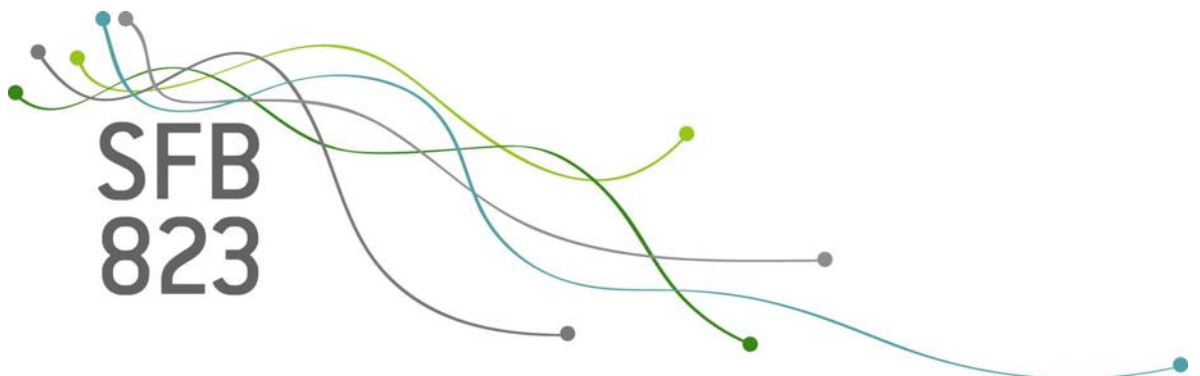SFB 823

Optimal designs for regression models with a constant coefficient of variation

Discussion Paper

Holger Dette, Werner G. Müller

Nr. 20/2012

SFB 823

Optimal designs for regression models with a constant coefficient of variation

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Werner G. Müller

Johannes Kepler Universität Linz

Institut für Angewandte Statistik

4040 Linz, Austria

email: werner.mueller@jku.at

May 29, 2012

Abstract

In this paper we consider the problem of constructing optimal designs for models with a constant coefficient of variation. We explore the special structure of the information matrix in these models and derive a characterization of optimal designs in the sense of Kiefer and Wolfowitz (1960). Besides locally optimal designs, Bayesian and standardized minimax optimal designs are also considered. Particular attention is spent on the problem of constructing D-optimal designs. The results are illustrated in several examples where optimal designs are calculated analytically and numerically.

Keyword and Phrases: optimal design; heteroscedasticity; constant coefficient of variation; polynomial regression

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)]. In many applications the data is heteroscedastic, which means that the variability of the response varies with the predictor. A substantial

1

simplification arises under the additional assumption of a constant coefficient of variation, which is a natural and popular one in engineering, medicine and various other fields of applications [see e.g. Chen et al. (1997) for a very prominent example in microarray analysis]. The purpose of the present paper is the construction of optimal designs for regression models with additive errors and a constant coefficient of variation.

In the general nonlinear regression model 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. A systematic approach to optimal design problems for heteroscedastic regression models was first 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. These authors mainly discuss the D-optimality criterion which was also investigated by Dette and Wong (1996) for a polynomial regression model with exponentially increasing variance functions. More recent work on this topic can be found in Fang and Wiens (2000), Atkinson (2008) or Dette and Holland-Letz (2009). However - to the best knowledge of the authors - optimal designs for nonlinear regression models with an additive error structure and a constant coefficient of variation have not been studied in the literature.

In Section 2 we introduce the necessary notation, derive the information matrix in regression models with a constant coefficient of variation and give a brief introduction into the theory of optimal approximate design. Particular attention is paid to the case of normally distributed responses, but the methodology can easily be transferred to other distributions.We discuss locally optimal designs in the sense of Chernoff (1953) and robust design strategies as introduced in Chaloner and Larntz (1989) and Dette (1997). Section 3 is devoted to the characterization of optimal designs by equivalence theorems in the sense of Kiefer and Wolfowitz (1960). Finally, in Sections 4 and 5 the general results are illustrated in several examples. In particular we present cases where the optimal designs for nonlinear regression models with a constant coefficient of variation can be calculated explicitly. We also discuss a few widely used nonlinear regression models for which efficient and optimal designs can be determined numerically.

2   Fisher information and optimal designs

For regression modeling under the assumption of constant coefficient of variation $\mu(x,\theta)/\sigma(x,\theta)$ we have in principle two possibilities. On the one hand we may assume that the errors in the model are multiplicative (and the observations being positive). In this case we can assume their distribution to be Gamma or lognormal [see Firth (1988) for a discussion] and we may end up with a generalized linear regression model for which numerous general results in optimal design are available [cf. Ford et al. (1992)]. Particularly Burridge and Sebastiani (1994) have concentrated on the constant coefficient of variation case and provide comparisons of D-optimal with some standard factorial designs.

On the other hand a model with additive errors allows more freedom in specification (e.g. also for negative observations), but has to our knowledge not been treated in the literature so far. Furthermore we investigate designs efficient for full maximum likelihood estimators in contrast to the quasi likelihood approach which is typically employed in generalized linear models.

2.1   Heteroscedastic nonlinear regression models

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

$$(2.1) \qquad \mathrm{E}[Y|x] = \mu(x,\theta) \ , \ \mathrm{Var}[Y|x] = \sigma^2(x,\theta)$$

and $\theta \in \mathbb{R}^k$ 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 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 and

$$(2.2) \qquad \Sigma_\theta = \mathrm{diag}(\sigma^2(x_1,\theta), \ldots, \sigma^2(x_n,\theta))$$

as the covariance matrix of the random vector $Y = (Y_1, \ldots, Y_n)^T$. Throughout this paper we assume that the responses at different experimental are independent and normally distributed (other distributions could be investigated by similar methods). Using this additional assumption the Fisher information of Y is given by the $k \times k$ matrix

$$(2.3) \qquad I = \frac{d\mu_\theta^T}{d\theta} \Sigma_\theta^{-1} \frac{d\mu_\theta}{d\theta} + \frac{1}{2} \frac{d\Sigma_\theta}{d\theta}^T \Sigma_\theta^{-2} \frac{d\Sigma_\theta}{d\theta}.$$

3

Following Kiefer (1974) 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); Pukelsheim (2006)]. If n observations can be taken a rounding procedure is applied to obtain integers n_j from the not necessarily integer valued quantities $w_j N$, and n_j observations are taken at experimental condition x_j $(j = 1, \ldots s)$ [see Pukelsheim and Rieder (1992)]. The analogue of the Fisher information matrix for an approximate design is given by the matrix

$$(2.4) \qquad M(\xi, \theta) = \int_{\mathcal{X}} I(x, \theta) d\xi(x)$$

where

$$(2.5) \quad I(x, \theta) = \frac{1}{\sigma^2(x, \theta)} \left(\frac{\partial \mu(x, \theta)}{\partial \theta} \right)^T \frac{\partial \mu(x, \theta)}{\partial \theta} + \frac{1}{2\sigma^4(x, \theta)} \left(\frac{\partial \sigma^2(x, \theta)}{\partial \theta} \right)^T \frac{\partial \sigma^2(x, \theta)}{\partial \theta}$$

denotes the Fisher information at the point x [see also Atkinson and Cook (1995)]. Consequently, an optimal (approximate) design minimizes an appropriate (convex) functional, say Φ, of the asymptotic covariance matrix $M^{-1}(\xi, \theta)$, and there are numerous criteria which can be used for discriminating between competing designs [see Silvey (1980) among many others].

In this paper we will mainly concentrate on optimal designs for estimating linear combinations of the parameter θ. To be precise, let $p > -1$ and let K denote a $k \times s$ matrix of rank $s \leq k$, then a design ξ^* is called locally Φ_p-optimal for estimating $K^T \theta$ if ξ^* minimizes

$$(2.6) \qquad \Phi_p(\xi, \theta) = (\operatorname{tr}((K^T M^{-1}(\xi, \theta) K)^p))^{1/p}$$

among all designs satisfying $\operatorname{range}(K) \subset \operatorname{range}(M(\xi, \theta))$ [see Kiefer (1974) or Pukelsheim (2006)]. Optimal designs with respect to the criteria (2.6) have been studied by numerous authors in the case of heteroscedastic data for which the Fisher information in (2.5) reduces to

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

2.2 Constant coefficient of variation

We now specialize the nonlinear regression model (2.1) to our case of interest. To be precise, let $\theta = (\alpha^T, \tau)$, where $\alpha \in \mathbb{R}^d$, $\tau \in \mathbb{R}^+$ and $d + 1 = k$. We assume that the regression function μ only depends on the parameter α and the standard deviation is proportional to the mean, that is

$$(2.7) \qquad \mu(x, \theta) = \mu(x, \alpha); \qquad \sigma(x, \theta) = \tau \mu(x, \alpha).$$

4

A straightforward calculation yields

$$(2.8) \qquad \frac{\partial \mu(x,\theta)}{\partial \theta} = \left(\frac{\partial \mu(x,\alpha)}{\partial \alpha}, 0 \right) \in \mathbb{R}^{d+1},$$

$$\frac{\partial \sigma^2(x,\theta)}{\partial \theta} = \left(2\tau^2 \frac{\partial \mu(x,\alpha)}{\partial \alpha} \mu(x,\alpha), 2\tau \mu^2(x,\alpha) \right) \in \mathbb{R}^{d+1},$$

and consequently we obtain for the Fisher information in (2.5)

$$(2.9) \qquad \frac{1}{\tau^2} \begin{pmatrix} \frac{(1+2\tau^2)\left(\frac{\partial \mu(x,\alpha)}{\partial \alpha}\right)^T \frac{\partial \mu(x,\alpha)}{\partial \alpha}}{\mu^2(x,\alpha)} & 2\tau \frac{\frac{\partial \mu(x,\alpha)}{\partial \alpha}}{\mu(x,\alpha)} \\ 2\tau \frac{\left(\frac{\partial \mu(x,\alpha)}{\partial \alpha}\right)^T}{\mu(x,\alpha)} & 2 \end{pmatrix}.$$

Note that these expressions only depend on the logarithmic derivative $\frac{\partial}{\partial \alpha} \log \mu(x,\alpha)$. Consequently, a particular simplification occurs for exponential models of the form

$$(2.10) \qquad \mu(x,\alpha) = e^{\eta(x,\alpha)},$$

for which the Fisher information in (2.9) reduces to

$$(2.11) \qquad I(x,\theta) = \frac{1}{\tau^2} \begin{pmatrix} (1+2\tau^2)\left(\frac{\partial \eta(x,\alpha)}{\partial \alpha}\right)^T \frac{\partial \eta(x,\alpha)}{\partial \alpha} & 2\tau \frac{\partial \eta(x,\alpha)}{\partial \alpha} \\ 2\tau \left(\frac{\partial \eta(x,\alpha)}{\partial \alpha}\right)^T & 2 \end{pmatrix}.$$

Note that the resulting optimal designs with respect to the criteria (2.6) are locally optimal in the sense of Chernoff (1953), because they require a specification of the unknown parameters. There are many situations where such preliminary knowledge is available, such that the application of locally optimal designs is well justified [a typical example are phase II dose finding trials, see Dette et al. (2008)]. However, the most important application of locally optimal designs is their use as benchmarks for commonly proposed designs. Moreover, they are the basis for more sophisticated design strategies, which require less precise knowledge about the model parameters, such as sequential, Bayesian or standardized minimax optimality criteria [see Chaloner and Verdinelli (1995) and Dette (1997) among others]. Optimal designs with respect to the latter criteria are called robust designs.

To be precise, let ξ_θ^* denote the locally Φ_p-optimal design and assume that the experimenter can specify a set Θ of possible parameters of the model. A design ξ_M^* is called standardized minimax Φ_p-optimal if it minimizes

$$(2.12) \qquad \Phi_{MM}(\xi) = \max\left\{ \frac{\Phi_p(\xi,\theta)}{\Phi_p(\xi_\theta^*,\theta)} \ \Big| \ \theta \in \Theta \right\}$$

among all designs ξ satisfying

$$(2.13) \qquad \text{range}(K) \subset \text{range}(M(\xi,\theta)) \quad \text{for all} \quad \theta \in \Theta.$$

Similarly, a design ξ_B^* is called Bayesian Φ_p-optimal if and only if it minimizes

(2.14)
$$\int_\Theta \log \Phi_p(\xi, \theta) \pi(d\theta)$$

among all designs satisfying (2.13), where π denotes a prior distribution for the parameter θ.

3 Optimality conditions

In this section we will derive checking conditions for the optimality of a given design. For this purpose observe that the Fisher Information in (2.5) can be represented in the form

(3.1)
$$I(x, \theta) = \sum_{j=1}^{2} f_j(x, \theta) f_j^T(x, \theta),$$

where the vectors f_1 and f_2 are given by

(3.2)
$$f_1(x, \theta) = \frac{1}{\sigma(t, \theta)} \left(\frac{\partial \mu(t, \theta)}{\partial \theta} \right)^T \; ; \quad f_2(x, \theta) = \frac{1}{\sqrt{2}\sigma^2(t, \theta)} \left(\frac{\partial \sigma^2(t, \theta)}{\partial \theta} \right)^T ,$$

respectively. The following result gives an equivalence theorem, which can be used to check the optimality of a candidate design.

Theorem 3.1 *The design ξ^* minimizes the criterion (2.6) if and only if there exist a generalized inverse of the matrix $M(\xi, \theta)$, such that the inequality*

$$d(x, \theta, \xi) = \left\{ \frac{1}{\sigma^2(x, \theta)} \frac{\partial \mu(x, \theta)}{\partial \theta} GK(K^T M^-(\xi^*, \theta)K)^{p-1} K^T G \left(\frac{\partial \mu(x, \theta)}{\partial \theta} \right)^T + \frac{1}{2\sigma^4(x, \theta)} \frac{\partial \sigma^2(x, \theta)}{\partial \theta} \right.$$

(3.3)
$$\left. \times GK(K^T M^-(\xi^*, \theta)K)^{p-1} K^T G \left(\frac{\partial \sigma^2(x, \theta)}{\partial \theta} \right)^T \right\} \left(\mathrm{tr}(K^T M^-(\xi^*, \theta)K)^p \right)^{-1} \leq 1$$

is satisfied for all $x \in \mathcal{X}$. Moreover, if ξ^* is locally Φ_p-optimal there is equality in (3.3) for all support points of the optimal design.

Proof. For fixed θ consider the set

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

of all information matrices of the form (2.4), where Ξ denote the set of all approximative designs on $\mathcal{X}$ and $I(x, \theta)$ is given by (3.1) and (3.2). $\mathcal{M}$ is obviously convex and it follows from Theorem

6

7.19 in Pukelsheim (2006) that the design ξ^* is locally Φ_p–optimal if and only if there exists a generalized inverse, say G, of the matrix $M(\xi^*, \theta)$ such that the inequality

$$\mathrm{tr}(AGK(K^T M^-(\xi^*, \theta)K)^{p-1}K^T G) \leq \mathrm{tr}(K^T M^-(\xi^*, \theta)K)^p$$

holds for all $A \in \mathcal{M}$, where there is equality for any matrix $A \in \mathcal{M}$ which minimizes $((K^T M^-(\xi^*, \theta)K)^p)^{1/p}$ in the set $\mathcal{M}$. Note that the family $\mathcal{M}$ is the convex hull of the set

$$\left\{ f_1(x, \theta)f_1^T(x, \theta) + f_2(x, \theta)f_2^T(x, \theta) \mid x \in \mathcal{X} \right\},$$

where f_1 and f_2 are defined in (3.2). Therefore the assertion of Theorem 3.1 follows by a standard argument of optimal design theory [see e.g. Silvey (1980)]. $\qquad\square$

The following theorem gives a corresponding result for Bayesian and standardized minimax optimality criterion and can be obtained by a combination of arguments presented in the proof of Theorem 3.1 with the results in Dette et al. (2007). The details are omitted for the sake of brevity.

Theorem 3.2

(a) *A design ξ_B^* is Bayesian Φ_p-optimal if and only if the inequality*

$$(3.4) \qquad\qquad u(x) = \int_\Theta d(x, \xi_B^*, \theta)\pi(d\theta) \leq 1$$

holds for all $x \in \mathcal{X}$, where the function d is defined in (3.3). Moreover, in this case there is equality in (3.4) for all support points of the design ξ_B^*.

(b) *The design ξ_M^* is standardized minimax Φ_p-optimal if and only if there exists a prior π^* on the set*

$$\left\{ \theta \in \Theta \ \Big| \ \frac{\Phi_p(\xi_M^*, \theta)}{\Phi_p(\xi_\theta^*, \theta)} = \Phi_{MM}(\xi_M^*) \right\},$$

such that the inequality

$$(3.5) \qquad\qquad \int_\Theta d(x, \xi_M^*, \theta)\pi^*(d\theta) \leq 1$$

holds for all $\theta \in \Theta$. Moreover, in this case there is equality in (3.5) for all support points of the design ξ_M^*.

Note that in the model with a constant coefficient of variation the function d in condition (3.3) of Theorem 3.1 reduces to

$$d(x,\theta,\xi) = \frac{1}{\tau^2\mu^2(x,\alpha)}\left\{\left(\frac{\partial\mu(x,\theta)}{\partial\alpha},0\right)GK(K^TM^-(\xi^*,\theta)K)^{p-1}K^TG\left(\frac{\partial\mu(x,\theta)}{\partial\alpha},0\right)^T\right.$$

$$(3.6) \qquad\qquad +2\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},\mu(x,\alpha)\right)GK(K^TM^-(\xi^*,\theta))^{p-1}K^TG\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},\mu(x,\alpha)\right)^T\left.\right\}.$$

In particular, the choice $K = I_{d+1}$, $p = 0$ corresponds to the estimation of all parameters and the D-optimality criterion and we obtain

$$(3.7) \qquad d(x,\theta,\xi) = \frac{1}{\tau^2\mu^2(x,\alpha)}\left\{\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},0\right)M^{-1}(\xi,\theta)\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},0\right)^T\right.$$

$$+2\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},\mu(x,\alpha)\right)M^{-1}(\xi,\theta)\left(\tau\frac{\partial\mu(x,\alpha)}{\partial\alpha},\mu(x,\alpha)\right)^T\left.\right\}.$$

4 Locally D-optimal designs

In this section, we investigate D-optimal designs for nonlinear regression models with a constant coefficient of variation. We consider first a common exponential model and then concentrate on two models, which are widely used in applications, namely the Michaelis-Menten model

$$(4.1) \qquad\qquad \mu_1(x,\theta) = \frac{\alpha_1 x}{\alpha_2 + x}$$

and the EMAX model

$$(4.2) \qquad\qquad \mu_2(x,\theta) = \alpha_0 + \frac{\alpha_1 x}{\alpha_2 + x},$$

where $x \in \mathcal{X} = [x_\ell, x_u] \subset \mathbb{R}_0^+$. Some applications of the Michaelis-Menten model can be found in Johansen (1984), Cornish-Browden (1995) or López et al. (2000) among many others. Usually the function μ_1 represents the velocity of a chemical reaction, the parameter $\alpha_1 \geq 0$ denotes the maximum velocity, the predictor x reflects the concentration of a substrate and the parameter $\alpha_2 \geq 0$ is the half-saturated constant, the concentration, where the velocity is half-maximal. Various applications of the EMAX model (4.2) in pharmaceutical or toxicological studies have been discussed by Danesi et al. (2002); Chien et al. (2005); Blake et al. (2008) among others. In model (4.2) the parameter $\alpha_0 \geq 0$ usually represents the response at placebo, $\alpha_1 \geq 0(\leq 0)$ is the maximum achievable increase (decrease) above the placebo response and $\alpha_2 \geq 0$ is the dose which produces 50% of the effect.

8

4.1 Exponential models

But consider first the common exponential regression model

$$(4.3) \qquad \mu(x, \alpha) = e^{\sum_{j=0}^{d-1} \alpha_j x^j}$$

where $x \in [a, b] \subset \mathbb{R}$. Optimal designs for special cases of the model (4.3) (more precisely the case $d = 2$) have been discussed in the literature under the assumption of a homoscedastic error [see Mukhopadhyaya and Haines (1995), Dette and Neugebauer (1997), Han and Chaloner (2003), Dette et al. (2006)]. In this section we derive analytic expressions for optimal designs in these models with a constant coefficient of variation. The following result shows that in this case the D-optimal designs can be determined explicitly for any degree of the polynomial in (4.3).

Theorem 4.1 *The D-optimal design for the exponential regression model* (4.3) *on the interval* $[a, b]$ *with a constant coefficient of variation puts equal masses at the points*

$$x_j = \frac{(b-a)s_j + b + a}{2}; \qquad j = 1, \ldots, d$$

where $s_1, \ldots, s_d$ *are the roots of the polynomial* $(x^2 - 1)L'_{d-1}(x)$ *and* $L_{d-1}(x)$ *is the Legendre polynomial of degree* $d-1$ *orthogonal with respect to the Lebesgue measure on the interval* $[-1, 1]$.

Proof. Observing the representation (2.10) we have $\eta(x, \alpha) = \sum_{j=0}^{d-1} \alpha_j x^j$ and it follows from (2.11) that the information matrix of a given design ξ can be represented as

$$(4.4) \qquad M(\xi, \theta) = \frac{1}{\tau^2} \begin{pmatrix} (1 + 2\tau^2)\underline{M}(\xi) & 2\tau\underline{m}(\xi) \\ 2\tau\underline{m}^T(\xi) & 2 \end{pmatrix},$$

where the matrix $\underline{M}(\xi)$ and the vector $\underline{m}(\xi)$ are given by

$$(4.5) \qquad \underline{M}(\xi) = (c_{i+j})_{i,j=0}^{d-1}; \qquad \underline{m}(\xi) = (c_i)_{i=0}^{d-1},$$

respectively, and

$$c_i = c_i(\xi) = \int_{\alpha}^{b} x d\xi(x); \qquad i = 0, 1, \ldots$$

denote the ith moment of the design ξ. Observing that $\underline{m}^T(\xi)$ coincides with the first row of the matrix $\underline{M}(\xi)$ it follows that $\underline{M}^{-1}(\xi)\underline{m}(\xi) = (1, 0, \ldots, 0)^T \in \mathbb{R}^d$ and we obtain for the determinant of the matrix $M(\xi, \theta)$ the representation

$$
\begin{aligned}
(4.6) \qquad \det(M(\xi, \theta)) &= (1 + 2\tau^2)^d \cdot \det(\underline{M}(\xi)) \cdot \det\left(2 - \frac{4\tau^2}{1 + 2\tau^2}\underline{m}^T(\xi)\underline{M}^{-1}(\xi)\underline{m}(\xi)\right) \\
&= 2(1 + 2\tau^2)^{d-1}\det\underline{M}(\xi).
\end{aligned}
$$

Consequently, the D-optimal design in a constant coefficient of variation model with mean (4.3) can be obtained by maximizing the determinant of the matrix $\underline{M}(\xi)$. Observing the representation of the matrix $\underline{M}(\xi)$ in (4.5) it follows that the optimal design coincides with the D-optimal design in a homoscedastic polynomial regression model of degree $d-1$. This design has been determined by Hoel (1958) and the assertion of Theorem 4.1 follows. $\qquad\square$

4.2 Michaelis Menten model

For the Michaelis Menten model (4.1) with a constant coefficient of variation the gradient of the expected response is given by

$$\frac{\partial}{\partial \alpha}\mu_1(x,\alpha) = \frac{x}{x+\alpha_2}\left(1, -\frac{\alpha_1}{x+\alpha_2}\right)$$

and consequently the Fisher information at the point $x \in [x_\ell, x_u]$ reduces to

(4.7) $I_1(x,\theta) = \dfrac{1}{\tau^2}\left(\begin{array}{cc} (1+2\tau^2)\left(\begin{array}{cc} \frac{1}{\alpha_1^2} & -\frac{1}{\alpha_1(x+\alpha_2)} \\ -\frac{1}{\alpha_1(x+\alpha_2)} & \frac{1}{(x+\alpha_2)^2} \end{array} \right) & 2\tau\left(\begin{array}{c} \frac{1}{\alpha_1} \\ -\frac{1}{x+\alpha_2} \end{array} \right) \\ 2\tau\left(\frac{1}{\alpha_1} - \frac{1}{x+\alpha_2}\right) & 2 \end{array} \right) \in \mathbb{R}^{3\times 3}.$

Note that in a model with a constant coefficient of variation the expectation and variance of the response at $x = 0$ vanishes, and consequently we assume without loss of generality $x_\ell > 0$. The following result specifies the locally D-optimal design

Theorem 4.2 *The locally D-optimal design for the Michaelis Menten model with a constant coefficient of variation has equal masses at the points x_u and x_ℓ.*

Proof. For the D-optimality criterion in a model with a constant coefficient of variation we obtain from the equivalence Theorem 3.1 and (3.7) the condition

(4.8) $\quad\dfrac{1}{\tau^2}\dfrac{1}{\mu_1^2(x,\alpha)}\dfrac{\partial}{\partial\alpha}\mu_1(x,\alpha)M^{11}(\xi,\theta)\left(\dfrac{\partial}{\partial\alpha}\mu_1(x,\alpha)\right)^T$

$$+\frac{2}{\tau^2\mu_1^2(x,\alpha)}\left(\tau\frac{\partial}{\partial\alpha}\mu_1(x,\alpha), \mu_1(x,\alpha)\right)M_1^{-1}(\xi,\theta)\left(\tau\frac{\partial}{\partial\alpha}\mu_1(x,\alpha), \mu_1(x,\alpha)\right)^T \le 3$$

for all $x \in [x_\ell, x_u]$, where $M^{11}(\xi,\theta)$ denotes the upper 2×2 block of the matrix

(4.9) $$M_1(\xi,\theta) = \int_{x\ell}^{x_u} I_1(x,\theta)d\xi(x)$$

10

and the Fisher information $I_1(x, \theta)$ is defined in (4.7). It is easy to see that all terms in (4.8) and (4.9) depend only on the expression

$$\frac{\partial}{\partial \alpha} \mu_1(x, \alpha) / \mu_1(x, \alpha) = \left(\frac{1}{\alpha_1}, -\frac{1}{x + \alpha_2} \right)$$

and therefore we introduce the transformation

$$z = \frac{1}{x + \alpha_2}$$

and the induced design space $\mathcal{Z} = [z_1, z_2] = [\frac{1}{x_u + \alpha_2}, \frac{1}{x_\ell + \alpha_2}]$. The corresponding candidate for the D-optimal design has equal masses at the points z_1 and z_2 and after a straightforward but tedious calculation the checking condition (4.8) reduces to

$$\frac{4z^2 + 3z_1^2 - 2z_1 z_2 + 3z_2^2 - 4z(z_1 + z_2)}{(z_1 - z_2)^2} = \frac{4(z - z_1)(z - z_2) + 3(z_1 - z_2^2)}{(z_1 - z_2)^2} \leq 3$$

Because this inequality is obviously satisfied for all $z \in [z_1, z_2]$ the assertion of Theorem 4.2 follows. $\qquad \square$

4.3   EMAX model

For the EMAX model (4.2) with a constant coefficient of variation we assume that $\alpha_0 + x_\ell > 0$ (if $x_\ell = \alpha_\ell = 0$ the variance at the left boundary point vanishes) and obtain

$$(4.10) \qquad \frac{\partial}{\partial \alpha} \mu_2(x, \alpha) / \mu_2(x, \alpha) = \left(\alpha_0 + \frac{\alpha_1 x}{\alpha_2 + x} \right)^{-1} \left(1, \frac{x}{\alpha_2 + x}, -\frac{\alpha_1 x}{(\alpha_2 + x)^2} \right)^T.$$

In general optimal designs cannot be found explicitly. Nevertheless, the following result establishes that Φ_p-optimal designs for the EMAX model with a constant coefficient of variation are always supported at at most 3 points.

Theorem 4.3 *Any locally Φ_p-optimal design for the EMAX model with a constant coefficient of variation is supported at at most 3 points. Moreover, if a locally Φ_p-optimal design is supported at 3 points, the boundary points x_ℓ and x_u of the design space must be support points.*

Proof. We introduce the transformation

$$(4.11) \qquad\qquad\qquad z = \frac{\alpha_1 x}{\alpha_2 + x}$$

and the induced design space $\mathcal{Z} = [z_\ell, z_u]$, where $z_\ell = \alpha_1 x_\ell/(\alpha_2 + x_\ell)$, $z_u = \alpha_1 x_u/(\alpha_2 + x_u)$. Then the gradient of the logarithm of the response in (4.10) can be represented as

$$(4.12) \qquad (\alpha_0 + z)^{-1}\left(1, \frac{z}{\alpha_1}, \frac{z(z - \alpha_1)}{\alpha_1 \alpha_2}\right)^T = B(\alpha_0 + z)^{-1}(1, z, z^2)^T$$

for some appropriate non-singular 3×3 matrix B. Consequently, it follows from (3.7) that the design ξ^* is Φ_p-optimal if and only if the inequality

$$(\alpha_0 + z)^{-2}\Big\{(1, z, z^2, 0)D_1(\xi, \theta)(1, z, z^2, 0)^T + (1, z, z^2, \alpha_0 + z)D_2(\xi, \theta)(1, z, z^2, \alpha_0 + z)^T\Big\} \leq 3$$

holds for all $z \in \mathcal{Z}$, where $D_1(\xi, \theta)$ and $D_2(\xi, \theta)$ denote appropriate matrices which do not depend on the variable $z \in \mathcal{Z}$ but on the coefficient τ. Now it is easy to see that this inequality is equivalent to the inequality

$$P_4(z, \xi, \theta) \leq 0$$

for all $z \in \mathcal{Z}$, where P_4 is an appropriate polynomial of degree 4. Therefore the assertion follows by a simple counting argument of the roots of the derivative of $P_4(z, \xi, \theta)$ with their corresponding multiplicities. $\qquad\square$

For a general $p > -1$ the Φ_p-optimal design has to be determined numerically. However, for the D-optimality criterion (i.e. $p = 0$) an explicit solution is available, which is given in the following result.

Theorem 4.4 *The locally D-optimal design for the EMAX model on the interval $[x_\ell, x_u]$ with a constant coefficient of variation has equal masses at the 3 points x_ℓ, x_u and*

$$x^* = \frac{-\alpha_0\alpha_2^2 + x_\ell x_u(\alpha_0 + \alpha_1) + \sqrt{(\alpha_2 + x_\ell)(\alpha_2 + x_u)(\alpha_0\alpha_2 + x_\ell(\alpha_1 + \alpha_0))(\alpha_0\alpha_2 + x_u(\alpha_1 + \alpha_0))}}{\alpha_0(2\alpha_2 + x_\ell + x_u) + \alpha_1(\alpha_2 + x_\ell + x_u)}.$$

Proof. By Theorem 4.3 the D-optimal design is supported at 3 points including the boundary points x_ℓ and x_u. Note that the standard argument for calculating the weights of the D-optimal design is not applicable here because we have 3 support points but the information matrix is a 4×4 matrix. For this purpose we use the transformation (4.11) and obtain for the Fisher information (2.9)

$$
\begin{aligned}
I(x, \theta) &= \frac{1}{\tau^2}\left(\frac{\partial \mu_2(x, \alpha)}{\partial \alpha}/\mu_2(x, \alpha), 0\right)\left(\frac{\partial \mu_2(x, \alpha)}{\partial \alpha}/\mu_2(x, \alpha), 0\right)^T \\
&\qquad + 2\left(\frac{\partial \mu_2(x, \alpha)}{\partial \alpha}/\mu_2(x, \alpha), \frac{1}{\tau}\right)\left(\frac{\partial \mu_2(x, \alpha)}{\partial \alpha}/\mu_2(x, \alpha), \frac{1}{\tau}\right)^T \\
&= \frac{1}{\tau^2}\tilde{B}(k(z), 0))(k(z), 0))^T)\tilde{B}^T + 2\tilde{B}(k(z), 1/\tau)(k(z), 1/\tau)^T\tilde{B}^T \\
&= \frac{1}{\tau^2}\tilde{B}\tilde{I}(z, \alpha_0)\tilde{B}^T
\end{aligned}
$$

12

where $k(z) = (1/(\alpha_0 + z), z/(\alpha_0 + z), z^2/(\alpha_0 + z))^T$, the 4×4 block-diagonal matrix $\tilde{B}$ is given by

$$\tilde{B} = \begin{pmatrix} B & 0 \\ 0 & 1 \end{pmatrix},$$

B is defined in (4.12) and the Fisher information on the space $\mathcal{Z} = [z_\ell, z_u]$ is defined by

(4.13)
$$\tilde{I}(z, \alpha_0) = \begin{pmatrix} (1 + 2\tau^2)I_{11}(z, \alpha) & 2\tau I_{12}(z, \alpha_0) \\ 2\tau I_{12}^T(z, \alpha_0) & 2 \end{pmatrix}.$$

Here the matrices $I_{11}(z, \alpha_0)$ and $I_{12}(z, \alpha_0)$ are given by

$$I_{11}(z, \alpha_0) = \frac{1}{(z + \alpha_0)^2} \begin{pmatrix} 1 & z & z^2 \\ z & z^2 & z^3 \\ z^2 & z^3 & z^4 \end{pmatrix} \quad \text{and} \quad I_{12}(z, \alpha_0) = \frac{1}{z + \alpha_0} \begin{pmatrix} 1 \\ z \\ z^2 \end{pmatrix},$$

respectively. We now consider the D-optimal design problem for the Fisher information (4.13) on the design space $\mathcal{Z} = [z_\ell, z_u]$ for which by Theorem 4.3 the optimal solution has 3 support points including the boundary points z_ℓ and z_u. Define

$$h(z, p_1, p_2) = \det(p_1 I(z_\ell, \alpha_0) + p_2 I(z, \alpha_0) + (1 - p_1 - p_2)I(z_u, \alpha_0)),$$

then maximizing h with respect to p_1, p_2 yields by tedious calculations $p_1 = p_2 = 1/3$. The function $h(z, \frac{1}{3}, \frac{1}{3})$ is given by

$$h(z) = 2\frac{(1 + t^2)^2(z_\ell - z)^2(z_\ell - z_u)^2(z - z_u)^2}{27(\alpha_0 + z_u)^2(\alpha_0 + z_\ell)^2(\alpha_0 + z)^2}$$

which is maximal at

$$z^* = -\alpha_0 + \sqrt{(\alpha_0 + z_\ell)(\alpha_0 + z_u)}.$$

Therefore the D-optimal design on the design space $\mathcal{Z} = [z_\ell, z_u]$ has equal masses at the points z_ℓ, z^* and z_u. Observing that $z_\ell = \alpha_1 x_\ell/(\alpha_2 + x_\ell)$, $z_u = \alpha_1 x_u/(\alpha_2 + x_u)$ the statement of the theorem now follows by transforming these results to the original design space via the transformation $x = \frac{\alpha_2 z}{\alpha_1 - z}$. $\qquad \square$

Note that in contrast to the homoscedastic case, when the locally D-optimal design depends only on the parameter α_2 [see Dette et al. (2010)], the D-optimal design for the EMAX model with a constant coefficient of variation depends on all parameters of the model. However, it is remarkable that the locally D-optimal designs do not depend on the size $\tau > 0$ of this coefficient.

5 Bayesian and standardized minimax D-optimal designs

For the Bayesian and standardized minimax optimality criteria the corresponding optimal designs have to be determined numerically in all cases of practical interest and for brevity we will only investigate robust designs for the EMAX model. Our main result shows showed that the Bayesian and standardized minimax D-optimal designs do not depend on the parameter τ.

Theorem 5.1 *The Bayesian and the standardized minimax D-optimal design for the EMAX model with a constant coefficient of variation $\tau > 0$ do not depend on the parameter τ.*

Proof. Recall the definition of the logarithmic derivative in (4.9) and introduce the transformation

$$x = \gamma(w) := \frac{\alpha_0 \alpha_2 w}{\alpha_1 - (\alpha_0 + \alpha_1)w}.$$

A straightforward calculation yields

$$
\begin{aligned}
\left(\frac{\partial}{\partial \alpha}\mu_2(x,\alpha)/\mu_2(x,\alpha)\right)^T &= \frac{1-w}{\alpha_0}\left(1, \frac{\alpha_0 w}{\alpha_1(1-w)}, -\frac{\alpha_0 w(\alpha_1 - (\alpha_o + \alpha_1)w)}{\alpha_1 \alpha_2(1-w)^2}\right)^T \\
&= D(1, w, \delta(w))^T
\end{aligned}
$$

where the function δ and the matrix D are defined by $\delta(w) = \frac{w^2}{1-w}$ and

$$
D = \begin{pmatrix}
\frac{1}{\alpha_0} & -\frac{1}{\alpha_0} & 0 \\
0 & \frac{1}{\alpha_1} & 0 \\
0 & -\frac{1}{\alpha_2} & \frac{\alpha_0}{\alpha_1 \alpha_2}
\end{pmatrix}
$$

respectively. Consequently, the determinant of the information matrix in the Bayesian and standardized minimax D-optimality criterion can represented as

$$|M(\xi,\theta)| = |\bar{D}|^2 \left| \int \bar{I}(w,\theta)d\xi(\gamma(w)) \right|$$

for some appropriate matrix $\bar{D}$, where

$$
\bar{I}(w,\theta) = \begin{pmatrix}
(1+2\tau^2)\bar{I}_{11}(w,\alpha) & 2\tau \bar{I}_{12}(w,\alpha) \\
2\tau \bar{I}_{12}^T(w,\alpha) & 2
\end{pmatrix}
$$

and

$$
\bar{I}_{11}(w,\alpha) = \begin{bmatrix}
1 & w & \delta(w) \\
w & w^2 & w\delta(w) \\
\delta(w) & w\delta(w) & \delta^2(w)
\end{bmatrix}, \quad
\bar{I}_{12}(w,\alpha) = \begin{pmatrix}
1 \\
w \\
\delta(w)
\end{pmatrix}.
$$

14

Therefore the determinant can be calculated by exactly the same arguments as given in the proof of Theorem 4.1 and a term depending on τ factorizes [see formula (4.6)]. Transforming back to original design space yields the assertion of assertion of Theorem 5.1. $\qquad\square$

For numerical illustration we considered a scenario motivated by the discussion in Dette et al. (2008), where the design space is given by $\mathcal{X} = [0, 150]$ and the parameters vary in the interval

(5.1) $\qquad\qquad\alpha_0 \in [0.5, 0.75]\qquad \alpha_1 \in [0.25, 0.75]\qquad \alpha_2 \in [15, 25]$

(5.2) $\qquad\qquad\alpha_0 \in [0.5, 1]\qquad \alpha_1 \in [0.25, 0.75]\qquad \alpha_2 \in [24, 26]$

(5.3) $\qquad\qquad\alpha_0 \in [0.5, 3]\qquad \alpha_1 \in [0.1, 2]\qquad \alpha_2 \in [20, 30]$

(5.4) $\qquad\qquad\alpha_0 \in [0.5, 2]\qquad \alpha_1 \in [0.2, 1]\qquad \alpha_2 \in [10, 40]$

where a uniform prior is used in the Bayesian optimality criterion. Some illustrative results are shown in Table 1.

support	Bayes			Minimax					
	x_1	x_2	x_3	x_1	x_2	x_3	w_1	w_2	w_3
(5.1)	0	15.310	150	0	14.157	150	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
(5.2)	0	12.300	150	0	11.559	150	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
(5.3)	0	15.301	150	0	13.392	150	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
(5.4)	0	15.009	150	0	12.037	150	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Table 1: *Weights and support points of standardized minimax and Bayesian D-optimal designs for the EMAX model (4.2) with a constant coefficient of variation.*

The optimality of the calculated robust D-optimal designs has been verified by the equivalence Theorem 3.2, more precisely in equality (3.4), and a representative plot is shown in Figure 1. We observe from Table 1 that the optimal designs with respect to the Bayesian and standardized minimax approach are very similar.
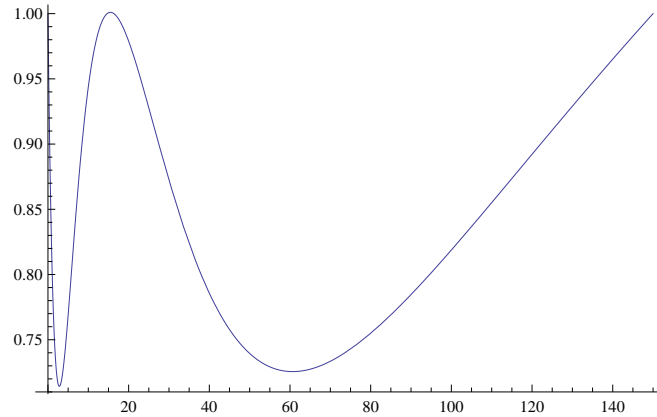
Figure 1: *The function u defined in (3.4) for a uniform prior on the cube (5.1)*

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. *Journal of the American Statistical Association*, 90:204–212.

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.

Blake, K., Madabushi, R., Derendorf, H., and Lima, J. (2008). Population pharmacodynamic model of bronchodilator response to inhaled albuterol in children and adults with asthma. *Chest*, 134(5):981–989.

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

Burridge, J. and Sebastiani, P. (1994). D-Optimal Designs for Generalised Linear Models with Variance Proportional to the Square of the Mean. *Biometrika*, 81(2):295–304.

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

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.

Chen, Y., Dougherty, E. R., and Bittner, M. L. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2(4):364–374.

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

Chien, J. Y., Friedrich, S., Heathman, M. A., de Alwis, D. P., and Sinha, V. (2005). Pharmacokinetics/pharmacodynamics and the stages of drug development: Role of modeling and simulation. *AAPS Journal*, 7(3):E544–E559.

Cornish-Browden, A. (1995). *Fundamentals of Enzyme Kinetics (Rev. Ed.)*. Portland Press, London.

Danesi, R., Innocenti, F., Fogli, S., Gennari, A., Baldini, E., Di Paolo, A., Salvadori, B., Bocci, G., Conte, P. F., and Del Tacca, M. (2002). Pharmacokinetics and pharmacodynamics of combination chemotherapy with paclitaxel and epirubicin in breast cancer patients. *British Journal of Clinical Pharmacology*, 53(5):508–518.

Dette, H. (1997). Designing 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., Haines, L., and Imhof, L. (2007). Maximin and Bayesian optimal designs for regression models. *Statistica Sinica*, 17:463–480.

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

Dette, H., Kiss, C., Bevanda, M., and Bretz, F. (2010). Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, 97(2):513–518.

Dette, H., Martinez Lopez, I., Ortiz Rodriguez, I., and Pepelyshev, A. (2006). Efficient design of experiment for exponential regression models. *Journal of Statistical Planning and Inference*, 136:4397–4418.

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

Dette, H. and Neugebauer, H. M. (1997). Bayesian D-optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, 60:331–349.

Dette, H. and Wong, W. K. (1996). Optimal Bayesian designs for models with partially specified heteroscedastic structure. *Annals of Statistics*, 24:2108–2127.

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.

Firth, D. (1988). Multiplicative Errors: Log-Normal or Gamma? *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):266–268.

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.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. *Annals of Mathematical Statistics*, 29:1134–1145.

Johansen, S. (1984). Functional relations, random coefficients and nonlinear regression, with application to kinetic data. *Lecture Notes in Statistics*, 22.

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

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

López, S., France, J., Gerrits, W. J., Dhanoa, M. S., Humphries, D. J., and Dijkstra, J. (2000). A generalized Michaelis-Menten equation for the analysis of growth. *Journal of Animal Science*, 78(7):1816–1828.

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

Mukhopadhyaya, S. and Haines, L. M. (1995). Bayesian D-optimal designs for the exponential growth model. *Journal of Statistical Planning and Inference*, 44(3):385–397.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

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

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.

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.