A geometric characterization of c-optimal designs for heteroscedastic regression

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Tim Holland-Letz

Ruhr-Universität Bochum

Medizinische Fakultät

44780 Bochum, Germany

email: tim.holland-letz@rub.de

November 21, 2008

Abstract

We consider the common nonlinear regression model where the variance as well as the mean is a parametric function of the explanatory variables. The c-optimal design problem is investigated in the case when the parameters of both the mean and the variance function are of interest. A geometric characterization of c-optimal designs in this context is presented, which generalizes the classical result of Elfving (1952) for c-optimal designs. As in Elfving's famous characterization c-optimal designs can be described as representations of boundary points of a convex set. However, in the case where there appear parameters of interest in the variance, the structure of the Elfving set is different. Roughly speaking the Elfving set corresponding to a heteroscedastic regression model is the convex hull of a set of ellipsoids induced by the underlying model and indexed by the design space. The c-optimal designs are characterized as representations of the points where the line in direction of the vector c intersects the boundary of the new Elfving set. The theory is illustrated in several examples including pharmacokinetic models with random effects.

Keyword and Phrases: c-optimal design, heteroscedastic regression, Elfving's theorem, pharmacokinetic models, random effects, locally optimal design, geometric characterization

AMS Subject Classification: 62K05

1 Introduction

Nonlinear regression models are widely used to describe the relation between several variables [see Seber and Wild (1989), Ratkowsky (1983, 1990)]. Because of the broad applicability of such models in many fields the problem of constructing optimum experimental designs for these models has found considerable interest in the literature. Early work has been done by Chernoff (1953) who introduced the concept of locally optimal designs. These designs require an initial guess of the unknown parameters of the model and are used as benchmarks for many commonly used designs. Locally optimum designs for nonlinear models have been discussed by numerous authors [see Ford et al. (1992), Box and Lucas (1959), Haines (1993), Haines (1995), Biedermann et al. (2006), López-Fidalgo and Wong (2002), Dette et al. (2004) among many others]. Most of the relevant literature discusses the design problem under the additional assumption of a constant variance but much less work has been done for models with an heteroscedastic error structure. Wong and Cook (1993) studied G-optimal designs for models, when heteroscedasticity is present in the data. A systematic approach to optimal design problems for heteroscedastic linear models was given by Atkinson and Cook (1995), who derived the necessary information matrices in the case where the variance as well as the mean depend on the parameters of the model and the explanatory variables. For other work on optimal designs for specific regression models with heteroscedastic errors and various optimality criteria we refer to King and Wong (1998), Ortiz and Rodríguez (1998), Fang and Wiens (2000), Brown and Wong (2000), Montepiedra and Wong (2001) and Atkinson (2008) among others.

The present paper is devoted to the local c-optimality criterion which determines the design such that a linear combination of the unknown parameters (specified by the vector c) has minimal variance [see e.g. Pukelsheim (1981), Pázman and Pronzato (2008)] . Under the assumption of homoscedasticity there exists a beautiful geometric characterization of locally c-optimal designs which is due to Elfving (1952) who considered the case of a linear model. Roughly speaking, the characterization of Elfving (1952) is possible, because the Fisher information can be represented in the form

$$(1.1) \qquad I(x, \theta) = f(x, \theta) f^T(x, \theta),$$

where $f(x, \theta)$ denotes a vector of functions corresponding to the particular model under consideration and θ is the vector of unknown parameters. The c-optimal design with support points $x_1, \ldots, x_m$ and weights $p_1, \ldots, p_m$ can be characterized as a representation of the point

$$(1.2) \qquad \lambda c = \sum_{r=1}^{m} p_r \varepsilon_r f(x_r, \theta)$$

where $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, 1\}$ and $\lambda > 0$ denotes a scaling factor such that the point λc is a boundary point of the Elfving set

(1.3) $$\mathcal{R}_1 = \text{conv} \{\varepsilon f(x, \theta) \mid x \in \mathcal{X}, \ \varepsilon \in \{-1, 1\}\}.$$

Here $\mathcal{X}$ denotes the design space, $\text{conv}(\mathcal{A})$ is the convex hull of a set $\mathcal{A} \subset \mathbb{R}^p$. This result has been applied by numerous authors to derive c-optimal designs in linear and non-linear regression models [see e.g. Studden (1968), Han and Chaloner (2003), Ford et al. (1992), Chernoff and Haitovsky (1990), Fan and Chaloner (2003), Dette et al. (2008) among many others]. For a review on Elfving's theorem we also refer to Fellmann (1999) and to the recent work of Studden (2005). On the other hand, in the case of heteroscedasticity, where both the mean and the variance depend on the explanatory variables and parameters of interest, the information matrix is usually of the form

(1.4) $$I(x, \theta) = \sum_{j=1}^{k} f_j(x, \theta) f_j^T(x, \theta)$$

[see Atkinson and Cook (1995)] with $k \geq 2$. Here $f_1, \ldots, f_k$ represent certain features of the nonlinear regression model under consideration (see Section 2 for more details) and a geometric characterization of the Elfving type is not available. It is the purpose of the present paper to fill this gap and to provide a useful characterization of c-optimal designs for regression models with Fisher information of the form (1.4).

In Section 2 we introduce the basic notation of a heteroscedastic nonlinear regression model and demonstrate that the Fisher information matrix has in fact the form (1.4). Section 3 contains our main results. We derive a general equivalence theorem for locally c-optimal designs in nonlinear regression models with Fisher information matrix given by (1.4). These results are used to derive a geometric characterization of locally c-optimal designs which generalize the classical result of Elving to regression models, where the mean as well as the variance depend on the explanatory variables and the parameters of interest. The corresponding Elving space has the same dimension as a classical Elving space but consists of the convex hull of a family of ellipsoids induced by the underlying model and indexed by the design space (see Section 3 for details). In the special case $k = 1$ the generalized Elfving set reduces to the classical set considered by Elfving (1952) and a similar result is obtained if the variance of the regression is a function of the mean. Finally, in Section 4 we illustrate the geometric characterization in several examples. In particular we demonstrate the applicability of the results in the Michaelis-Menten model with exponentially decreasing variance and in a nonlinear random effect model used in toxicokinetics, where the Fisher information has a similar structure as in (1.4).

2 c-optimal designs for heteroscedastic regression models

Consider the common nonlinear regression model where at a point x a response Y is observed with expectation

$$(2.1) \qquad \mathrm{E}[Y|x] = \mu(x, \theta)$$

and variance

$$(2.2) \qquad \mathrm{Var}[Y|x] = \sigma^2(x, \theta).$$

Here $\theta \in \mathbb{R}^p$ denotes the vector of unknown parameters. Note that we do not exclude the case, where μ and σ^2 depend on different subsets of the parameter vector, that is $\theta = (\theta_1, \theta_2)$, $\mu(x, \theta) = \mu(x, \theta_1)$; $\sigma^2(x, \theta) = \sigma^2(x, \theta_2)$. We assume that n independent observations $Y_1, \ldots, Y_n$ are available under experimental conditions $x_1, \ldots, x_n \in \mathcal{X}$, where $\mathcal{X}$ denotes the design space. We define $\mu_\theta = (\mu(x_1, \theta), \ldots, \mu(x_n, \theta))^T$ as the vector of the expected responses, $\sigma_\theta^2 = (\sigma^2(x_1, \theta), \ldots, \sigma^2(x_n, \theta))^T$ as the vector of variances and

$$(2.3) \qquad \Sigma_\theta = \mathrm{diag}(\sigma^2(x_1, \theta), \ldots, \sigma^2(x_n, \theta))$$

as the covariance matrix of the random vector $Y = (Y_1, \ldots, Y_n)^T$. Under the additional assumption of a normal distribution the Fisher information of Y is given by the $k \times k$ matrix

$$(2.4) \qquad I = \frac{d\mu_\theta^T}{d\theta} \Sigma_\theta^{-1} \frac{d\mu_\theta}{d\theta} + \frac{1}{2} \frac{d\sigma_\theta^2}{d\theta}^T \Sigma_\theta^{-2} \frac{d\sigma_\theta^2}{d\theta}.$$

An approximate experimental design is a discrete probability measure with masses $w_1, \ldots, w_m$ at points $x_1, \ldots, x_m \in \mathcal{X}$. These points define the distinct experimental conditions at which observations have to be taken and $w_1, \ldots, w_m > 0$, $\sum_{j=1}^m w_j = 1$ are positive weights representing the proportions of total observations taken at the corresponding points [see Silvey (1980), Atkinson and Donev (1992), Pukelsheim (1993), Randall et al. (2007)]. If N observations can be taken a rounding procedure is applied to obtain integers r_j from the not necessarily integer valued quantities $w_j N$ $(j = 1, \ldots, m)$ [see Pukelsheim and Rieder (1992)]. If the assumption of a normal distribution is made the analogue of the Fisher information matrix for an approximate design is the matrix

$$(2.5) \qquad M(\xi, \theta) = \int_{\mathcal{X}} I(x, \theta) d\xi(x) \in \mathbb{R}^{p \times p},$$

where

$$(2.6) \quad I(x, \theta) = \frac{1}{\sigma^2(x, \theta)} \left(\frac{\partial \mu(x, \theta)}{\partial \theta} \right)^T \frac{\partial \mu(x, \theta)}{\partial \theta} + \frac{1}{2\sigma^4(x, \theta)} \left(\frac{\partial \sigma^2(x, \theta)}{\partial \theta} \right)^T \frac{\partial \sigma^2(x, \theta)}{\partial \theta}$$

denotes the Fisher information at the point x [see also Atkinson and Cook (1995) or Atkinson (2008)].

Under some additional assumptions of regularity it can be shown that the asymptotic covariance matrix of the maximum likelihood estimate for the parameter θ is given by

$$(2.7) \qquad \frac{1}{N} M^{-1}(\xi, \theta)$$

[see Jennrich (1969)]. An optimal design for estimating the parameter θ minimizes an appropriate function of this matrix and numerous criteria have been proposed for this purpose [see Silvey (1980), Pukelsheim (1993) or Randall et al. (2007)]. In this paper we consider the local c-optimality criterion, which determines the design ξ such that the expression

$$(2.8) \qquad c^T M^-(\xi, \theta) c,$$

is minimal, where $c \in \mathbb{R}^p$ is a given vector, the minimum is calculated among all designs for which $c^T \theta$ is estimable, that is $c \in \mathrm{Range}(M(\xi, \theta))$, and $M^-(\xi, \theta)$ denotes a generalized inverse of the matrix $M(\xi, \theta)$. The expression (2.8) is approximately proportional to the variance of the maximum likelihood estimate for the linear combination $c^T \theta$. It is shown in Pukelsheim (1993) that the expression (2.8) does not depend on the specific choice of the generalized inverse if the vector c is estimable by the design ξ. Note that the criterion (2.8) is a local optimality criterion in the sense that it requires the specification of the unknown parameter θ [see Chernoff (1953)]. Locally optimal designs are commonly used as benchmarks for given designs used in applications. Moreover, in some cases knowledge about the parameter θ is available from previous studies and locally c-optimal designs are robust with respect to misspecifications of θ [see Dette et al. (2008)]. For robustifications of locally optimal designs the interested reader is referred to the work of Chaloner and Verdinelli (1995), Dette (1995) or Müller and Pázman (1998) among many others.

Note that the Fisher information (2.4) is of the form

$$(2.9) \qquad I(x, \theta) = \sum_{\ell=1}^{k} f_\ell(x, \theta) f_\ell^T(x, \theta) ,$$

where $k = 2$ and

$$(2.10) \qquad f_1(x, \theta) = \frac{1}{\sigma(t, \theta)} \left(\frac{\partial \mu(t, \theta)}{\partial \theta} \right)^T \;;\; f_2(x, \theta) = \frac{1}{\sqrt{2}\sigma^2(t, \theta)} \left(\frac{\partial \sigma^2(t, \theta)}{\partial \theta} \right)^T .$$

Because the formulation of our main results does not yield any additional complication in the subsequent discussion, we consider in the following models with a Fisher information of the form

(2.9). We call a design ξ_c minimizing $c^T M^-(\xi, \theta) c$ a locally c-optimal for a model with Fisher information matrix of the form (2.9), where the minimum is taken over the class of all designs for which $c^T \theta$ is estimable. The corresponding statements for the heteroscedastic nonlinear regression models are then derived as special cases.

3 Elfving's Theorem for heteroscedastic models

The following result allows an easy verification of c-optimality for a given design.

Theorem 3.1. *A design ξ_c is locally c-optimal in a model with Fisher information matrix of the form (2.9) if and only if there exists a generalized inverse G of the matrix $M(\xi_c, \theta)$ such that the inequality*

(3.1)
$$\sum_{\ell=1}^{k} \left(\frac{c^T G f_\ell(x)}{c^T M^-(\xi, \theta) c} \right)^2 \leq 1$$

holds for all $x \in \mathcal{X}$. Moreover, there is equality in (3.1) at any support point of the design ξ_c.

Proof of Theorem 3.1. Let Ξ denote the set of all approximative designs on $\mathcal{X}$ and for fixed θ let
$$\mathcal{M} = \{M(\xi, \theta) \mid \xi \in \Xi\} \subset \mathbb{R}^{p \times p}$$
denote the set of all information matrices of the form (2.5), where $I(x, \theta) = \sum_{\ell=1}^{k} f_\ell(x, \theta) f_\ell^T(x, \theta)$. $\mathcal{M}$ is obviously convex and the information matrix $M(\xi_c, \theta)$ of a locally c-optimal design for which the linear combination $c^T \theta$ is estimable (i.e. $c \in \text{Range} \ (M(\xi, \theta))$ maximizes the function $(c^T M^- c)^{-1}$ in the set $\mathcal{M} \cap \mathcal{A}_c$, where $\mathcal{A}_c = \{M(\xi, \theta)) \in \mathcal{M} \mid c \in \text{Range}(M(\xi, \theta))\}$. 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 $M(\xi_c, \theta)$ such that the inequality
$$\text{tr}(AGcc^T G^T) \leq c^T M^-(\xi_c, \theta) c$$
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
$$\left\{ \sum_{\ell=1}^{k} f_\ell(x, \theta) f_\ell^T(x, \theta) \ \middle| \ x \in \mathcal{X} \right\},$$
and therefore the assertion of Theorem 3.1 follows by a standard argument of optimal design theory [see e.g. Silvey (1980)]. □

The following result gives the corresponding statement for the case of the nonlinear heteroscedastic regression model introduced in Section 2. It is a direct consequence of Theorem 3.1 observing that in this case we have $k = 2$ and the two functions f_1 and f_2 are given by (2.10).

Corollary 3.2. *Under the assumption of a normally distributed error a design ξ_c is c-optimal for the nonlinear heteroscedastic regression model (2.1) and (2.2) if and only if there exists a generalized inverse G of the matrix $M(\xi_c, \theta)$ such that the inequality*

$$(3.2) \qquad \frac{1}{\sigma^2(x,\theta)}\left(\frac{\partial\mu(x,\theta)}{\partial\theta}Gc\right)^2 + \frac{1}{2\sigma^4(x,\theta)}\left(\frac{\partial\sigma^2(x,\theta)}{\partial\theta}Gc\right)^2 \le c^T M^-(\xi_c,\theta)c$$

holds for all $x \in \mathcal{X}$. Moreover, there is equality in (3.2) at any support point of the design ξ_c.

In the following discussion we will use the equivalence Theorem 3.1 to derive a geometric characterization of locally c-optimal designs in nonlinear regression models with Fisher information of the form (2.9), which generalizes the classical result of Elving in an interesting direction. For this purpose we define a generalized Elving set by

$$(3.3) \qquad \mathcal{R}_k = \mathrm{conv}\left\{\sum_{j=1}^{k}\varepsilon_j f_j(x,\theta) \;\middle|\; x \in \mathcal{X}; \sum_{j=1}^{k}\varepsilon_j^2 = 1\right\}$$

and obtain the following result.

Theorem 3.3. *A design $\xi_c = \{x_r, p_r\}_{r=1}^{m}$ is locally c-optimal in a model with Fisher information matrix of the form (2.9) if and only if there exist constants $\gamma > 0$, $\varepsilon_{1,1}, \ldots, \varepsilon_{1,m}, \ldots, \varepsilon_{k,1}, \ldots, \varepsilon_{k,m}$ satisfying*

$$(3.4) \qquad \sum_{\ell=1}^{k}\varepsilon_{\ell r}^2 = 1 \;; \qquad r = 1, \ldots, m$$

such that the point $\gamma c \in \mathbb{R}^p$ lies on the boundary of the generalized Elfving set $\mathcal{R}_k$ defined in (3.3) and has the representation

$$(3.5) \qquad \gamma c = \sum_{r=1}^{m} p_r \left\{\sum_{\ell=1}^{k}\varepsilon_{\ell r}f_\ell(x_r)\right\} \in \partial\,\mathcal{R}_k.$$

Proof. Assume that the design $\xi_c = \{x_r; p_r\}_{r=1}^{m}$ minimizes $c^T M^-(\xi,\theta)c$ or equivalently satisfies Theorem 3.1. We define $\gamma^2 = (c^T Gc)^{-1}$, $d = \gamma Gc \in \mathbb{R}^p$, then it follows from $c \in \mathrm{Range}(M(\xi,\theta))$

and the representation (2.9) that

$$\gamma c = M(\xi_c, \theta)d \;=\; \sum_{r=1}^{m} p_r \left\{ \sum_{\ell=1}^{k} f_\ell(x_r, \theta) f_\ell^T(x_r, \theta) \right\} d$$

$$=\; \sum_{r=1}^{m} p_r \left\{ \sum_{\ell=1}^{k} f_\ell(x_r, \theta) \varepsilon_{\ell r} \right\},$$

where we have used the notation $\varepsilon_{\ell r} = f_\ell^T(x_r)d$ $(\ell = 1, \ldots, k; \; r = 1, \ldots, m)$. By Theorem 3.1 there is equality in (3.1) for each support point x_r, which implies (observing the definition of γ and $\varepsilon_{\ell r}$)

$$\sum_{\ell=1}^{k} \varepsilon_{\ell r}^2 = 1 \; ; \qquad r = 1, \ldots, m.$$

Consequently, the conditions (3.4) and (3.5) are satisfied and it remains to show that $\gamma c \in \partial \mathcal{R}_k$. For this purpose we note that we have by Cauchy's inequality and Theorem 3.1

$$(3.6) \qquad \left(d^T \left\{ \sum_{\ell=1}^{k} \varepsilon_\ell f_\ell(x, \theta) \right\} \right)^2 \le \sum_{\ell=1}^{k} (d^T f_\ell(x, \theta))^2 \sum_{\ell=1}^{k} \varepsilon_\ell^2 \le \sum_{\ell=1}^{k} \varepsilon_\ell^2 = 1$$

for all $x \in \mathcal{X}$ and $\varepsilon_1, \ldots, \varepsilon_\ell$ satisfying (3.4). Moreover, $\gamma c^T d = 1$ by the definition of the constant γ and the vector d. Consequently, the vector d defines a supporting hyperplane to the generalized Elfving set $\mathcal{R}_k$ at the point γc, which implies $\gamma c \in \partial \mathcal{R}_k$

In order to prove the converse assume that $\gamma c \in \partial \mathcal{R}_k$ and that (3.4) and (3.5) are satisfied. In this case there exists a supporting hyperplane, say $d \in \mathbb{R}^p$, to the generalized Elfving set $\mathcal{R}_k$ at the boundary point γc, that is

$$(3.7) \qquad\qquad\qquad \gamma c^T d = 1 \, ,$$

and the inequality

$$(3.8) \qquad\qquad\qquad \left| d^T \left\{ \sum_{\ell=1}^{k} \varepsilon_\ell f_\ell(x, \theta) \right\} \right| \le 1$$

holds for all $x \in \mathcal{X}$, $\varepsilon_1, \ldots, \varepsilon_k$ satisfying (3.4). Defining

$$\varepsilon_\ell(x) = \frac{d^T f_\ell(x, \theta)}{\sqrt{\sum_{j=1}^{k} (d^T f_j(x, \theta))^2}} \, , \qquad \ell = 1, \ldots, k$$

we have $\sum_{\ell=1}^{k} \varepsilon_\ell^2(x) = 1$ and obtain from the inequality (3.8)

$$(3.9) \qquad\qquad \sum_{\ell=1}^{k} (d^T f_\ell(x))^2 = \left| \sum_{\ell=1}^{k} d^T \{ \varepsilon_\ell(x) f_\ell(x, \theta) \} \right|^2 \le 1$$

8

for all $x \in \mathcal{X}$. On the other hand, it follows from (3.5) and (3.7) that

$$1 = \gamma c^T d = \sum_{r=1}^{m} p_r d^T \left\{ \sum_{\ell=1}^{k} \varepsilon_{\ell r} f_\ell(x_r, \theta) \right\},$$

which implies (using the inequality (3.9)) for $x_1, \ldots, x_m$

$$(3.10) \qquad d^T \left\{ \sum_{\ell=1}^{k} \varepsilon_{\ell r} f_\ell(x_r, \theta) \right\} = 1 \qquad r = 1, \ldots, m.$$

Consequently, we obtain using the Cauchy-Schwarz inequality

$$1 = \left(\sum_{\ell=1}^{k} \varepsilon_{\ell r} d^T f_\ell(x_r, \theta) \right)^2 \leq \sum_{\ell=1}^{k} \varepsilon_{\ell r}^2 \sum_{\ell=1}^{k} \left(d^T f_\ell(x_r, \theta) \right)^2 \leq 1$$

for each $r = 1, \ldots, m$, which gives

$$(3.11) \qquad \varepsilon_{\ell r} = \lambda_r d^T f_\ell(x_r, \theta) \qquad \ell = 1, \ldots, k; \ r = 1, \ldots, m$$

for some constants $\lambda_1, \ldots, \lambda_m$. Now inserting (3.11) in (3.10) yields

$$1 = \lambda_r \sum_{\ell=1}^{k} \left(d^T f_\ell(x_r, \theta) \right)^2 = \lambda_r \qquad r = 1, \ldots m,$$

and combining (3.11) and (3.5) gives

$$\gamma c = \sum_{r=1}^{m} p_r \left\{ \sum_{\ell=1}^{k} f_\ell(x_r, \theta) f_\ell^T(x_r, \theta) \right\} d = M(\xi_c, \theta) d.$$

It follows from Searle (1982) that there exists a generalized inverse of the matrix $M(\xi_c, \theta)$ such that $d = \gamma G c$ and we have from (3.9) for all $x \in \mathcal{X}$

$$(3.12) \qquad \gamma^2 \sum_{\ell=1}^{k} (c^T G f_\ell(x))^2 \leq 1.$$

By (3.7) we have $\gamma^2 = c^T G c = c^T M^{-1}(\xi_c, \theta) c$ and the inequality (3.12) reduces to (3.1). Consequently the c-optimality of the design $\xi_c = \{x_r; p_r\}_{r=1}^{m}$ follows from Theorem 3.1. $\qquad \square$

Note that in the case $k = 1$ the generalized Elfving set $\mathcal{R}_k$ defined in (3.3) reduces to the classical Elfving set (1.3) introduced by Elfving (1952). Similarly, Theorem 3.3 reduces to the classical Elfving Theorem for models in which the Fisher information can be represented by $I(x, \theta) = f_1(x, \theta) f_1^T(x, \theta)$ [see Pukelsheim (1993), Chapter 2, or Studden (2005)]. In the following

9

Corollary we specify the geometric characterization of Theorem 3.3 in the special case of nonlinear heteroscedastic regression models where we have $k = 2$ and the functions f_1 and f_2 are given by (2.10).

Corollary 3.4. *Consider the nonlinear heteroscedastic regression model defined by (2.1) and (2.2) with the additional assumption of a normal distribution, and define*

$$(3.13) \quad \mathcal{R} \quad = \quad conv \left\{ \frac{\varepsilon_1}{\sigma(x,\theta)} \left(\frac{\partial \mu(x,\theta)}{\partial \theta} \right)^T + \frac{\varepsilon_2}{\sqrt{2}\sigma^2(x,\theta)} \left(\frac{\partial \sigma^2(x,\theta)}{\partial \theta} \right)^T \Big| x \in \mathcal{X}, \; \varepsilon_1^2 + \varepsilon_2^2 = 1 \right\}.$$

A design $\xi_c = \{x_r; p_r\}_{r=1}^m$ is locally c-optimal if and only if there exist constants $\gamma > 0$ and $\varepsilon_{1,1}, \ldots, \varepsilon_{1,m}, \varepsilon_{2,1}, \ldots, \varepsilon_{2,m}$ satisfying $\varepsilon_{1r}^2 + \varepsilon_{2r}^2 = 1 \; (r = 1, \ldots, m)$ such that the point

$$\gamma c = \sum_{r=1}^m p_r \left\{ \frac{\varepsilon_{1r}}{\sigma(x_r,\theta)} \left(\frac{\partial \mu(x_r,\theta)}{\partial \theta} \right)^T + \frac{\varepsilon_{2r}}{\sqrt{2}\sigma^2(x_r,\theta)} \left(\frac{\partial \sigma^2(x_r,\theta)}{\partial \theta} \right)^T \right\}$$

is a boundary point of the set $\mathcal{R}$.

Remark 3.5. Consider the case where mean and variance in the regression model are related by a known link function, say ℓ, that is

$$(3.14) \qquad\qquad \sigma^2(x,\theta) = \ell(\mu(x,\theta)).$$

Under this assumption a straightforward calculation shows that the Elfving set (3.13) reduces to

$$\mathcal{R} \quad = \quad conv \left\{ \frac{\varepsilon_1}{\sqrt{\ell(\mu(x,\theta))}} \left(\frac{\partial \mu(x,\theta)}{\partial \theta} \right)^T + \frac{\varepsilon_2 \ell'(\mu(x,\theta))}{\sqrt{2\ell^2(\mu((x,\theta))}} \left(\frac{\partial \mu(x,\theta)}{\partial \theta} \right)^T \Big| x \in \mathcal{X}, \; \varepsilon_1^2 + \varepsilon_2^2 = 1 \right\}$$

$$= \quad conv \left\{ \left(\frac{\varepsilon_1}{\sqrt{\ell(\mu(x,\theta))}} + \frac{\varepsilon_2 \ell'(\mu(x,\theta))}{\sqrt{2\ell^2(\mu(x,\theta))}} \right) \left(\frac{\partial \mu(x,\theta)}{\partial \theta} \right)^T \Big| x \in \mathcal{X}, \; \varepsilon_1^2 + \varepsilon_2^2 = 1 \right\}.$$

Now using the values

$$\varepsilon_1 = \varepsilon \frac{1}{\sqrt{1 + \omega^2(x)}} \;; \quad \varepsilon_2 = \varepsilon \frac{\omega(x)}{\sqrt{1 + \omega^2(x)}} \;; \quad \omega(x) = \frac{\ell'(\mu(x,\theta))}{\sqrt{2\ell(\mu(x,\theta))}}$$

for some $\varepsilon \in \{-1, 1\}$ it can be easily shown that $\mathcal{R}_1 \subset \mathcal{R}$, where the set $\mathcal{R}_1$ is defined by

$$\mathcal{R}_1 = conv \left\{ \varepsilon \sqrt{\frac{1}{\ell(\mu(x,\theta))} + \frac{1}{2} \left(\frac{\ell'(\mu(x,\theta))}{\ell(\mu(x,\theta))} \right)^2} \left(\frac{\partial \mu}{\partial \theta}(x,\theta) \right)^T \Big| x \in \mathcal{X}, \quad \varepsilon \in \{-1, 1\} \right\},$$

and corresponds to a classical Elfving set of the form (1.3). Similarly, it follows observing the inequality

$$\left| \frac{\varepsilon_1}{\sqrt{\ell(\mu,(x,\theta))}} + \frac{\varepsilon_2 \ell'(\mu(x,\theta))}{\sqrt{2\ell^2(\mu(x,\theta))}} \right| \leq \sqrt{ \frac{1}{\ell(\mu(x,\theta))} + \frac{1}{2} \left(\frac{\ell'(\mu(x,\theta))}{\ell(\mu(x,\theta))} \right)^2 }$$

whenever $\varepsilon_1^2 + \varepsilon_2^2 = 1$, that $\mathcal{R} \subset \mathcal{R}_1$, which implies $\mathcal{R} = \mathcal{R}_1$. Thus for nonlinear heteroscedastic regression models, where the conditional expectation and variance are related by a known link function by (3.14), the Elfving set $\mathcal{R}$ reduces to an Elfving set of the form (1.3) and the classical result of Elfving (1952) can be used to characterize locally c-optimal designs.

4   Examples

In this section we will illustrate the geometric characterization by two examples. In the first example we discuss a nonlinear heteroscedastic regression model. Our second example considers a population model and indicates that an information matrix of the form (2.9) may also appear in other situations.

4.1   A heteroscedastic Michaelis-Menten model

As a first example we consider a common model in enzyme kinetics with an heteroscedastic error structure, that is

(4.1)
$$Y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \sqrt{e^{-\theta_3 x_i}} \varepsilon_i \; ; \;\; i = 1, \ldots, n,$$

where errors are independent identically and normal distributed with mean 0 and variance $\sigma^2 > 0$. The parameters to be estimated are $\theta = (\theta_1, \theta_2, \theta_3)$. Clearly, $Var(Y_i) = \tilde{\sigma}^2(x_i, \theta) := e^{-\theta_3 x_i} \sigma^2$, which means that the model is heteroscedastic. Some design issues for the Michaelis-Menten model have been discussed by López-Fidalgo and Wong (2002). In the following we illustrate the geometric characterization of Section 3 by constructing optimal designs for estimating the minimum effective dose in the model (4.1).

For this purpose we assume for the parameters $\theta = (3, 1.7, 0.1), \;\; \sigma^2 = 1$. A straightforward calculation shows that

$$\frac{\partial f(x,\theta)}{\partial \theta} = \left(\frac{x}{\theta_2 + x}, \frac{-\theta_1 x}{(\theta_2 + x)^2}, 0 \right)^T, \quad \frac{\partial \tilde{\sigma}^2(x,\theta)}{\partial \theta} = (0, 0, -xe^{-\theta_3 x})^T,$$

which yields

$$(4.2) \qquad f_1(x, \theta) \;=\; \frac{1}{\sigma e^{-\theta_3 x/2}} \left(\frac{x}{\theta_2 + x}, \frac{-\theta_1 x}{(\theta_2 + x)^2}, 0 \right)^T$$

$$(4.3) \qquad f_2(x, \theta) \;=\; \frac{1}{\sqrt{2\sigma^2 e^{-\theta_3 x}}} (0, 0, -x e^{-\theta_3 x})^T$$

The corresponding generalized Elfving space is as depicted in Figure 2. Suppose we are interested in estimating the minimum effective dose x_{min}, i.e. the smallest value of x resulting in an expected value $E[Y|x] = E$. The solution of this equation is given by

$$x_{min} = \frac{E\theta_2}{\theta_1 - E}.$$

An optimal design for estimating the minimum effective dose minimizes the variance of the estimate for x_{min}. Consequently, if maximum likelihood is used to estimate θ and x_{min} is estimated by $\hat{x}_{min} = E\hat{\theta}_2/(\hat{\theta}_1 - E)$ an optimal design for estimating the minimum effective dose is a locally c-optimal design problem for the vector

$$c = \left(\frac{-E\theta_2}{(\theta_1 - E)^2}, \frac{E}{\theta_1 - E}, 0 \right)^T ,$$

which is marked as the blue line in Figure 1. Let $E = 1$ and the maximum possible observation $x_{max} = 10$, that is $\mathcal{X} = [0, 10]$. From the figure we obtain as optimal design for estimating the minimum effective dose the two point design

$$\xi_c = \left(\begin{array}{cc} 1.4 & 10 \\ 0.952 & 0.048 \end{array} \right).$$

The optimality of this design can be verified by Theorem 3.1.

4.2 A random effect nonlinear regression model

An information matrix of the form (2.9) appears also in the case of a nonlinear model with mixed effects and homoscedastic error structure, which has been intensively discussed in the toxicokinetics and pharmacokinetics literature [see e.g. Beatty and Pigeorsch (1997) or Retout and Mentré (2003) among others]. To be precise we consider a special case of a nonlinear mixed effects model which appears in population toxicokinetics, i.e.

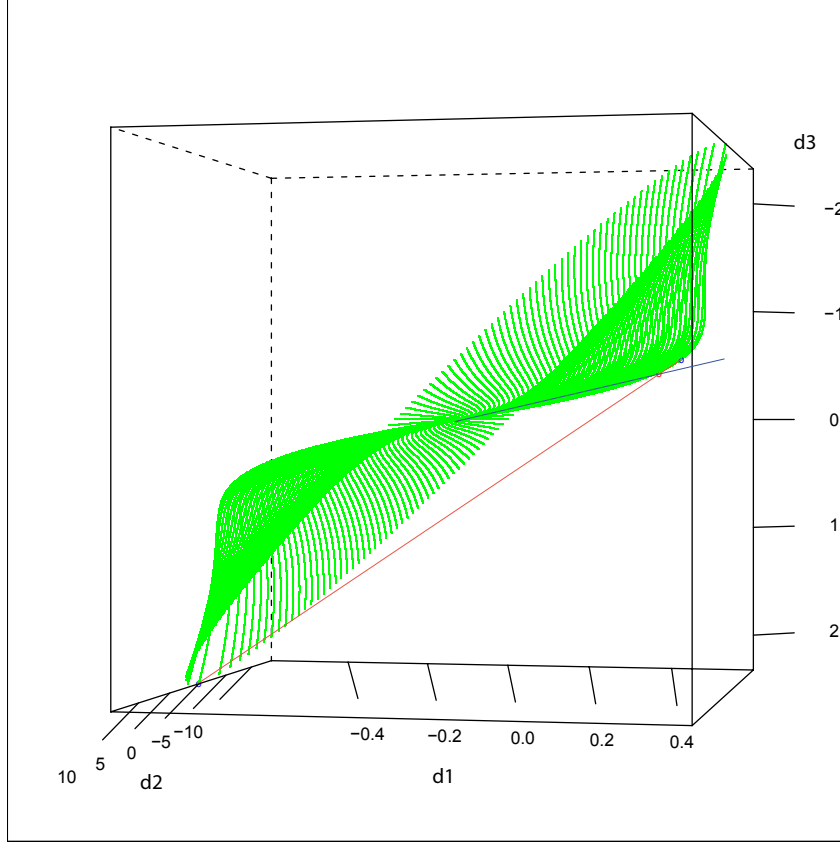$$Y_i = f(x_i, b_i) + \varepsilon_i \; ; \; i = 1, \dots, n \,,$$

Figure 1: *The Elfving space $\mathcal{R}$ defined in (3.3) for a Michaelis-Menten model (4.1) with heteroscedastic error structure. The functions f_1 and f_2 are given by (4.2) and (4.3), respectively.*

where the errors are independent identically and normally distributed with mean 0 and variance $\sigma^2 > 0$. The quantities $b_1, \ldots b_n \sim \mathcal{N}(\theta, \Omega)$ denote here p-dimensional independent normally distributed random variables representing the effect of the corresponding subject under investigation [see Beatty and Pigeorsch (1997), Ette et al. (1995), Cayen and Black (1993)]. We assume that the random variables $b_1, \ldots, b_n$ and the vector $(\varepsilon_1, \ldots, \varepsilon_n)^T$ are independent. The variance of the random variable Y_i can be approximated by

$$\mathrm{Var}(Y_i) \approx \tilde{\sigma}^2(x_i, \theta) := \frac{\partial f(x_i, \theta)}{\partial \theta}^T \Omega \frac{\partial f(x_i, \theta)}{\partial \theta} + \sigma^2.$$

It now follows by similar arguments as in Retout and Mentré (2003) that the Fisher information matrix for the parameter θ at the point x can be approximated by

$$I(x, \theta) = \frac{1}{\tilde{\sigma}^2(x, \theta)} \frac{\partial f(x, \theta)^T}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta} + \frac{1}{2\tilde{\sigma}^4(x, \theta)} \frac{\partial \tilde{\sigma}^2(x, \theta)^T}{d\theta} \frac{\partial \tilde{\sigma}^2(x, \theta)}{\partial \theta},$$

13

which corresponds to the case (2.9) with $k = 2$ and

$$f_1(x, \theta) = \frac{1}{\tilde{\sigma}(x, \theta)} \frac{\partial f(x, \theta)}{\partial \theta}, \qquad f_2(x, \theta) = \frac{1}{\sqrt{2}\tilde{\sigma}^2(x, \theta)} \frac{\partial \tilde{\sigma}^2(x, \theta)}{\partial \theta} \ .$$

Consider for example the simple first order elimination model

(4.4) $$Y = \theta_1 e^{-\theta_2 x} + \varepsilon, \qquad x \in \mathcal{X} = [0, \infty) \,,$$

which is widely used in pharmacokinetics [e.g. Rowland (1993)]. For the parameters we assume

$$\theta = (30, 1.7) \,, \quad \Omega = diag(\omega_1, \omega_2) = diag(1, 0.1) \ \text{and} \ \sigma^2 = 0.04$$

A straightforward calculation shows that

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

$$\frac{\partial^2 f(x, \theta)}{\partial \theta_1 d\theta_2} = -x e^{-\theta_2 x}, \ \frac{\partial^2 f(x, \theta)}{\partial \theta_1^2} = 0, \ \frac{\partial^2 f(x, \theta)}{\partial \theta_2^2} = x^2 e^{-\theta_2 x},$$

which yields

(4.5) $$f_1(x, \theta) \;=\; \frac{1}{\tilde{\sigma}(x, \theta)} (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})$$

(4.6) $$f_2(x, \theta) \;=\; \frac{\sqrt{2}}{\tilde{\sigma}^2(x, \theta)} \sum_{m=1}^{2} \left(\frac{\partial^2 f(x, \theta)}{\partial \theta_m \partial \theta} \frac{\partial f(x, \theta)}{\partial \theta_m} \right) \omega_m \ .$$

The corresponding generalized Elfving space is as depicted in Figure 2. If we are interested in the optimal design for estimating the area under the curve, i.e.

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

we obtain a locally c-optimal design problem for the vector $c = (1/\theta_2, -\theta_1/\theta_2^2)^T$, which is marked as the blue line in Figure 2. From this Figure we obtain as locally c-optimal design for the estimation of the area under the curve the two point design

$$\xi_c = \begin{pmatrix} 0.13 & 1.99 \\ 0.29 & 0.71 \end{pmatrix}.$$

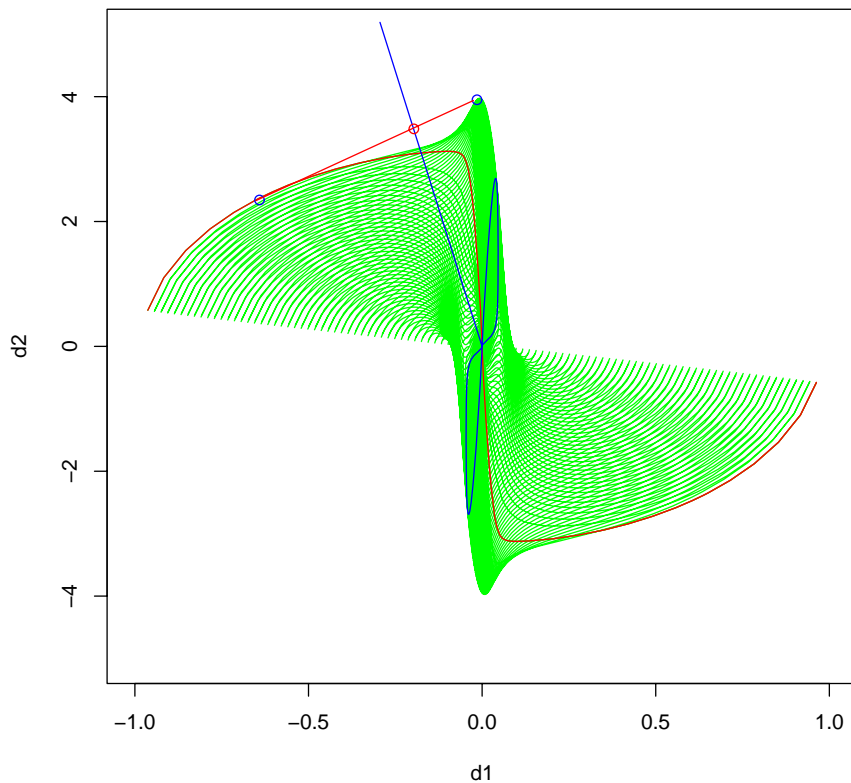The optimality of this design can also be verified by Theorem 3.1.

Figure 2: *The Elfving space $\mathcal{R}$ defined in (3.3) for a simple population model (4.4). The functions f_1 and f_2 are given by (4.5) and (4.6), respectively.*

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.

Atkinson, A. C. and Cook, R. D. (1995). *D*-optimum designs for heteroscedastic linear models. *J. Amer. Statist. Assoc.*, 90:204–212.

Atkinson, A. C. and Donev, A. (1992). *Optimum Experimental Designs.* Clarendon Press, Oxford.

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

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.

Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika*, 46:77–90.

Brown, L. D. and Wong, W. K. (2000). An algorithmic construction of optimal minimax designs for heteroscedastic linear models. *J. Statist. Plann. Inference*, 85:103–114.

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. *Statist. Sci.*, 10(3):273–304.

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

Chernoff, H. and Haitovsky, Y. (1990). Locally optimal design for comparing two probabilities from binomial data subject to misclassification. *Biometrika*, 77:797–805.

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., Bretz, F., Pepelyshev, A., and Pinheiro, J. C. (2008). Optimal designs for dose finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.

Dette, H., Melas, V. B., and Pepelyshev, A. (2004). Optimal designs for a class of nonlinear regression models. *Ann. Statist.*, 32:2142–2167.

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.

Fan, S. K. and Chaloner, K. (2003). A geometric method for singular c-optimal designs. *J. Statist. Plann. Inference*, 113:249–257.

Fang, Z. and Wiens, D. P. (2000). Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. *Journal of the American Statistical Association*, 95:807–818.

Fellmann, J. (1999). Gustav Elfving's contribution to the emergence of the optimal experimental design theory. *Statistical Science*, 14(2):197–200.

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. (1993). Optimal design for nonlinear regression models. *Communications in Statistics - Theory and Methods*, 22:1613–1627.

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

Han, C. and Chaloner, K. (2003). D-and c-optimal designs for exponential regression models used in pharmacokinetics and viral dynamics. *Journal of Statistical Planning and Inference*, 115:585–601.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40:633–643.

King, J. and Wong, W. K. (1998). Optimal minimax designs for prediction in heteroscedastic models. *J. Statist. Plann. Inference*, 69:371–383.

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

Montepiedra, G. and Wong, W. K. (2001). A new design criterion when heteroscedasticity is ignored. *Ann. Inst. Statist. Math.*, 53:418–426.

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

Ortiz, I. and Rodríguez, C. (1998). Optimal designs with respect to elfving's partial minimax criterion for heteroscedastic polynomial regression. *Test*, 7:347–360.

Pázman, A. and Pronzato, L. (2008). *Asymptotic normality of nonlinear least squares under singular experimental designs.* Optimal Design and Related Areas in Optimization and Statistics, Springer Optimization and Its Application, to appear. Springer-Verlag, Heidelberg.

Pukelsheim, F. (1981). On c-optimal design measures. *Mathematische Operationsforschung und Statistik, Series Statistics*, 12:13–20.

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.

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

Ratkowsky, D. A. (1983). *Nonlinear regression modeling: a unified practical approach.* Marcel Dekker, New York.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models.* Dekker, New York.

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.

Searle, S. R. (1982). *Matrix Algebra useful for Statistics. Wiley Series in Probability and Mathematical Statistics.* Wiley and Sons, New York.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression.* John Wiley and Sons Inc., New York, NY.

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

Studden, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.*, 39:1435–1447.

Studden, W. J. (2005). Elfving's theorem revisited. *J. Statist. Plann. Inference*, 130:85–94.

Wong, W. K. and Cook, R. D. (1993). Heteroscedastic G-optimal designs. *Journal of Royal Statistical Society*, 55:871–880.