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

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

Andrey Pepelyshev
University of Sheffield
Department of Probability & Statistics
Sheffield, U.K.
email: a.pepelyshev@sheffield.ac.uk

28 July 2009

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 can be 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 in 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, mixed models, locally optimal design, geometric characterization
AMS Subject Classification: 62K05

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 [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 an 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 uncorrelated case the optimal design problem for dependent data is intrinsically more difficult. Most authors use asymptotic arguments to determine efficient designs [see Sacks and Ylvisaker (1968), Bickel and Herzberg (1979), Näther (1985), Dette et al. (2009), Müller and Pázman (2003) among others]. In particular for the case of dependent data the powerful equivalence theorem [Pukelsheim (1993)] and geometric representations [Elfving (1952)] from the uncorrelated case are not available.

In the present paper we try to fill this gap for the c-optimality criterion, which determines the design such that the variance of a linear combination of the parameters (specified by the vector c) is minimal. Note that many commonly used criteria (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) 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 $\tilde{f}_l$, $l = 1, ..., m(m+1)/2$ such that the total Fisher information matrix for the set of m observations on this patient can be written in the form

$$I(\theta) = \sum_{l=1}^{m(m+1)/2} u_l \tilde{f}_l \tilde{f}_l^T, \tag{1.1}$$

where the quantities u_l can take the values of $-1, 0, 1$ only. For this representation we introduce in addition to the original design space for the individual observations a design space of m observations for each patient. Using this representation, we can derive an equivalence theorem for c-optimal designs using the general theory in Pukelsheim (1993) and apply recent results of Dette and Holland-Letz (2009) to 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.

2 An Elfving representation for models with correlated observations

We begin our discussion with the linear case where the results are slightly more transparent. The nonlinear case can easily be reduced to this situation (see Remark 2.4), 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

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

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

$$V_i = diag(\sigma_1, ..., \sigma_m)(c(x_{ir}, x_{is}))_{r,s=1,...,m}\ diag(\sigma_1, \ldots, \sigma_m) \in \mathbb{R}^{m\times m}$$

on the diagonal. We now write $F_i = (f(x_{i1}), ..., f(x_{im})) \in \mathbb{R}^{k\times m}$ as the design matrix for individual $i, i = 1, ..., n$, define the matrix $F = (F_1, \ldots, F_n) = (f(x_{11}), ..., f(x_{nm})) \in \mathbb{R}^{k\times nm}$ as the design matrix corresponding to all patients, and denote by $v_{lj}(x_i)$ the Element in the position (l, j) of the matrix V_i^{-1}. The information matrix (inverse covariance matrix) of the weighted least squares estimate for the parameter θ can be expressed as

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

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

$F_i V_i^{-1} F_i^T$, $i = 1, ..., n$. For this purpose we collect all experimental conditions corresponding to one individual in a vector $x_i = (x_{i1}, ..., x_{im}) \in \mathbb{R}^m$ and consider $\mathcal{X}^m$ as design space. An exact design is characterized by a tuple $(x_i, n_i)_{i=1}^n$, where $x_i \in \mathcal{X}^m$ 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 $x_i = (x_{i1}, ..., x_{im})^T \quad (i = 1, ..., p)$. 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*

$$F_i V_i^{-1} F_i^T = \sum_{l=1}^{m(m+1)/2} u_l \tilde{f}_l(x_i) \tilde{f}_l(x_i)^T \ , \tag{2.3}$$

where the functions $\tilde{f}_l : \mathcal{X}^m \to \mathbb{R}^k$ and the constants $u_l \in \{-1, 0, 1\}$ are defined in equations (2.7) and (2.8) below, respectively, $l = 1, ..., m(m+1)/2$.

Proof: Let $V_i^{-1} = (v_{lj}(x_i))_{l,j=1}^m$ denote the inverse of the matrix V_i, then a straightforward calculation yields

$$\begin{aligned} F_i V_i^{-1} F_i^T &= \sum_{l=1}^m \sum_{j=1}^m f(x_{il}) f(x_{ij})^T v_{lj}(x_i) \\ &= \sum_{l=1}^m \sum_{j>l}^m [f(x_{il}) + sgn(v_{lj}(x_i)) f(x_{ij})][f(x_{il}) + sgn(v_{lj}(x_i)) f(x_{ij})]^T |v_{lj}(x_i)| \\ &\quad + \sum_{l=1}^m f(x_{il}) f(x_{il})^T (v_{ll}(x_i) - \sum_{j \neq l}^m |v_{lj}(x_i)|) \\ &= \sum_{l=1}^m \sum_{j>l}^m g_{lj}(x_i) g_{lj}(x_i)^T + \sum_{l=1}^m s_l h_l(x_i) h_l(x_i)^T, \end{aligned}$$

where the functions $s_l : \mathcal{X}^m \to \{-1, 0, 1\}$ and $g_{lj}, h_l : \mathcal{X}^m \to \mathbb{R}^k$ are defined by

$$s_l(x_i) = s_l(x_{i1}, ..., x_{im}) = sgn(v_{ll}(x_i)) - \sum_{j \neq l}^m |v_{lj}(x_i)|), \qquad l = 1, \ldots, m \tag{2.4}$$

$$g_{lj}(x_i) = g_{lj}(x_{i1}, ..., x_{im}) = (f(x_{il}) + sgn(v_{lj}(x_i)) f(x_{ij})) \sqrt{|v_{lj}(x_i)|}, \quad l, j = 1, \ldots, m \tag{2.5}$$

and

$$h_l(x_i) = h_l(x_{i1}, ..., x_{im}) = f(x_{il}) \sqrt{|v_{ll}(x_i) - \sum_{j \neq l}^m |v_{lj}(x_i)||}, \qquad l = 1, \ldots, m, \tag{2.6}$$

respectively. With the notation

$$\tilde{f}_l = \begin{cases} h_l & \text{if } l = 1, ..., m \\ g_{1,l-m+1} & \text{if } l = m+1, ..., 2m-1 \\ g_{2,l-2m+3} & \text{if } l = 2m, ..., 3m-3 \\ \vdots & \vdots \\ g_{m-1,m} & \text{if } l = m(m+1)/2 \end{cases} \tag{2.7}$$

and

$$u_l = \begin{cases} s_l & \text{if } l = 1, ..., m \\ 1 & \text{if } l = m+1, ..., m(m+1)/2 \end{cases} \tag{2.8}$$

we can express the information matrix as

$$F_i V_i^{-1} F_i^T = \sum_{l=1}^{m(m+1)/2} u_l \tilde{f}_l(x_i) \tilde{f}_l(x_i)^T, \tag{2.9}$$

which completes the proof of Lemma 2.1. □

Using Lemma 2.1 the total information matrix for an exact design of m observations each on n subjects can therefore be written as

$$M = \sum_{i=1}^{n} \sum_{l=1}^{m(m+1)/2} u_l \tilde{f}_l(x_i) \tilde{f}_l(x_i)^T. \tag{2.10}$$

Following Kiefer (1974) we define an approximate design as a probability measure ξ on $\mathcal{X}^m$ with finite support. Similarly to (2.10) the information matrix of an approximate design ξ using p different sets of m single subject measurements (with weights $\xi(x_i) = \xi(x_{i1}, ..., x_{im})$ at the points x_i) can be expressed as

$$M(\xi) = \sum_{i=1}^{p} \sum_{l=1}^{m(m+1)/2} u_l \tilde{f}_l(x_i) \tilde{f}_l(x_i)^T \xi(x_i). \tag{2.11}$$

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

Recall that for a given vector $c \in \mathbb{R}^k$ an approximate design ξ_c is called c-optimal if and only if $c \in$ Range $(M(\xi_c))$ and ξ_c minimizes the expression $c^T M^-(\xi) c$, where $M^-(\xi)$ denotes the generalized inverse of the matrix $M(\xi)$ (note that this expression is approximately proportional

to the variance of the weighted least squares estimate for the linear combination $c^T\theta$). We can now use the representation (2.11) to derive a condition, which can be used to check the optimality of a given approximate design. 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 ξ_c is c-optimal in a regression model with information matrix of the form (2.11) if and only if there exists a generalized inverse G of the matrix $M(\xi_c)$ such that the inequality*

$$\sum_{\ell=1}^{m(m+1)/2} u_\ell \frac{\left(c^T G \tilde{f}_\ell(x)\right)^2}{c^T M^-(\xi_c,\theta)c} \leq 1 \tag{2.12}$$

holds for all $x \in \mathcal{X}^m$. Moreover, there is equality in (2.12) at any support point of the design ξ_c.

Proof of Theorem 2.1. Let Ξ denote the set of all approximative designs on $\mathcal{X}^m$ and let

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

denote the set of all information matrices of the form (2.11). $\mathcal{M}$ is obviously 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. $c \in$ Range $(M(\xi))$] maximizes the function $(c^T M^- c)^{-1}$ in the set $\mathcal{M} \cap \mathcal{A}_c$, where

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

Consequently it follows from Theorem 7.19 in Pukelsheim (1993) 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

$$\text{tr}(AGcc^TG^T) \leq c^T M^-(\xi_c)c$$

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

$$\left\{ \sum_{\ell=1}^{m(m+1)/2} u_\ell \tilde{f}_\ell(x)\tilde{f}_\ell^T(x) \,\Big|\, x \in \mathcal{X}^m \right\},$$

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

If it can be shown that $u_\ell \geq 0$ for all $\ell = 1, ..., m(m+1)/2$ 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 (2.11), which generalizes the

classical result of Elfving (1952) to the case of dependent data. For this purpose we define a generalized Elfving set by

$$\mathcal{R}_{(m(m+1)/2)} = \text{conv}\left\{ \sum_{\ell=1}^{m(m+1)/2} \varepsilon_\ell \tilde{f}_\ell(x\theta) \,\Big|\, x \in \mathcal{X}^m; \sum_{\ell=1}^{m(m+1)/2} \varepsilon_\ell^2 = 1 \right\} \tag{2.13}$$

[note that the set $\mathcal{R}_{(m(m+1)/2)}$ reduces for $m = 1$ to the classical Elfving space considered by Elfving (1952)]. Additionally, we assume that the quantities u_ℓ defined in (2.8) are nonnegative for all $\ell = 1 \dots m(m+1)/2$, i.e.

$$v_{ll}(x_i) - \sum_{j \neq l}^{m} |v_{lj}(x_i)| \geq 0, \qquad l = 1, \dots, m, \qquad i = 1, \dots, n. \tag{2.14}$$

Theorem 2.2 *Assume that (2.14) is satisfied. A design $\xi_c = \{x_r, p_r\}_{r=1}^p$ is locally c-optimal in a model with information matrix of the form (2.11) if and only if there exist constants $\gamma > 0$, $\varepsilon_{11}, \dots, \varepsilon_{1p}, \dots, \varepsilon_{(m(m+1)/2)1}, \dots, \varepsilon_{(m(m+1)/2)p}$ satisfying*

$$\sum_{\ell=1}^{m(m+1)/2} \varepsilon_{\ell r}^2 = 1 \; ; \qquad r = 1, \dots, p, \tag{2.15}$$

such that the point $\gamma c \in \mathbb{R}^k$ lies on the boundary of the generalized Elfving set $\mathcal{R}_{(m(m+1)/2)}$ defined in (2.13) and has the representation

$$\gamma c = \sum_{r=1}^{p} p_r \left\{ \sum_{\ell=1}^{(m(m+1)/2)} \varepsilon_{\ell r} \tilde{f}_\ell(x_r) \right\} \in \partial\, \mathcal{R}_{(m(m+1)/2)}. \tag{2.16}$$

Proof. From assumption (2.14) it follows that $u_\ell \in \{0, 1\}$ and consequently the information matrix at the experimental condition $x = (x_1, \dots, x_m)$ is of the form

$$I(x) = \sum_{\{\ell | u_\ell = 1\}} \tilde{f}_\ell(x) \tilde{f}_\ell(x)^T. \tag{2.17}$$

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 (2.17). □

Remark 2.3 If $m = 2$ observations can be taken for each patient and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then assumption (2.14) is always satisfied, because

$$V_i^{-1} = \frac{1}{\sigma^2 |V_i|} \begin{pmatrix} 1 & -c(x_{i1}, x_{i2}) \\ -c(x_{i1}, x_{i2}) & 1 \end{pmatrix} \tag{2.18}$$

and $|c(x_{i1}, x_{i2})| \leq 1$.

Remark 2.4 The results can easily be generalized to nonlinear fixed effects models of the form

$$Y_{ij} = \eta(x_{ij}, \theta) + \epsilon_{ij} \quad i = 1, .., n; j = 1, ..., m, \tag{2.19}$$

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 weighted least squares is asymptotically given by (2.2) where $F_i = (f(x_{i1}), ..., f(x_{im})) \in \mathbb{R}^{k \times m}$ and the vector f is given by

$$f(t) = \frac{\partial}{\partial \theta} \eta(t, \theta) \tag{2.20}$$

Following Chernoff (1953) we assume that a preliminary guess for the unknown parameter θ is available. In this case the information matrix in (2.10) is well defined and all results of this section remain correct for the nonlinear model (2.19) using the identification (2.20). In particular locally c-optimal designs can be characterized by the appropriately modified equivalence Theorem 2.1 and the geometric characterization in Theorem 2.2.

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 uncorrelated case, 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 could be derived along the lines of Dette (1996), who considered the uncorrelated case, i.e. $m = 1$. However, these investigations are extremely complicated and will be devoted to future research.

3 Examples

We will demonstrate the application of the geometric characterization of Elfving type in two examples, a simple 2 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 a linear example we consider a two parameter fixed effects quadratic model, where m observations are taken for each of the n patients, that is

$$y_{ij} = \theta_1 x_{ij} + \theta_2 x_{ij}^2 + \epsilon_{ij} \quad i = 1, ..., n, j = 1, ..., m. \tag{3.1}$$

We begin with the case $m = 2$ and assume that observations corresponding to the same patients are correlated with covariance function $cov(\epsilon_{i1}, \epsilon_{i2}) = \sigma^2 c(x_{i1}, x_{i2}) = \sigma^2 \lambda^{|x_{i1}-x_{i2}|}$, i.e.

$$V_i^{-1} = \sigma^{-2} \begin{pmatrix} 1 & \lambda^{|x_{i1}-x_{i2}|} \\ \lambda^{|x_{i1}-x_{i2}|} & 1 \end{pmatrix}^{-1} = \begin{pmatrix} v_{11}(x_i) & v_{12}(x_i) \\ v_{21}(x_i) & v_{22}(x_i) \end{pmatrix}.$$

In this situation we have $f(x) = (x, x^2)^T$ and by Remark 2.3 the assumption (2.14) is satisfied, which yields $u_\ell = 1 \quad (\ell = 1, 2, 3)$. Consequently, the information matrix can be written as a sum of $m(m+1)/2 = 3$ terms using the functions

$$\tilde{f}_1(x_{i1}, x_{i2}) = h_1(x_{i1}, x_{i2}) = f(x_{i1}) \sqrt{|v_{11}(x_i) - |v_{12}(x_i)||} \tag{3.2}$$

$$\tilde{f}_2(x_{i1}, x_{i2}) = h_2(x_{i1}, x_{i2}) = f(x_2) \sqrt{|v_{22}(x_i) - |v_{12}(x_i)||} \tag{3.3}$$

$$\tilde{f}_3(x_{i1}, x_{i2}) = g_{12}(x_{i1}, x_{i2}) = (f(x_{i1}) + sgn(v_{12}(x_i)) f(x_{i2})) \sqrt{|v_{12}(x_i)|}. \tag{3.4}$$

For the choice of parameters $\lambda = 0.6$, $\sigma^2 = 0.04$ and the design space $\mathcal{X} = [0, 2]$ the corresponding generalized Elfving set $\mathcal{R}_3$ defined by (2.13) is depicted in Figure 1. Every pixel in the figure is induced by a point measurement set $x \in \mathcal{X}^m$ ($m = 2$), where the function $\tilde{f}_\ell$ and the quantities ε_ℓ in (2.13) are evaluated at a (dense) grid. 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

$$p(x) = \sum_{\ell=1}^{m(m+1)/2} \epsilon_\ell \tilde{f}_\ell(x); \quad \sum_{\ell=1}^{m(m+1)/2} \epsilon_\ell^2 = 1$$

or several points $p(x_1), ..., p(x_p)$ of this type, where $p \leq k$ and k represent the number of parameters in the model (here $k = 2$). Each point $x_j = (x_{j1}, ..., x_{jm}) \in \mathcal{X}^m$ corresponds to a set of measurements (in the concrete example we have $m = 2$) per patient. 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 γc, 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, the right part of Figure 1 yields the colours blue and red 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. Therefore the locally c-optimal design advises the experimenter to use two different individual measurement sets, that is:

$$\xi_c = \begin{pmatrix} (0.0, 0.8) & (1.2, 2.00) \\ 0.48 & 0.52 \end{pmatrix}. \tag{3.5}$$

This means that 48% of the patients are treated at experimental conditions $x_{11} = 0, x_{12} = 0.8$ and 52% are treated at $x_{21} = 1.2$ and $x_{22} = 2.0$. Note that in concrete applications the value of the components can be determined from the exact red/green/blue value of the corresponding pixel of the points in the representation (2.16) using appropriate graphic software.

Alternatively, we can use the figure to determine any 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 $r \in \mathcal{R}_3$. The support points are then given as the solution of the system of equations

$$\max_{\epsilon_1,\ldots,\epsilon_3} \sum_{l=1}^{3} \epsilon_l \tilde{f}_l(x_1, x_2) d = 1, \quad \sum_{i=1}^{3} \epsilon_i^2 = 1$$

[see the proof of Theorem 3.3 in Dette and Holland-Letz (2009)]. This yields an alternative derivation of the design (3.5). In both cases the optimality of this design can also 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). In this case we have $m(m+1)/2 = 6$ and, writing $x_i = (x_{i1}, x_{i2}, x_{i3})$, the functions $\tilde{f}_l$ used in the representation (2.10) are given by

$$\begin{aligned}
\tilde{f}_1(x_i) = h_1(x_i) &= f(x_{i1}) \sqrt{|v_{11}(x_i) - |v_{12}(x_i)| - |v_{13}(x_i)||}, \\
\tilde{f}_2(x_i) = h_2(x_i) &= f(x_{i2}) \sqrt{|v_{22}(x_i) - |v_{21}(x_i)| - |v_{23}(x_i)||}, \\
\tilde{f}_3(x_i) = h_3(x_i) &= f(x_{i3}) \sqrt{|v_{33}(x_i) - |v_{31}(x_i)| - |v_{32}(x_i)||}, \\
\tilde{f}_4(x_i) = g_{12}(x_i) &= (f(x_{i1}) + sgn(v_{12}(x_i)) f(x_{i2})) \sqrt{|v_{12}(x_i)|}, \\
\tilde{f}_5(x_i) = g_{13}(x_i) &= (f(x_{i1}) + sgn(v_{13}(x_i)) f(x_{i3})) \sqrt{|v_{13}(x_i)|}, \\
\tilde{f}_6(x_i) = g_{23}(x_i) &= (f(x_{i2}) + sgn(v_{23}(x_i)) f(x_{i3})) \sqrt{|v_{23}(x_i)|},
\end{aligned}$$

where $f(x) = (x, x^2)^T$ denotes the vector of regression functions and the matrix $V_i^{-1} = (v_{lk}(x_i))_{l,k=1}^3$ is defined by

$$V_i^{-1} = \sigma^{-2} \begin{pmatrix} 1 & \lambda^{|x_{i1}-x_{i2}|} & \lambda^{|x_{i1}-x_{i3}|} \\ \lambda^{|x_{i2}-x_{i1}|} & 1 & \lambda^{|x_{i2}-x_{i3}|} \\ \lambda^{|x_{i3}-x_{i1}|} & \lambda^{|x_{i3}-x_{i2}|} & 1 \end{pmatrix}^{-1}.$$

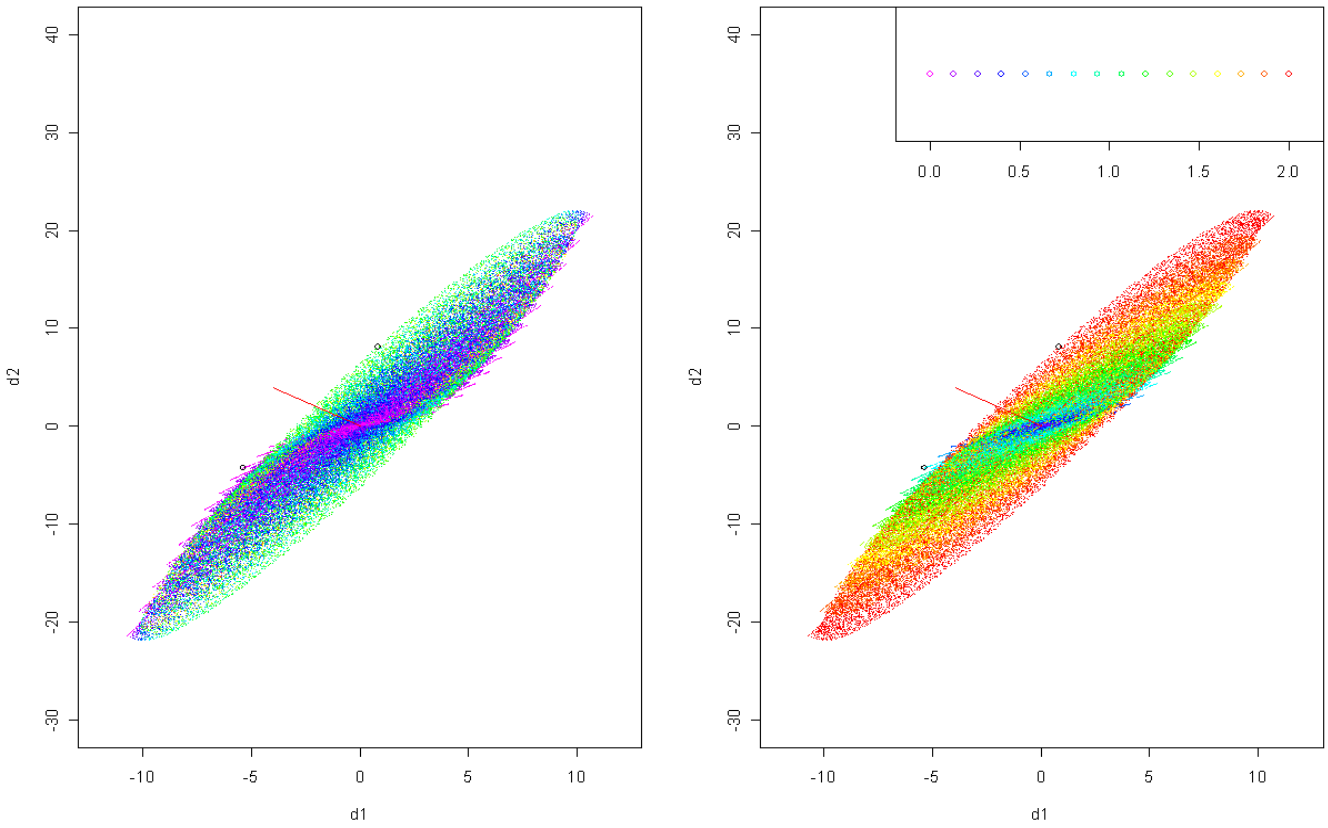

Figure 1: *The Elfving set $\mathcal{R}_3$ defined in (2.13) for a quadratic regression model (3.1) with two observations per patient. The functions $\tilde{f}_1$, $\tilde{f}_2$ and $\tilde{f}_3$ are given by (3.2), (3.3) and (3.4), respectively. The vector c is depicted by the red line, while the two black circles denote the points used in the Elfving representation (2.16).*

We can verify that assumption (2.14) is satisfied for this correlation structure (note that this is not the case for all correlation matrices if $m > 2$).

The corresponding Elfving set is depicted in Figure 2. As $m = 3$ here, three subfigures are needed, each corresponding to one of the components $x_i = (x_{i1}, x_{i2}, x_{i3})$. In this case only one point is used in the Elfving representation (2.16) 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.

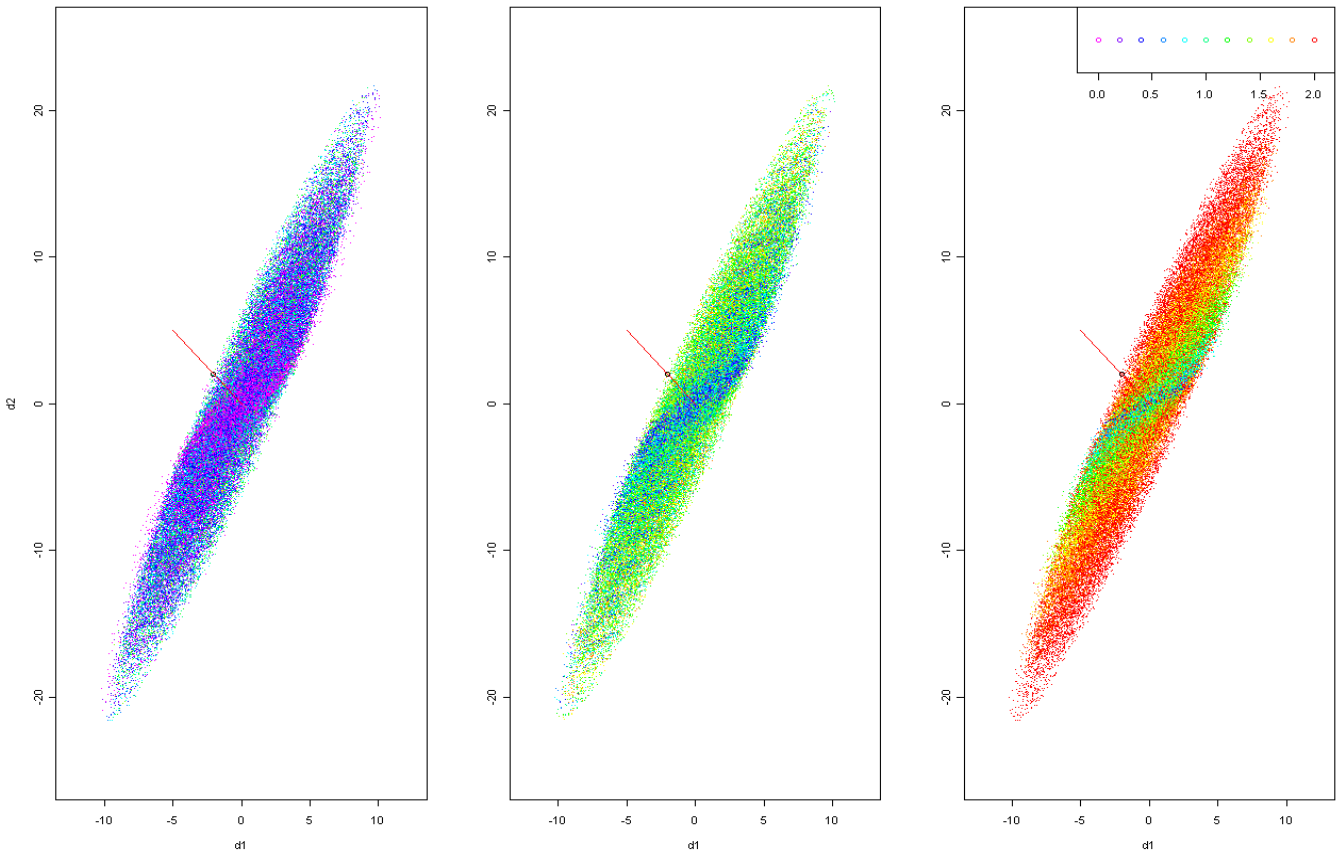

Figure 2: *The Elfving set $\mathcal{R}_6$ defined in (2.13) for the quadratic regression model (3.1) with 3 observations per patient. The vector c is depicted by the red line, while the black circle shows the point used in the Elfving representation (2.16).*

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, i.e.

$$Y_{ij} = \eta(x_{ij}, b_i) + \varepsilon_{ij} \quad i = 1, \ldots, n, j = 1, \ldots, m, \tag{3.6}$$

where $\eta : \mathcal{X} \times \mathbb{R}^k \to \mathbb{R}$ is a known function and the errors $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{im})$ for each patient are normally distributed with mean 0 and covariance matrix $W_i \in \mathbb{R}^{m \times m}$, $i = 1, ..., n$. The quantities $b_1, \ldots b_n \sim \mathcal{N}(\theta, \Omega)$ 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 $b_1, \ldots, b_n$ and the vector $(\varepsilon_{11}, \ldots, \varepsilon_{nm})^T$ 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

$$\eta(x,b) \approx \eta(x,\theta) + f(x,\theta)(b-\theta)^T, \tag{3.7}$$

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 2.4) the nonlinear model (3.6) is approximated by the linear model (3.7). For the construction of the design we assume that knowledge about the parameter θ 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 (3.6) is approximated by replacing the functions f in model (2.1) with $f(x) = f(x,b)|_{b=\theta}$. The variance of the random vector $Y_i = (Y_{i1}, ..., Y_{im})$ now includes the variance caused by the random effect and can be approximated by

$$\mathrm{Var}(Y_i) = V_i \approx F_i^T \Omega F_i + W_i \qquad i = 1, \ldots, n.$$

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

$$Y_{ij} = b_{i1} e^{-b_{i2} x_{ij}} + \varepsilon_{ij}, \qquad x_{ij} \in \mathcal{X} = [0,2], i = 1, ...n, j = 1, 2\,, \tag{3.8}$$

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

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

A straightforward calculation shows that

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

Therefore, we have $f(x) = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})$ and the three functions $\tilde{f}_1, \tilde{f}_2, \tilde{f}_3$ are defined in a similar manner as illustrated in Example 1. Moreover, it can be easily checked that assumption (2.14) is satisfied. The detailed calculations are omitted for the sake of brevity.
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, i.e.

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

it is easy to see that this 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 (2.16), 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 means that all patients should be treated under experimental conditions $x_1 = 0$ and $x_2 = 2$. The optimality of this design can also be verified by Theorem 2.1.

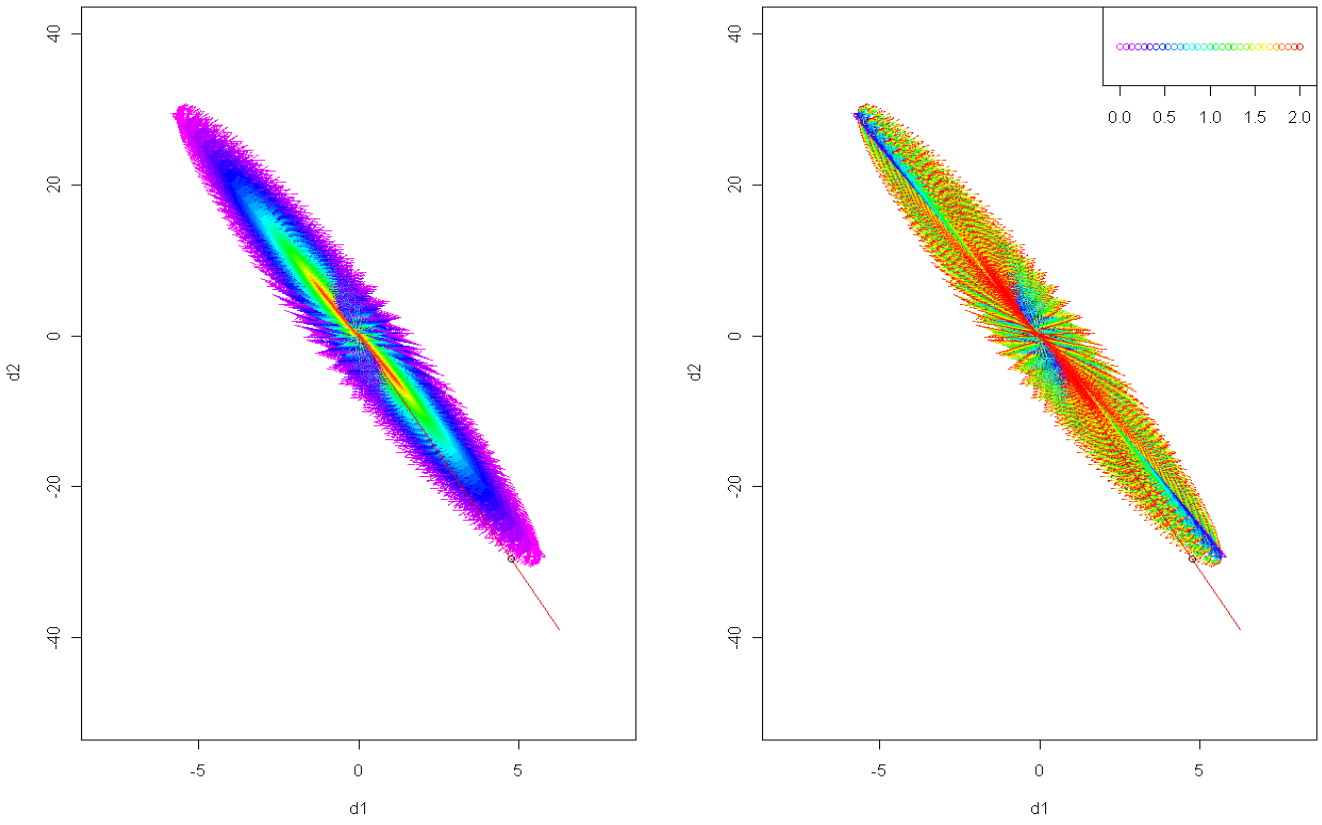

Figure 3: *The Elfving space $\mathcal{R}_3$ defined in (2.13) for the first order elimination model with 2 observations per patient.*

Acknowledgements 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.

References

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49:325–337.

Atkinson, A. C. and Haines, L. M. (1996). Designs for nonlinear and generalized linear models. In Ghosh, S. and Rao, C. R., editors, *Handbook of Statistics 13, Design and Analysis of Experiments*, pages 437–475. North-Holland Publishing Co., Amsterdam.

Beatty, D. A. and Piegorsch, W. W. (1997). Optimal statistical design for toxicokinetic studies. *Statistical Methods in Medical Research*, 6:359–376.

Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann. Statist.*, 7(1):77–95.

Cayen, M. and Black, H. (1993). *Role of toxicokinetics in dose selection for carcinogenicity studies.* Welling, P., de la Iglesia, F: Drug toxicokinetics. Marcel Dekker, New York, N. Y.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science.*, 10:273–304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.*, 24:586–602.

Colombo, S., Buclin, T., Cavassini, M., Decosterd, L., Telenti, A., Biollaz, J., and Csajka, C. (2006). Population pharmacokinetics of atazanavir in patients with human immunodeficiency virus infection. *Antimicrobial Agents and Chemotherapy*, 50.

Dette, H. (1995). Designing of experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

Dette, H. (1996). A note on bayesian c- and D-optimal designs in nonlinear regression models. *Annals of Statistics*, 24:1225–1234.

Dette, H. and Holland-Letz, T. (2009). A geometric characterization of c-optimal designs for heteroscedastic regression. *To appear in: Annals of Statistics.*

Dette, H., Pepelyshev, A., and Holland-Letz, T. (2009). Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics. *Submitted for publication.*

Elfving, G. (1952). Optimal allocation in linear regression theory. *Annals of Mathematical Statistics*, 23:255–262.

Ette, E., Kelman, A., Howie, C., and Whiting, B. (1995). Analysis of animal pharmacokinetic data: Performance of the one point per animal design. *Journal of Pharmacokinetics and Biopharmaceutics*, 23:551–566.

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.

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

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

Müller, W. and Pázman, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika*, 90(2):423–434.

Näther, W. (1985). Exact design for regression models with correlated errors. *Statistics*, 16(4):479–484.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. John Wiley & Sons, New York.

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

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.

Rowland, M. (1993). *Clinical Pharmacokinetics: Concepts and Applications.* Williams and Wilkins, Baltimore.

Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.*, 39:49–69.

Schmelter, T. (2007). Considerations on group-wise identical designs for linear mixed models. *Journal of Statistical Planning and Inference*, 137:4003–4010.

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