Published in final edited form as:

J R Stat Soc Series B Stat Methodol. 2011 March 1; 73(2): 239–252. doi:10.1111/j.1467-9868.2010.00757.x.

A geometric characterization of optimal designs for regression models with correlated observations

Tim Holland-Letz,

Ruhr-Universität Bochum, Medizinische Fakultät, 44780 Bochum, Germany

Holger Dette, and

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

Andrey Pepelyshev

University of Sheffield, Department of Probability & Statistics, Sheffield, U.K.

Tim Holland-Letz: tim.holland-letz@rub.de; Holger Dette: holger.dette@rub.de; Andrey Pepelyshev: a.pepelyshev@sheffield.ac.uk

Abstract

We consider the problem of optimal design of experiments for random effects models, especially population models, where a small number of correlated observations can be taken on each individual, while the observations corresponding to different individuals are assumed to be uncorrelated. We focus on c -optimal design problems and show that the classical equivalence theorem and the famous geometric characterization of Elfving (1952) from the case of uncorrelated data can be adapted to the problem of selecting optimal sets of observations for the n individual patients. The theory is demonstrated by finding optimal designs for a linear model with correlated observations and a nonlinear random effects population model, which is commonly used in pharmacokinetics.

Keyword and Phrases

c -optimal design; correlated observations; Elfving's theorem; pharmacokinetic models; random effects; locally optimal design; geometric characterization

1 Introduction

It is a common situation in pharmacokinetic trials that only a very small number of measurements can be taken on a single patient, but a larger number of n different patients are available [Sheiner et al. (1977), Schmelter (2007), Colombo et al. (2006)]. In this situation it is impossible to reliably estimate parameters of interest for each patient. However, often these individual parameters are not of primary interest, because it is assumed that the individual parameters are realizations of some global distribution. Therefore, the main aim of the experiment is the estimation of the mean and/or variance of this distribution. This results in a random effects model and is called the population approach [Retout and Mentré (2003)]. Unfortunately, the common random effect causes measurements at different timepoints on a single patient to be correlated, therefore most of the commonly used tools of classical optimal design theory are not applicable in this context. Compared to the case of uncorrelated observations the optimal design problem for dependent data is intrinsically more difficult. Most authors used asymptotic arguments to determine efficient designs [see Sacks and Ylvisaker (1968), Bickel and Herzberg (1979), Näther (1985), Müller and Pázman (2003) among others]. In general the powerful equivalence theorem [Pukelsheim (2006)] cannot be transferred to the case of dependent

data [for an equivalence theorem in the case of repeated and correlated observations at the same measurement setting see Fedorov (1972)]. Note also that geometric characterizations of optimal designs are only available for the uncorrelated case [Elfving (1952), Ford et al. (1992), Haines (1995) or Studden (2005)].

The purpose of the present paper is to derive a geometric characterization of c -optimal designs, which minimize the variance of a linear combination of the estimates of the parameters (specified by a given vector c), for models with correlated observations. Note that many commonly used objectives in the statistical analysis (as designing the experiment for the estimation of the area under the curve, the maximum concentration or, in dose finding studies, the minimal effective dose) yield optimality criteria which are special cases of the c -optimality criterion [see Atkinson et al. (1993)]. In the following sections we show that if the number of available observations is the same for each patient, the total information of all observations on a single patient, accounting for correlations, can be expressed as a sum of information matrices in the usual form for uncorrelated observations. More precisely, if m observations are available for each patient, there exist vector valued functions $\tilde{f}_\ell$, $\ell = 1, \dots, m$ such that the total information matrix for the set of m observations on this patient can be written in the form

$$I = \sum_{\ell=1}^m \tilde{f}_\ell \tilde{f}_\ell^T. \quad (1.1)$$

For this representation we introduce for the individual observations a design space of m observations for each patient in addition to the original design space. Using this representation, we can derive an equivalence theorem for c -optimal designs using the general theory in Pukelsheim (2006) and obtain a geometric characterization of c -optimal designs for the problem of allocating the n available patients to different sets of m individual observations. As a result we obtain a generalization of the famous result of Elfving to the case of dependent data.

The theoretical details are presented in Section 2. In Section 3 we demonstrate the application of these ideas in two examples, a linear model and a basic nonlinear model taken from population pharmacokinetics. Finally, some technical details are given in an Appendix. For the sake of brevity this paper concentrates on the geometric characterization of locally c -optimal designs. However, it is worthwhile to mention that in the case of uncorrelated observations Elfving-type characterizations are also available for other optimality criteria, including D -, E - and Bayesian optimality criteria [see Dette (1993a,b) and Dette (1996)]. An interesting problem of future research is the extension of the methodology developed in the present paper to obtain geometric characterizations of optimal designs for correlated observations with respect to these criteria.

2 An Elfving representation for models with correlated observations

We begin our discussion with the case of linear models where the results are slightly more transparent. The analysis of nonlinear models can easily be reduced to this situation (see Remark 2.1), while the case of random effect models is discussed in Section 3. Assume that m observations can be taken, each on a number of n individuals in the linear model

$$Y_{ij} = \theta^T f(x_{ij}) + \varepsilon_{ij}; \quad i = 1, \dots, n, j = 1, \dots, m, \quad (2.1)$$

where Y_{ij} denotes the j -th observation on the i -th individual and x_{ij} is the experimental condition corresponding to this observation, which is chosen from a compact interval $\chi \subset \mathbb{R}$. We use $x_i = (x_{i1}, \dots, x_{im})$ to denote all experimental conditions corresponding to the individual i . The vector $\theta = (\theta_1, \dots, \theta_k)^T \in \Theta \subset \mathbb{R}^k$ is the vector of parameters to be estimated, $f(x) = (f_1(x), \dots, f_k(x))^T$ denotes a vector of known functions and ε_{ij} denotes a random error term with expectation 0 and variance $\sigma^2(x_{ij})$ ($i = 1, \dots, n, j = 1, \dots, m$) depending on x_{ij} . Observations on the same individual are assumed to be correlated, with $\text{corr}(\varepsilon_{ij}, \varepsilon_{ij*}) = c(x_{ij}, x_{ij*})$, while data corresponding to different individuals are assumed to be independent, i.e. $\text{corr}(\varepsilon_{ij}, \varepsilon_{i*j*}) = 0$, whenever $i \neq i^*$. We express the total covariance matrix of errors as the block diagonal matrix $V = \text{diag}(V_1, \dots, V_n) \in (\mathbb{R}^{m \times m})^n$, with matrices

$$V_i = \text{diag}(\sigma(x_{i1}), \dots, \sigma(x_{im}))(c(x_{ir}, x_{is}))_{r,s=1,\dots,m} \text{diag}(\sigma(x_{i1}), \dots, \sigma(x_{im})) \in \mathbb{R}^{m \times m}$$

on the diagonal. We now write $F_i = (f(x_{i1}), \dots, f(x_{im})) \in \mathbb{R}^{k \times m}$ as the design matrix for individual i , $i = 1, \dots, n$ and define the matrix $F = (F_1, \dots, F_n) = (f(x_{11}), \dots, f(x_{nm})) \in \mathbb{R}^{k \times nm}$ as the design matrix corresponding to all patients. The information matrix (proportional to the inverse of the covariance matrix) of the generalized least squares estimate for the parameter θ can be expressed as

$$M = FV^{-1}F^T = \sum_{i=1}^n F_i V_i^{-1} F_i^T. \quad (2.2)$$

The following arguments demonstrate that this expression can be rewritten in a form closer to the usual form of information matrices, which is obtained in the case of uncorrelated observations. We begin with an alternative representation for the individual information matrices $F_i V_i^{-1} F_i^T$, $i = 1, \dots, n$. For this purpose we collect all experimental conditions corresponding to one individual in a vector $x_i = (x_{i1}, \dots, x_{im}) \in \mathbb{R}^m$ and consider χ^m as a design space. An exact design is characterized by a tuple $(x_i, n_i)_{i=1}^p$, where $x_i \in \chi^m$ and $n_i \in \mathbb{N}$ such that $\sum_{i=1}^p n_i = n$. This means that n_i of the n patients are treated under the experimental condition $x_i = (x_{i1}, \dots, x_{im})^T$ ($i = 1, \dots, p$). Our first result provides the information matrix corresponding to one observation at the experimental condition x_i .

Lemma 2.1

An information matrix of the form $F_i V_i^{-1} F_i^T$ can also be expressed as

$$F_i V_i^{-1} F_i^T = \sum_{\ell=1}^m \tilde{f}_\ell(x_i) \tilde{f}_\ell(x_i)^T, \quad (2.3)$$

where the functions $\tilde{f}_\ell: \chi^m \rightarrow \mathbb{R}^k$ are defined as the columns of the $k \times m$ -matrix $F^* = F_i V_i^{-1/2}$.

Note that the vectors $\tilde{f}_\ell(x_i)$ in the representation (2.3) are defined implicitly (at least for $k \geq 4$) and can be easily calculated by a singular value decomposition of the matrices V_i . Using Lemma 2.1 the total information matrix for an exact design of m observations each on n subjects can therefore be written as

$$M = \sum_{i=1}^n \sum_{\ell=1}^m \tilde{f}_{\ell}(x_i) \tilde{f}_{\ell}(x_i)^T. \quad (2.4)$$

Following Kiefer (1974) we define an approximate design as a probability measure ζ on χ^m with finite support. Similarly to (2.4) the information matrix of an approximate design ζ using p different sets of m single subject measurements (with weights $\zeta(x_i) = \zeta(x_{i1}, \dots, x_{im})$ at the points x_i) can be expressed as

$$M(\xi) = \sum_{i=1}^p \sum_{\ell=1}^m \tilde{f}_{\ell}(x_i) \tilde{f}_{\ell}(x_i)^T \xi(x_i). \quad (2.5)$$

If ζ puts masses $\zeta_i = \zeta(x_i)$ at points $x_i \in \chi^m (i=1, \dots, p, \sum_{i=1}^p \xi_i=1)$ this means that approximately $n_i \approx n\zeta_i$ patients have to be treated under experimental conditions $x_i = (x_{i1}, \dots, x_{im})$ ($i = 1, \dots, p$). In practice the integers n_i are obtained by an appropriate rounding procedure from the quantities $n\zeta_i$ [see for example Pukelsheim and Rieder (1992)]. Note that the design space here is χ^m , i.e. the space of all possible m -observation sets.

Recall that for a given vector $c \in \mathbb{R}^k$ an approximate design ζ_c is called c -optimal if and only if $c \in \text{Range}(M(\zeta_c))$ and ζ_c minimizes the expression $c^T M^-(\zeta) c$, where $M^-(\zeta)$ denotes the generalized inverse of the matrix $M(\zeta)$ (note that this expression is approximately proportional to the variance of the generalized least squares estimate for the linear combination $c^T \theta$). We can now use the general theory developed in Pukelsheim (2006) and the representation (2.5) to derive a condition, which can be used to check the optimality of a given approximate design with respect to a given optimality criterion. In the special case of c -optimal designs, i.e. designs which are optimal for the estimation of a linear combination $c^T \theta$ of the parameters ($c \in \mathbb{R}^k$), we obtain the following result.

Theorem 2.1

A design ζ_c is c -optimal in a regression model with information matrix of the form (2.5) if and only if there exists a generalized inverse G of the matrix $M(\zeta_c)$ such that the inequality

$$\sum_{\ell=1}^m \frac{(c^T G \tilde{f}_{\ell}(x))^2}{c^T M^-(\xi_c) c} \leq 1 \quad (2.6)$$

holds for all $x \in \chi^m$. Moreover, there is equality in (2.6) at any support point of the design ζ_c .

We can use this theorem to apply Theorem 3.3 of Dette and Holland-Letz (2009) and derive a geometric characterization of c -optimal designs for models with information matrices of the form (2.5), which generalizes the classical result of Elfving (1952) to the case of dependent data. We define a generalized Elfving set by

$$\mathcal{R}_m = \text{conv} \left\{ \sum_{\ell=1}^m \varepsilon_{\ell} \tilde{f}_{\ell}(x) \mid x \in \mathcal{X}^m; \sum_{\ell=1}^m \varepsilon_{\ell}^2 = 1 \right\} \quad (2.7)$$

[note that the set $\mathcal{R}_m$ reduces for $m = 1$ to the classical Elfving space considered by Elfving (1952)]. We can now formulate our main theorem.

Theorem 2.2

A design $\xi_c = \{x_r, p_r\}_{r=1}^p$ is locally c -optimal in a model with information matrix of the form (2.5) if and only if there exist constants $\gamma > 0, \varepsilon_{11}, \dots, \varepsilon_{1p}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{mp}$ satisfying

$$\sum_{\ell=1}^m \varepsilon_{\ell r}^2 = 1; \quad r=1, \dots, p, \quad (2.8)$$

such that the point $\gamma c \in \mathbb{R}^k$ lies on the boundary of the generalized Elfving set $\mathcal{R}_m$ defined in (2.7) and has the representation

$$\gamma c = \sum_{r=1}^p p_r \left\{ \sum_{\ell=1}^m \varepsilon_{\ell r} \tilde{f}_{\ell}(x_r) \right\} \in \partial \mathcal{R}_m. \quad (2.9)$$

Remark 2.1

Theorem 2.1 and 2.2 can easily be generalized to the case of nonlinear fixed effects models of the form

$$Y_{ij} = \eta(x_{ij}, \theta) + \varepsilon_{ij} \quad i=1, \dots, n; j=1, \dots, m, \quad (2.10)$$

where η denotes a (not necessarily linear) function defined on $\mathcal{X} \times \Theta$. A rather detailed review and numerous references on optimal designs for nonlinear models can be found in Atkinson and Haines (1996). In the situation considered in this paper, standard results on nonlinear regression models show that the covariance matrix of the nonlinear generalized least squares can asymptotically be approximated by (2.2) where $F_i = (f(x_{i1}), \dots, f(x_{im})) \in \mathbb{R}^{k \times m}$ and the vector f is given by

$$f(x) = \frac{\partial}{\partial \theta} \eta(x, \theta) \quad (2.11)$$

Here the function f depends on the unknown parameter θ . Following Chernoff (1953) we assume that a preliminary guess for θ is available. In this case the information matrix in (2.4) is well defined and all results of this section remain correct for the nonlinear model (2.10). In particular locally c -optimal designs can be characterized by the equivalence Theorem 2.1 and the geometric characterization in Theorem 2.2, using the modified design matrix $F_i = (f(x_{i1}), \dots, f(x_{im}))$ with functions $f(x)$ defined in (2.11).

The concept of locally optimal designs has been criticized due to its sensitivity with respect to misspecification of the unknown parameter. Robust optimal designs could be obtained using a Bayesian or minimax approach [see e.g. Chaloner and Verdinelli (1995), Dette (1995), Müller and Pázman (1998)]. A geometric method of constructing Bayesian optimal designs for one-parameter models and a two-point prior distribution is given by Haines (1995) for the case of uncorrelated observations, but its generalization to models with more parameters, arbitrary prior distributions or correlated observations seems to be difficult. A generalization of Elfving's characterization to these more sophisticated criteria may be derived as a generalization of results of Dette (1996), who considered the case of uncorrelated observations. However, these investigations are extremely complicated and are deferred to future research.

3 Examples

We will demonstrate the application of the geometric characterization of Elfving type in two examples, a simple two parameter fixed effects polynomial model with intrinsically correlated observations and a nonlinear population model which is commonly used in pharmacokinetics.

3.1 Quadratic regression

As an example of a linear model we consider a two parameter fixed effects quadratic model without intercept, where m observations are taken for each of the n patients, that is

$$Y_{ij} = \theta_1 x_{ij} + \theta_2 x_{ij}^2 + \varepsilon_{ij} \quad i=1, \dots, n, j=1, \dots, m. \quad (3.1)$$

We begin with the case $m = 2$ and assume that observations corresponding to the same patients are correlated with covariance function $\text{cov}(\varepsilon_{i1}, \varepsilon_{i2}) = \sigma^2 c(x_{i1}, x_{i2}) = \sigma^2 \lambda^{|x_{i1} - x_{i2}|}$ $\lambda \in [0, 1]$. We obtain

$$V_i = \sigma^2 \begin{pmatrix} 1 & \lambda^{|x_{i1} - x_{i2}|} \\ \lambda^{|x_{i1} - x_{i2}|} & 1 \end{pmatrix}.$$

The eigenvalues of the matrix $V_i^{1/2}$ are given by $\sqrt{1 - \lambda^{|x_{i1} - x_{i2}|}}$ and $\sqrt{1 + \lambda^{|x_{i1} - x_{i2}|}}$ with corresponding eigenvectors $(1, -1)^T$ and $(1, 1)^T$, respectively. Therefore we obtain for the matrix $V_i^{-1/2}$ the representation

$$V_i^{-1/2} = \frac{1}{2\sigma} \begin{pmatrix} \frac{1}{\sqrt{1 - \lambda^{|x_{i1} - x_{i2}|}}} + \frac{1}{\sqrt{1 + \lambda^{|x_{i1} - x_{i2}|}}} & -\frac{1}{\sqrt{1 - \lambda^{|x_{i1} - x_{i2}|}}} + \frac{1}{\sqrt{1 + \lambda^{|x_{i1} - x_{i2}|}}} \\ -\frac{1}{\sqrt{1 - \lambda^{|x_{i1} - x_{i2}|}}} + \frac{1}{\sqrt{1 + \lambda^{|x_{i1} - x_{i2}|}}} & \frac{1}{\sqrt{1 - \lambda^{|x_{i1} - x_{i2}|}}} + \frac{1}{\sqrt{1 + \lambda^{|x_{i1} - x_{i2}|}}} \end{pmatrix}.$$

For the model (3.1) we have $f(x) = (x, x^2)^T$, $F_i = (f(x_{i1}), f(x_{i2}))$ and it follows from Lemma 2.1 that the functions $\tilde{f}_\ell(x_i)$ ($\ell = 1, 2$) are given by the columns of the matrix $F_i^* = F_i V_i^{-1/2}$ that is

$$\begin{aligned}\tilde{f}_1(x_i) &= \frac{1}{2\sigma} \left(\frac{x_{i1} - x_{i2}}{\sqrt{1 - \lambda|x_{i1} - x_{i2}|}} + \frac{x_{i1} + x_{i2}}{\sqrt{1 + \lambda|x_{i1} - x_{i2}|}}, \frac{x_{i1}^2 - x_{i2}^2}{\sqrt{1 - \lambda|x_{i1} - x_{i2}|}} + \frac{x_{i1}^2 + x_{i2}^2}{\sqrt{1 + \lambda|x_{i1} - x_{i2}|}} \right)^T, \\ \tilde{f}_2(x_i) &= \frac{1}{2\sigma} \left(\frac{x_{i2} - x_{i1}}{\sqrt{1 - \lambda|x_{i1} - x_{i2}|}} + \frac{x_{i2} + x_{i1}}{\sqrt{1 + \lambda|x_{i1} - x_{i2}|}}, \frac{x_{i2}^2 - x_{i1}^2}{\sqrt{1 - \lambda|x_{i1} - x_{i2}|}} + \frac{x_{i2}^2 + x_{i1}^2}{\sqrt{1 + \lambda|x_{i1} - x_{i2}|}} \right)^T.\end{aligned}$$

For the choice of parameters $\lambda = 0.6$, $\sigma^2 = 0.04$ and the design space $\chi = [0, 2]$ the corresponding generalized Elfving set $\mathcal{R}_2$ defined by (2.7) is depicted in Figure 1. Every pixel in the figure is induced by a measurement set $x \in \chi^m$ ($m = 2$), where the functions f_ℓ and the quantities ε_ℓ in (2.7) are evaluated at a (dense) grid. Note that a more dense grid would yield a smoother figure. Both parts of the figure represent the same Elfving space, but the coloring in the left part corresponds to potential values of the first measurement x_1 of $x = (x_1, x_2)$, while the coloring in the right part corresponds to the second measurement x_2 (see the legend of Figure 1).

Suppose we want to estimate the linear combination $c^T\theta$ defined by the vector $c = (-1, 1)^T$, which is marked as the red line in Figure 1. The optimal sets of measurements are those which can be used to construct the point of the intersection of the boundary of the Elfving space with the line in the direction of the vector c . This representation may require a single point of the form

$$p(x) = \sum_{\ell=1}^m \varepsilon_\ell \tilde{f}_\ell(x); \sum_{\ell=1}^m \varepsilon_\ell^2 = 1$$

or several points $p(x_1), \dots, p(x_p)$ of this type, where $p \leq k$ and k represent the number of parameters in the model (here $k = 2$). Each point $x_j = (x_{j1}, \dots, x_{jm}) \in \chi^m$ corresponds to a set of m measurements per patient (for model (3.1) we have $m = 2$). The weights used in the convex combination yield the weights of the optimal design, i.e. the proportions of total observations taken at the corresponding point x_j . The actual components x_{j1} and x_{j2} of the point x_j can be determined from the coloring of the point $p(x_j)$ in the left and right part of Figure 1, respectively. Thus, we can easily determine the support points graphically. For example, from Figure 1 we observe that two points, say x_1 and x_2 , are required to represent the boundary point γ_c , which are marked by two circles. From the left part of the Figure we obtain that the colour of x_1 is pink, while the colour of the second point is green, and from the legend in the right upper part of the figure we obtain the values $x_{11} = 0.0$ and $x_{21} = 1.2$ for the first components of x_1 and x_2 , respectively. Similarly, in the right part of Figure 1 we can observe blue and red colours for the two points, which yields $x_{12} = 0.8$ and $x_{22} = 2.0$ for the second components of x_1 and x_2 , respectively. Of course, in concrete applications the value of the components can be determined more precisely from the exact red/green/blue value of the corresponding pixel of the points in the representation (2.9) using appropriate software. Therefore the locally c -optimal design is given by

$$\xi_i = \begin{pmatrix} (0.0, 0.8) & (1.2, 2.00) \\ 0.48 & 0.52 \end{pmatrix} \quad (3.2)$$

and advises the experimenter to use two different individual measurement sets. This means that 48% of the patients should be treated at experimental conditions $x_{11} = 0$, $x_{12} = 0.8$ and 52% should be treated at $x_{21} = 1.2$ and $x_{22} = 2.0$. Note that it seems counterintuitive to use

measurements at the point $x = 0$, which carries no information in a model with uncorrelated observations. However, this heuristic argument is not necessarily true in the case of correlated observations.

Alternatively, we can use the figure to determine a hyperplane H supporting the Elfving space at the point γc . This plane is defined through a vector $d = (d_1, d_2)^T$ fulfilling $d^T z = 1$ for all $z \in H$, $(\gamma c)^T d = 1$ and $r^T d \leq 1$ for all $r \in \mathcal{R}_2$. The support points are then given as the solution of the system of equations

$$\max_{\varepsilon_1, \varepsilon_2} \sum_{\ell=1}^2 \varepsilon_{\ell} \tilde{f}_{\ell}(x_1, x_2) d = 1, \quad \sum_{i=1}^2 \varepsilon_i^2 = 1$$

[see the proof of Theorem 3.3 in Dette and Holland-Letz (2009)]. This yields an alternative derivation of the design (3.2). Note that the optimality of this design can be verified by Theorem 2.1.

We now suppose that $m = 3$ observations are available for each individual in the quadratic regression model (3.1). The design matrix is thus given by $F_i = (f(x_{i1}), f(x_{i2}), f(x_{i3}))$ and the functions $\tilde{f}_{\ell}(x_i)$, $\ell = 1, 2, 3$ can be determined similarly as in the case of $m = 2$ observations per individual.

The corresponding Elfving set is depicted in Figure 2. As $m = 3$ here, three subfigures are needed, each subfigure corresponds to one of the components $x_i = (x_{i1}, x_{i2}, x_{i3})$. We can observe that only one point is used in the Elfving representation (2.9) and we obtain by a similar reasoning as in the first part of this example that for $c = (-1, 1)^T$ the c -optimal design is given by

$$\xi_c = \begin{pmatrix} (0.0, 1.0, 2.0) \\ 1 \end{pmatrix}.$$

This means that all individuals have to be treated at experimental conditions 0, 1.0 and 2.0.

3.2 A nonlinear population model

In order to demonstrate the applicability of the methodology to population pharmacokinetic models, we consider a generic nonlinear random effects model

$$Y_{ij} = \eta(x_{ij}, b_i) + \varepsilon_{ij} \quad i=1, \dots, n, j=1, \dots, m, \quad (3.3)$$

where $\eta : \chi \times \mathbb{R}^k \rightarrow \mathbb{R}$ is a known function and the errors $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})$ for each patient are normally distributed with mean 0 and covariance matrix $W_i \in \mathbb{R}^{m \times m}$, $i = 1, \dots, n$. The quantities $b_1, \dots, b_n \sim \mathcal{N}(\theta, \Omega)$ denote k -dimensional independent normally distributed random variables with mean θ and covariance matrix Ω representing the effect of the corresponding subject under investigation [see Beatty and Piegorsch (1997), Ette et al. (1995), Cayen and Black (1993)]. We also assume that the random variables $b_1, \dots, b_n$ and the vector $(\varepsilon_{11}, \dots, \varepsilon_{nm})^T$ are independent.

Due to the nonlinearity of the model an explicit representation of the corresponding Fisher information matrix cannot be derived. Following Retout and Mentré (2003) we propose to use a first-order Taylor expansion to derive an approximation of this matrix. Assuming differentiability of the regression function we use the expansion

$$\eta(x, b) \approx \eta(x, \theta) + f(x, \theta)^T (b - \theta), \quad (3.4)$$

where

$$f(x, b) = \frac{\partial \eta(x, b)}{\partial b}$$

denotes the gradient of the regression function with respect to b . This means that similarly to the case of fixed effects nonlinear models (see Remark 2.1) the nonlinear model (3.3) is approximated by the linear model (3.4). For the construction of the optimal design we assume that knowledge about the parameters θ and Ω is available from previous or similar experiments and consider the determination of locally optimal designs [see Chernoff (1953)]. As a consequence, the covariance matrix of the nonlinear least squares estimate in the model (3.3) is approximated by replacing the function f in model (2.1) with $f(x) = f(x, b)|_{b=\theta}$. The variance of the random vector $Y_i = (Y_{i1}, \dots, Y_{im})$ now includes the variance caused by the random effect and can be approximated by

$$\text{Var}(Y_i) = V_i \approx F_i^T \Omega F_i + W_i \quad i=1, \dots, n.$$

Consider for example the simple first order elimination model with two observations for each subject ($b_i = (b_{i1}, b_{i2})$)

$$Y_{ij} = b_{i1} e^{-b_{i2} x_{ij}} + \varepsilon_{ij}, \quad x_{ij} \in \mathcal{X} = [0, 2], i=1, \dots, n, \quad j=1, 2, \quad (3.5)$$

which is widely used in pharmacokinetics [see e.g. Rowland (1993)]. We assume that the errors ε_{ij} are homoscedastic and uncorrelated with variance $\sigma^2 > 0$, that is $V_i \approx F_i^T \Omega F_i + \sigma^2 I_m$ and for the parameters we consider the case

$$\theta = (5, 0.8), \quad \Omega = \text{diag}(1, 0.1) \quad \text{and} \quad \sigma^2 = 0.04.$$

A straightforward calculation shows that

$$\frac{\partial \eta(x, \theta)}{\partial \theta} = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T.$$

Therefore, we have $f(x) = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T$ and the functions $\tilde{f}_1, \tilde{f}_2$ are defined in a similar manner as it is illustrated in Example 1.

The corresponding generalized Elfving set is depicted in Figure 3. If we are interested in the optimal design for estimating the area under the curve

$$AUC = \int_0^\infty \theta_1 e^{-\theta_2 x} dx = \frac{\theta_1}{\theta_2},$$

it is easy to see that this objective corresponds to a locally c -optimal design problem for the vector $c = (1/\theta_2, -\theta_1/\theta_2^2)^T$, which is marked as the red line in Figure 3. From this Figure it can be seen that only one point is needed in the Elfving representation (2.9), and we obtain by a similar reasoning as in Section 3.1 that the locally c -optimal design for the estimation of the area under the curve is given by

$$\xi_c = \begin{pmatrix} (0.0, 2.0) \\ 1 \end{pmatrix}.$$

This design means that all patients should be treated under experimental conditions $x_1 = 0.0$ and $x_2 = 2.0$. The optimality of this design can also be verified by Theorem 2.1.

Acknowledgments

The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG), the BMBF Project SKAVOE and the NIH grant award IR01GM072876:01A1. The third author was partially supported by EPSRC grant EP/D048893/1. The authors would also like to thank three anonymous referees and the associate editor for some very constructive comments which led to substantial improvements of an earlier version of this manuscript.

4 Appendix: Proofs of main results

4.1 Proof of Lemma 2.1

Let $V_i = U_i D_i U_i^T$ denote the eigenvalue decomposition of the positive semidefinite matrix V_i , with U_i the matrix of eigenvectors and $D_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{im})$ the diagonal matrix of eigenvalues of V_i . Then the matrix

$$V_i^{-1/2} = U_i \text{diag}(\lambda_{i1}^{-1/2}, \dots, \lambda_{im}^{-1/2}) U_i^T$$

is the root of V_i^{-1} , i.e. $V_i^{-1/2} V_i^{-1/2} = V_i^{-1}$, and we obtain

$$F_i V_i^{-1} F_i^T = F_i V_i^{-1/2} V_i^{-1/2} F_i^T.$$

Defining

$$F_i^* := F_i V_i^{-1/2} \in \mathbb{R}^{k \times m}$$

and writing $\tilde{f}_\ell(x_i)$ for the ℓ -th column of F_i^* , we have

$$F_i V_i^{-1} F_i^T = F_i^* (F_i^*)^T = \sum_{i=1}^m \tilde{f}_\ell(x_i) \tilde{f}_\ell(x_i)^T,$$

which completes the proof of Lemma 2.1.

4.2 Proof of Theorem 2.1

Let Ξ denote the set of all approximative designs on χ^m and let

$$\mathcal{M} = \{M(\xi) | \xi \in \Xi\} \subset \mathbb{R}^{k \times k}$$

denote the set of all information matrices of the form (2.5). The set $\mathcal{M}$ is convex and the information matrix $M(\xi_c)$ of a locally c -optimal design for which the linear combination $c^T \theta$ is estimable [i.e. $c \in \text{Range}(M(\xi))$] maximizes the function $(c^T M^- c)^{-1}$ in the set $\mathcal{M} \cap \mathcal{A}_c$, where

$$\mathcal{A}_c = \{M(\xi) \in \mathcal{M} | c \in \text{Range}(M(\xi))\}.$$

Consequently we obtain from Theorem 7.19 in Pukelsheim (2006) that the design ξ_c is c -optimal if and only if there exists a generalized inverse, say G , of the matrix $M(\xi_c)$ such that the inequality

$$\text{tr}(AGc c^T G^T) \leq c^T M^-(\xi_c) c$$

holds for all $A \in \mathcal{M}$, where there is equality for any matrix $A \in \mathcal{M}$ which maximizes $(c^T M^- c)^{-1}$ in the set $\mathcal{M}$. Note that the family $\mathcal{M}$ is the convex hull of the set

$$\left\{ \sum_{\ell=1}^m \tilde{f}_\ell(x) \tilde{f}_\ell(x)^T \mid x \in \chi^m \right\},$$

and therefore the assertion of Theorem 2.1 follows by a standard argument of optimal design theory [see e.g. Silvey (1980)].

4.3 Proof of Theorem 2.2

Recall that the information matrix at the experimental condition $x = (x_1, \dots, x_m)$ is of the form

$$I(x) = \sum_{\ell=1}^m \tilde{f}_{\ell}(x) \tilde{f}_{\ell}(x)^T. \quad (4.1)$$

Therefore, the result is a direct consequence of Theorem 3.3 in Dette and Holland-Letz (2009), which presents a geometric characterization of Elfving type for c -optimal designs in models with an information matrix of the form (4.1).

References

- Atkinson AC, Chaloner K, Herzberg AM, Juritz J. Optimum experimental designs for properties of a compartmental model. *Biometrics*. 1993; 49:325–337. [PubMed: 8369370]
- Atkinson, AC.; Haines, LM. Designs for nonlinear and generalized linear models. In: Ghosh, S.; Rao, CR., editors. *Handbook of Statistics 13, Design and Analysis of Experiments*. North-Holland Publishing Co.; Amsterdam: 1996. p. 437–475.
- Beatty DA, Piegorsch WW. Optimal statistical design for toxicokinetic studies. *Statistical Methods in Medical Research*. 1997; 6:359–376. [PubMed: 9447654]
- Bickel PJ, Herzberg AM. Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann Statist*. 1979; 7(1):77–95.
- Cayen, M.; Black, H. Role of toxicokinetics in dose selection for carcinogenicity studies. In: Welling, P.; de la Iglesia, F., editors. *Drug toxicokinetics*. Marcel Dekker; New York, N. Y.: 1993.
- Chaloner K, Verdinelli I. Bayesian experimental design: A review. *Statistical Science*. 1995; 10:273–304.
- Chernoff H. Locally optimal designs for estimating parameters. *Ann Math Statist*. 1953; 24:586–602.
- Colombo S, Buclin T, Cavassini M, Decosterd L, Telenti A, Biollaz J, Csajka C. Population pharmacokinetics of atazanavir in patients with human immunodeficiency virus infection. *Antimicrobial Agents and Chemotherapy*. 2006; 50
- Dette H. Elfving's Theorem for D -optimality. *Annals of Statistics*. 1993a; 21:753–766.
- Dette H. A new interpretation of optimality for E -optimal designs in linear regression models. *Metrika*. 1993b; 40:37–50.
- Dette H. Designing of experiments with respect to “standardized” optimality criteria. *Journal of the Royal Statistical Society, Ser B*. 1995; 59:97–110.
- Dette H. A note on bayesian c - and D -optimal designs in nonlinear regression models. *Annals of Statistics*. 1996; 24:1225–1234.
- Dette H, Holland-Letz T. A geometric characterization of c -optimal designs for heteroscedastic regression. *Ann Statist*. 2009; 37(6B):4088–4103.
- Elfving G. Optimal allocation in linear regression theory. *Annals of Mathematical Statistics*. 1952; 23:255–262.
- Ette E, Kelman A, Howie C, Whiting B. Analysis of animal pharmacokinetic data: Performance of the one point per animal design. *Journal of Pharmacokinetics and Biopharmaceutics*. 1995; 23:551–566. [PubMed: 8733946]
- Fedorov, VV. *Theory of Optimal Experiments*. Academic Press; New York: 1972.
- Ford I, Torsney B, Wu CFJ. The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Ser B*. 1992; 54:569–583.
- Haines LM. A geometric approach to optimal design for one-parameter non-linear models. *Journal of the Royal Statistical Society, Series B*. 1995; 57(3):575–598.
- Kiefer J. General equivalence theory for optimum designs. *Annals of Statistics*. 1974; 2:849–879.
- Müller CH, Pázman A. Applications of necessary and sufficient conditions for maximum efficient designs. *Metrika*. 1998; 48:1–19.
- Müller W, Pázman A. Measures for designs in experiments with correlated errors. *Biometrika*. 2003; 90(2):423–434.
- Näthner W. Exact design for regression models with correlated errors. *Statistics*. 1985; 16(4):479–484.

- Pukelsheim, F. Optimal Design of Experiments, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics; Philadelphia, PA: 2006.
- Pukelsheim F, Rieder S. Efficient rounding of approximate designs. *Biometrika*. 1992; 79:763–770.
- Retout S, Mentré F. Further developments of the Fisher information matrix in nonlinear mixed-effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*. 2003; 13:209–227. [PubMed: 12729390]
- Rowland, M. Clinical Pharmacokinetics: Concepts and Applications. Williams and Wilkins; Baltimore: 1993.
- Sacks J, Ylvisaker ND. Designs for regression problems with correlated errors; many parameters. *Ann Math Statist*. 1968; 39:49–69.
- Schmelter T. Considerations on group-wise identical designs for linear mixed models. *Journal of Statistical Planning and Inference*. 2007; 137:4003–4010.
- Sheiner L, Rosenberg B, Marathe V. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of Pharmacokinetics and Biopharmaceutics*. 1977; 5:445–479. [PubMed: 925881]
- Silvey, SD. Optimal Design. Chapman and Hall; London: 1980.
- Studden WJ. Elfving's theorem revisited. *Journal of Statistical Planning and Inference*. 2005; 130:85–94.

Figure 1.

The Elfving set $\mathcal{R}_2$ defined in (2.7) for a quadratic regression model (3.1) with two observations per patient. The axes d_1 and d_2 represent the first and second dimension of the two dimensional elements of this set. The vector c is depicted by the red line, while the two black circles denote the points used in the Elfving representation (2.9).

Figure 2.

The Elfving set $\mathcal{R}_3$ defined in (2.7) for the quadratic regression model (3.1) with 3 observations per patient. The axes d_1 and d_2 represent the first and second dimension of the two dimensional elements of this set. The vector c is depicted by the red line, while the black circle shows the point used in the Elfving representation (2.9).

Figure 3.

The Elfving set $\mathcal{R}_2$ defined in (2.7) for the first order elimination model with 2 observations per patient. The axes $d1$ and $d2$ represent the first and second dimension of the two dimensional elements of this set. The vector c is depicted by the red line.