(x_i)$ for any element of $\tilde{\mathcal{X}}^2$ has the same structure as in the previous example, see equation (2.2). Again, we will need the first and second derivative of the concentration function (3.5) to calculate the matrix V_i and its derivative given in equations (3.2) and (3.3). We can then calculate $A(\boldsymbol{x}_i)$, $i = 1, \ldots, r$, $M(\boldsymbol{w})$ and $\nabla\phi_{M(\boldsymbol{w})}$ for any given design $\boldsymbol{w}$.

We again start with an equally spaced design $\boldsymbol{w}^0$ and use algorithm (2.16). Results after 100 iterations are shown in Figure 2 (left part).

With an appropriate rounding we obtain a design, which advises the experimenter to take 18%, 20%, 10%, 27% and 25% of the observations at the experimental conditions $\{(0,0),(12.5,0)\}$, $\{(0,0),(25,1)\}$, $\{(0,0),(100,0)\}$, $\{(12.5,0),(100,1)\}$ and $\{(25,1),(100,0)\}$, respectively. In the right part of Figure 2 we observe that an efficiency close to 100% is reached very quickly. For comparison, we calculated the the D-efficiencies of the three candidate designs initially proposed, which are 73%, 80% and 78%. Therefore a substantial improvement could be achieved by the application of optimal designs.

The situation for D_s-optimality is very similar. The only difference is the gradient which is now given by (2.14). In this case the design found by the multiplicative algorithm is a four point design and advices the experimenter to take 28%, 19%, 14% and 40% of the observations at the experimental conditions $\{(0,0),(12.5,0)\}$, $\{(12.5,0),(25,1)\}$, $\{(12.5,0),(100,1\}$ and $\{(25,1),(100,0)\}$, respectively. The efficiencies of the three candidate designs are 59%, 71% and 66%, so again a substantial improvement can be achieved by the application of optimal designs.
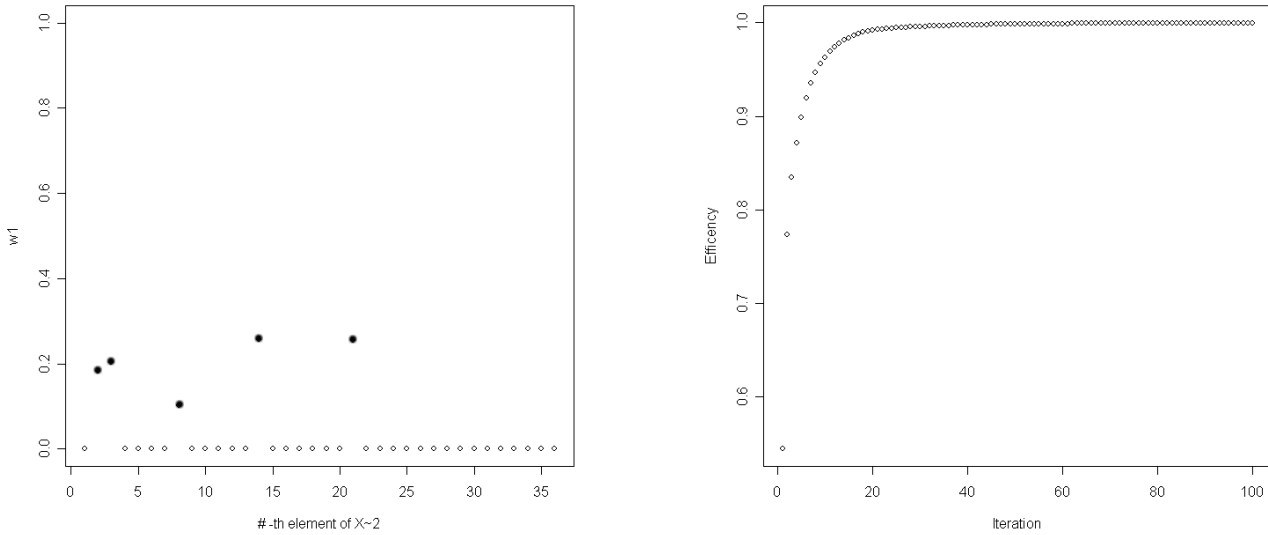
Figure 2: *Left panel: Distribution of weights after 100 iterations of the multiplicative algorithm* (2.16) *in the dose finding cross-over model* (3.5) *with two observations per patient, for D-optimality. Right panel: Lower bound* (2.12) *for the efficiency of the design after each iteration*

4 Multiplicative algorithm for designs with protected runs

In many practical applications it is necessary to have at least a pre specified number of observations at certain experimental conditions. For example, in a dose finding study it is reasonable to assign at least a certain proportion, say for example 10%, of the patients to the dose level most likely to be used in the production run. In the example from the introduction, such a constraint was in place, the direct comparison between dose level 50 in one and in two applications was requested to be included with a minimum weight of 25%.

Similarly, if the experiment is conducted in two (or more) stages with N_1 and N_2 observations in stage 1 and 2, respectively, it is common practice to use the information from the first stage in experimental design of the second stage as well. A typical example for this situation are also dose finding studies, where the initial trials will usually be done using small sample sizes (to check toxicity and tolerance). The total information of the experiment is then given by

$$M(\boldsymbol{w}) = \frac{N_1}{N_1 + N_2} M(\boldsymbol{w}_1) + \frac{N_2}{N_1 + N_2} M(\boldsymbol{w}_2),$$

13

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ denote the approximate designs corresponding to the first and second stage, respectively. It was pointed out in Pukelsheim (2006), chapter 11.5 and 11.6, that the determination of an optimal design $\boldsymbol{w}_2^*$ for the second stage corresponds to the determination of an optimal design $\boldsymbol{w}^*$ for the total experiment, where the class of designs under consideration consists of all designs which have a minimum mass at the support point of the design $\boldsymbol{w}_1$, that is

$$(4.1) \qquad \boldsymbol{w} \geq \frac{N_1}{N_1 + N_2} \boldsymbol{w}_1.$$

In this section we will develop multiplicative algorithms for this type of problems. More precisely, we can consider the situation where the optimal design $\boldsymbol{w}^* = (w_1^*, \ldots, w_r^*)$ has to satisfy certain constraints

$$(4.2) \qquad a(\boldsymbol{x}_i) \;\; \leq \;\; w_i^*, \quad i = 1, \ldots, r.$$

Here for each $i = 1, \ldots r$ the quantity $a(\boldsymbol{x}_i)$ denotes a lower bound for the percentage of the total which should at least be taken under experimental condition $\boldsymbol{x}_i$, $\quad i = 1, \ldots, r$ (if there is no restriction at $\boldsymbol{x}_i$ we put $a(\boldsymbol{x}_i) = 0$). We will use the notation $\tilde{\Xi} \subseteq \Xi$ for the set of all designs fulfilling these constraints and call these designs admissible in the following discussion. For the inclusion of constraints of the type (4.2) let $\alpha = \sum_{i=1}^{r} a(\boldsymbol{x}_i)$ be the sum of all required weights in an admissible design ξ. In the following we will introduce a "modified" design space which includes only those designs fulfilling (4.2). To do so, we associate with the original design space $\mathcal{X}^m$ a set $\tilde{\mathcal{X}}^m = (\nu_1, \ldots, \nu_r)$. Each element ν_i of $\tilde{\mathcal{X}}^m$ is itself a design on $\mathcal{X}^m$, with weights $a(\boldsymbol{x}_j)$, $\quad j = 1, \ldots, r$ for all points from the measurement range $\mathcal{X}^m$ and an additional weight $(1 - \alpha)$ for the point $\boldsymbol{x}_i$. The effective weights of a design $\nu_i \in \tilde{\mathcal{X}}^m$ are thus given by

$$(4.3) \qquad \nu_i(\boldsymbol{x}_j) = \begin{cases} a(\boldsymbol{x}_j) & \text{if } j \neq i \\ a(\boldsymbol{x}_j) + (1 - \alpha) & \text{if } j = i \end{cases}.$$

Thus, there is exactly one element in the "new" design space $\tilde{\mathcal{X}}^m$ for each element of the given design space $\mathcal{X}^m$. Any design $\mathbf{u} = (u_1, \ldots, u_q)$ on the "new" design space $\tilde{\mathcal{X}}^m$ with q support points and given weights $u(\nu_i)$, $\quad i = 1, \ldots, q$ induces a design $\boldsymbol{w}$ on $\mathcal{X}^m$, where the weights for the point $\boldsymbol{x}_s$ are given by

$$(4.4) \qquad w(\boldsymbol{x}_s) = (1 - \alpha)u_s + \sum_{j=1}^{q} u_j a(\boldsymbol{x}_s) = (1 - \alpha)u_s + a(x_s) \geq a(\boldsymbol{x}_s).$$

Thus, any such design will fulfill the constraints $a(\boldsymbol{x}_i) \leq w_i$, $\quad i = 1, \ldots, r$. Conversely, any design on $\mathcal{X}^m$ satisfying (4.2) defines a design on $\tilde{\mathcal{X}}^m$. Note that the space of designs on $\tilde{\mathcal{X}}^m$

retains the convexity of the original space of designs and contains all admissible designs (i.e. all elements of $\tilde{\Xi}$). The information matrix for any element of $\tilde{\mathcal{X}}^m$ is given by

$$(4.5) \qquad A(\nu_i) = (1-\alpha)A(\boldsymbol{x}_i) + \sum_{j=1}^{r} a(\boldsymbol{x}_j)A(\boldsymbol{x}_j).$$

On this modified design space we can now apply the multiplicative algorithm in exactly the same way as described in Section 3.

We illustrate this idea by the calculation of D-optimal designs for the dose finding problem described in the introduction. We assume that we have to take at least 25% of the observations at the combination $\{(50,0),(50,1)\}$ (alternative 1, actual constraint in the trial), or alternatively 10% each at the two combinations $\{(0,0),(50,0)\}$ and $\{(100,0),(100,1)\}$ (alternative 2). The resulting designs for the remaining 75% respectively 80% of observations are the following:

Alternative 1:

$$\xi = \begin{pmatrix} (0,0),(12.5,0) & (0,0),(25,1) & (0,0),(100,0) & (12.5,0),(100,1) & (25,1),(100,0) \\ 0.15 & 0.19 & 0.08 & 0.19 & 0.14 \end{pmatrix}.$$

Alternative 2:

$$\xi = \begin{pmatrix} (0,0),(25,1) & (0,0),(25,0) & (12.5,0),(100,1) & (25,1),(100,0) \\ 0.25 & 0.12 & 0.17 & 0.26 \end{pmatrix}.$$

As expected, the relatively noninformative point $\{(50,0),(50,1)\}$ in alternative 1 changes very little in the rest of the design, while in alternative 2 the center point is no longer needed due to the already available information. In all cases, efficiencies of the designs after 200 iterations are above 99.9%.

5 Appendix: Proof of Theorem 2.1

Let $\operatorname{tr}(\nabla\phi(M(\boldsymbol{w}))M(\boldsymbol{w})) \neq 0$, consider the r-dimensional simplex

$$(5.1) \qquad \mathcal{W} = \{w : w_i > 0, \sum_{i=1}^{r} w_i = 1\},$$

and define for matrices $M(\boldsymbol{w}_1), M(\boldsymbol{w}_2), \boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{W}$ the directional derivative

$$D_\phi(\boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{\partial}{\partial\alpha}\phi((1-\alpha)M(\boldsymbol{w}_1) + \alpha M(\boldsymbol{w}_2))\Big|_{\alpha=0+},$$

then it is easy to see that

$$D_\phi(\boldsymbol{w}_1, \boldsymbol{w}_2) = \text{tr}\left(\nabla\phi(M(\boldsymbol{w}_1))(M(\boldsymbol{w}_2) - M(\boldsymbol{w}_1))\right).$$

Now the function

$$g(\alpha) = \phi((1-\alpha)M(\boldsymbol{w}_1) + \alpha M(\boldsymbol{w}_2))$$

is concave, which implies

$$\begin{aligned}
\phi(M(\boldsymbol{w}_2)) - \phi(M(\boldsymbol{w}_1)) &= g(1) - g(0) \le g'(0) = D_\phi(\boldsymbol{w}_1, \boldsymbol{w}_2) \\
&= \text{tr}(\nabla\phi(M(\boldsymbol{w}_1))(M(\boldsymbol{w}_2) - M(\boldsymbol{w}_1))).
\end{aligned}$$

Using positive homogenity of the information functions and interpreting $-\text{tr}(\nabla\phi(M(\boldsymbol{w}_1))M(\boldsymbol{w}_1))$ as the directional derivative of $\boldsymbol{w_1}$ in the direction of 0, we obtain the identity

$$\nabla\phi(M(\boldsymbol{w}_1)) = \phi(M(\boldsymbol{w}_1))\frac{\nabla\phi(M(\boldsymbol{w}_1))}{\text{tr}(\nabla\phi(M(\boldsymbol{w}_1))M(\boldsymbol{w}_1))}.$$

Inserting this into the preceding statement yields

$$\frac{\phi(M(\boldsymbol{w}_2))}{\phi(M(\boldsymbol{w}_1))} \le \frac{\text{tr}(\nabla\phi(M(\boldsymbol{w}_1))M(\boldsymbol{w}_2))}{\text{tr}(\nabla\phi(M(\boldsymbol{w}_1))M(\boldsymbol{w}_1))}$$

for all $\boldsymbol{w}_1, \boldsymbol{w}_2$, and the assertion follows using $\boldsymbol{w}_1 = \boldsymbol{w}$, $\boldsymbol{w}_2 = \boldsymbol{w}^*$ $(i = 1, \ldots, r)$ and from the observation that for any vector of weights $\tilde{\boldsymbol{w}}$ we have

$$\text{tr}(\nabla\phi(\text{M}(\boldsymbol{w}_1))\text{M}(\tilde{\boldsymbol{w}})) \le \max_{i=1}^{r}(\text{tr}(\nabla\phi(\text{M}(\boldsymbol{w}_1))\text{A}(\boldsymbol{x_i}))).$$

References

Atkinson, A. C. (2008). Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. *Journal of Statistical Planning and Inference*, 138(9):2595–2606.

Biedermann, S., Dette, H., and Zhu, W. (2006). Optimal designs for dose-response models with restricted design spaces. *Journal of the American Statistical Association*, 101:747–759.

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., and Hsu, C.-H. e. a. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials (with discussion). *J Biopharm Stat*, 17:965–995.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.

Chang, H. and Ying, Z. (2009). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *Annals of Statistics*, 37:1466–1488.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.

Dette, H. (1996). Lower bounds for efficiencies with applications. In Brunner, E. and Denker, M., editors, *Research Developments in Probability and Statistics: Festschrift zum 65sten Geburtstag von M.L. Puri*, pages 111–124. VSP Utrecht.

Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

Dette, H. and Holland-Letz, T. (2009). A geometric characterization of c-optimal designs for heteroscedastic regression. *Annals of Statistics*, 37(6B):4088–4103.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2008). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics and Data Analysis*, 53(2):312–320.

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

Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Ser. B*, 54:569–583.

Haines, L. M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *Journal of the Royal Statistical Society, Series B*, 57(3):575–598.

Harman, R. and Pronzato, L. (2007). Improvements on removing nonoptimal support points in D-optimum design algorithms. *Stat. & Prob. Letters*, 77(1):90–94.

Harman, R. and Trnovska, M. (2009). Approximate d-optimal designs of experiments on the convex hull of a finite number of information matrices. *Mathematica Slovaca*, 59:693–704.

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

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Springer, New York.

López-Fidalgo, J. and Wong, W. K. (2002). Design issues for the Michaelis-Menten model. *Journal of Theoretical Biology*, 215:1–11.

Mandal, S. and Torsney, B. (2006). Construction of optimal designs using a clustering approach. *J. Statist. Plann. Inference*, 136:1120–1134.

Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84(2):429–442.

Müller, C. H. and Pázman, A. (1998). Applications of necessary and sufficient conditions for maximum efficient designs. *Metrika*, 48:1–19.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* D. Reidel Publishing Company, Dordrecht.

Pukelsheim, F. (2006). *Optimal Design of Experiments, Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, 79:763–770.

Randall, T., Donev, A., and Atkinson, A. C. (2007). *Optimum Experimental Designs, with SAS.* Oxford University Press, Oxford.

Retout, S. and Mentré, F. (2003). Further developments of the Fisher information matrix in nonlinear mixed-effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 13:209–227.

Retout, S., Mentré, F., and Bruno, R. (2002). Fisher information matrix for nonlinear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine*, 21:2623–2639.

Sheiner, L. B., Rosenberg, B., and Marathe, V. (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of Pharmacokinetics and Biopharmaceutics*, 5:445–479.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Silvey, S. D., Titterington, D. M., and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. *Comm. Statist., Theory Meth.*, 14:1379–1389.

Ting, N. (2006). *Dose Finding in Drug Development*. Springer, New York.

Titterington, D. M. (1976). Algorithms for computing d-optimal design on finite design spaces. *Proc. of the 1976 Conf. on Information Science and Systems, John Hopkins University*, 3:213–216.

Wynn, H. P. (1972). Results in the construction of d-optimum experimental designs. *Journal of the Royal Statistical Society B*, 34:133–147.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *Annals of Statistics*, 38:1593–1606.