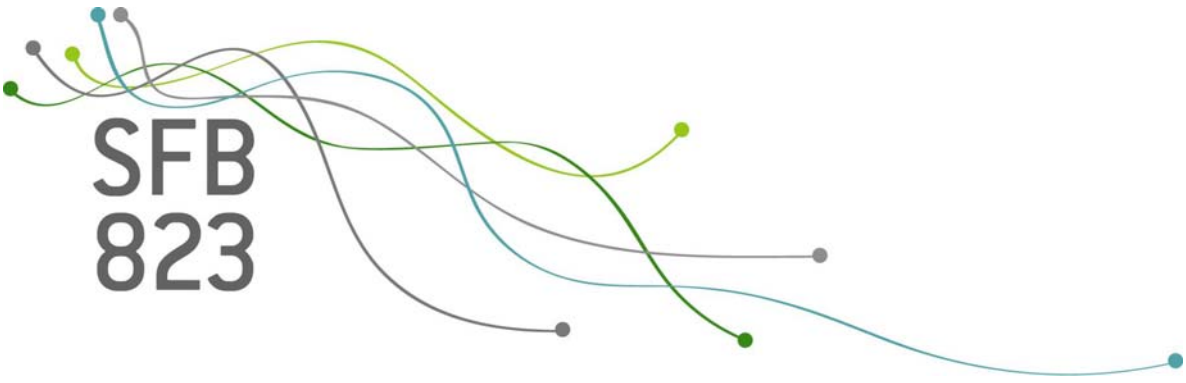
SFB 823

Discussion Paper

Optimal designs for nonlinear regression models with respect to non-informative priors

Ina Burghaus, Holger Dette

Nr. 41/2013



Optimal designs for nonlinear regression models with respect to non-informative priors

Ina Burghaus
Ruhr-Universität Bochum
Abteilung für Medizinische Informatik,
Biometrie und Epidemiologie
44780 Bochum, Germany
e-mail: ina.burghaus@rub.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: holger.dette@rub.de

October 23, 2013

Abstract

In nonlinear regression models the Fisher information depends on the parameters of the model. Consequently, optimal designs maximizing some functional of the information matrix cannot be implemented directly but require some preliminary knowledge about the unknown parameters. Bayesian optimality criteria provide an attractive solution to this problem. These criteria depend sensitively on a reasonable specification of a prior distribution for the model parameters which might not be available in all applications. In this paper we investigate Bayesian optimality criteria with non-informative prior distributions. In particular, we study the Jeffreys and the Berger-Bernardo prior for which the corresponding optimality criteria are not necessarily concave. Several examples are investigated where optimal designs with respect to the new criteria are calculated and compared to Bayesian optimal designs based on a uniform and a functional uniform prior.

Keywords: optimal design; Bayesian optimality criteria; non-informative prior; Jeffreys prior; reference prior; polynomial regression; canonical moments; heteroscedasticity

1 Introduction

Nonlinear regression models provide an important tool to describe the relation between a response and a predictor and have many applications in engineering, physics, biology, economics and medicine, among others [see Ratkowsky (1983)]. It is well known that a good design can improve the accuracy of the statistical analysis substantially and numerous authors have worked

on the problem of constructing optimal designs for nonlinear regression models. An intrinsic difficulty of these optimization problems consists in the fact that the Fisher information, say $I(x, \boldsymbol{\theta})$, at an experimental condition x depends on the unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ of the model. A common approach in the literature is to assume some prior knowledge of the unknown parameter, which can be used for the construction of optimal designs. Chernoff (1953) proposed the concept of local optimality where a fixed value of the unknown parameter is specified, and a design is determined by maximizing a functional of the information matrix for this specified parameter.

Since this pioneering work numerous authors have constructed locally optimal designs for various regression models [see He et al. (1996), Khuri et al. (2006), Fang and Hedayat (2008), Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011), among many others]. On the other hand, the concept of local optimality has been criticized by several authors, because it depends sensitively on a precise specification of the unknown parameters and can lead to inefficient designs if these parameters are misspecified [see for example Dette et al. (2013), Example 2.1]. As a robust alternative Pronzato and Walter (1985) and Chaloner and Larntz (1989) proposed Bayesian optimal designs which maximize an expectation of the information criterion with respect to a prior distribution for the unknown parameters [see also Chaloner and Verdinelli (1995) for a review]. Bayesian optimal designs for various prior distributions have been discussed by numerous authors [see Haines (1995), Dette and Neugebauer (1997), Han and Chaloner (2003) or Braess and Dette (2007) among others]. However, there exist many applications where the specification of a prior distribution is difficult and several authors advocate the use of a uniform prior as a pragmatic approach if no preliminary knowledge about the unknown parameter is available. In a recent paper it was pointed out by Bornkamp (2012) that for several models the use of a uniform prior as a non-informative prior does not yield reasonable designs. This author proposed the concept of a functional uniform prior in order to construct Bayesian optimality criteria with respect to non-informative prior distributions.

In this paper we consider two alternative criteria for the construction of Bayesian optimal designs with respect to non-informative prior distributions. Roughly speaking, the criteria maximize the predicted Kullback-Leibler distance between the prior and the posterior distribution for the unknown parameter of the model with respect to the choice of the experimental design, where – in contrast to the classical approach to Bayesian optimality – the prior distribution depends also on the design of experiment. The criteria are introduced in Section 2, which also gives an introduction into the field of optimal experimental design. Here it is demonstrated that Bayesian optimal design problems corresponding to non-informative priors are in general not convex. Necessary conditions for the optimality of a given design are also derived. In Section 4 we use the theory of canonical moments which is introduced in Section 3 [see also Dette and Studden (1997)] in order to determine saturated Bayesian optimal designs with respect to non-informative priors for polynomial regression models with a heteroscedastic error structure. Finally, in Section 5 we consider two frequently used nonlinear regression models and compare the optimal designs with respect to the new criteria proposed in this paper with optimal designs with respect to "classical" Bayesian optimality criteria based on a uniform and a functional uniform distribution.

2 Optimal design and non-informative priors

An approximate design is defined as a probability measure ξ on the design space $\mathcal{X}$ with finite support [see Kiefer (1974)]. If the design ξ has masses ξ_i at the points x_i $(i = 1, \ldots, m)$ and N observations can be made by the experimenter, this means that the quantities $\xi_i N$ are rounded to integers, say n_i, satisfying $\sum_{i=1}^m N_i = N$, and the experimenter takes N_i observations at each location x_i $(i = 1, \ldots, m)$. The corresponding design with masses N_i/N at the points x_i $(i = 1, \ldots, m)$ will be denoted as exact design ξ_N. Assume that ξ_N is an exact design with masses N_i/N at points x_i $(i = 1, \ldots, m)$ and that N_i independent observations $Y_{i1}, \ldots, Y_{iN_i}$ are taken at each x_i with density

$$p(y_{ij}|\boldsymbol{\theta}, x_i) \; ; \quad j = 1, \ldots, N_i, \; i = 1, \ldots, m; \tag{2.1}$$

such that

$$\lim_{N\to\infty} \frac{N_i}{N} = \xi_i > 0, \qquad i = 1, \ldots, m, \tag{2.2}$$

where $\boldsymbol{\theta} \in \Theta$ is a k-dimensional parameter. If ξ_N denotes the design with masses N_i/N at x_i $(i = 1, \ldots, m)$ we define by

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \xi_N) = \prod_{i=1}^m \prod_{j=1}^{N_i} p(y_{ij}|\boldsymbol{\theta}, x_i)$$

the joint density of the N-dimensional vector $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{mN_m})^T$. In the following we assume that the prior distribution for the parameter $\boldsymbol{\theta}$ may depend on the design (such as the Jeffreys prior) and consider the problem of maximizing the expected Kullback-Leibler distance between the prior and posterior distribution with respect to the choice of the design ξ_N, that is

$$U(\xi_N) = \int \log\Big(\frac{p(\boldsymbol{\theta}|\boldsymbol{y}, \xi_N)}{p(\boldsymbol{\theta}|\xi_N)}\Big) \cdot p(\boldsymbol{y}, \boldsymbol{\theta}|\xi_N) \; d\boldsymbol{\theta} dy. \tag{2.3}$$

Here $p(\boldsymbol{\theta}|\xi_N)$ denotes the density of the prior distribution of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\boldsymbol{y}, \xi_N)$ the density of the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{y}$ and $p(\boldsymbol{y}, \boldsymbol{\theta}|\xi_N)$ is the density of the joint distribution of $(\boldsymbol{Y}, \boldsymbol{\theta})$. Note that all distributions may depend on the design ξ_N.
Under regularity assumptions it can be shown by similar arguments as in Chaloner and Verdinelli (1995) that the expected Kullback-Leibler distance can be approximated by

$$\begin{aligned} U(\xi_N) \; \approx \; & -\frac{k}{2}\log(2\pi) - \frac{k}{2} + \frac{1}{2}\int \log\Big(|NM(\boldsymbol{\theta}, \xi_N|\Big) p(\boldsymbol{\theta}|\xi_N) \; d\boldsymbol{\theta} \\ & - \int \log p(\boldsymbol{\theta}|\xi_N) \; p(\boldsymbol{\theta}|\xi_N) \; d\boldsymbol{\theta}, \end{aligned} \tag{2.4}$$

where

$$M(\xi_N, \boldsymbol{\theta}) = \int \Big(\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}|\boldsymbol{\theta}, x)\Big)\Big(\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}|\boldsymbol{\theta}, x)\Big)^T p(\boldsymbol{y}|\boldsymbol{\theta}, x) dy \xi_N(dx) \tag{2.5}$$

denotes the Fisher information matrix. If the prior distribution of $\boldsymbol{\theta}$ does not depend on the design, then the criterion for Bayesian D-optimality arises, i.e.

$$\Phi_D(\xi) = \int \log(|M(\xi, \boldsymbol{\theta})|) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{2.6}$$

We call the designs maximizing the criterion (2.6) *Bayesian D-optimal designs with respect to the prior p*. A noninformative prior often used in applications is the uniform prior, i.e.

$$p_{\text{uni}}(\boldsymbol{\theta}) \propto 1. \tag{2.7}$$

Bornkamp (2012) pointed out some deficits of this prior and proposed Bayesian D-optimal designs with respect to functional uniform priors

$$p_{\text{funct}}(\boldsymbol{\theta}) = \frac{\int_{\mathcal{X}} |M(\delta_x, \boldsymbol{\theta})|^{1/2} dx}{\int \int_{\mathcal{X}} |M(\delta_x, \boldsymbol{\theta})|^{1/2} dx d\boldsymbol{\theta}}, \tag{2.8}$$

where here and throughout this paper δ_x denotes the Dirac measure at the point $x \in \mathcal{X}$. As stated in Chaloner and Verdinelli (1995) a necessary and sufficient condition for Bayesian D-optimality is given by the following theorem.

Theorem 2.1 *A design ξ^* is Bayesian D-optimal if and only if the inequality*

$$\int \text{tr}\big\{ M^{-1}(\xi^*, \boldsymbol{\theta}) M(\xi_x, \boldsymbol{\theta}) \big\} p(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq k \tag{2.9}$$

holds for all $x \in \mathcal{X}$. Moreover, there is equality for all support points of the design ξ^*.

In the context of Bayesian analysis priors depending on the design are frequently used. A typical example is the Jeffreys prior [see Jeffreys (1946)]

$$p^J(\boldsymbol{\theta}|\xi) = \frac{|M(\boldsymbol{\theta}, \xi)|^{1/2}}{\int |M(\boldsymbol{t}, \xi)|^{1/2} d\boldsymbol{t}} \approx \frac{|NM(\boldsymbol{\theta}, \xi_N)|^{1/2}}{\int |NM(\boldsymbol{t}, \xi_N)|^{1/2} d\boldsymbol{t}}. \tag{2.10}$$

Using the Jeffreys prior the expression (2.4) reduces to

$$U(\xi_N) \approx V(\xi) = -\frac{k}{2} \log(2\pi) - \frac{k}{2} + \log(\int |NM(\xi, \boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}).$$

Consequently, we call an approximate design ξ *Bayesian optimal with respect to the Jeffreys prior* if ξ maximizes the functional

$$\Phi_J(\xi) = \int |M(\xi, \boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}, \tag{2.11}$$

where we assume throughout this paper that the integral in (2.11) is finite for all approximate designs (sufficient for this property are compactness assumptions regarding the parameter space

and continuity of the information matrix with respect to the parameter). This criterion for the choice of an experimental design has been sporadically discussed in the literature before [see Polson (1992) or Firth and Hinde (1997a,b)].
An intrinsic difficulty in these optimization problems consists in the fact that the criterion Φ_J is in general not convex. Consequently, standard optimal design theory based on convex optimization is not directly applicable. Nevertheless, the following results provide a necessary condition for optimality with respect to this criterion. A proof can be found in Firth and Hinde (1997b).

Theorem 2.2 *If a design ξ^* is Bayesian optimal with respect to the Jeffreys prior, then the inequality*

$$\int \mathrm{tr}\big\{M^{-1}(\xi^*,\boldsymbol{\theta})M(\xi_x,\boldsymbol{\theta})\big\}|M(\xi^*,\boldsymbol{\theta})|^{1/2}d\boldsymbol{\theta} \leq k\int |M(\xi^*,\boldsymbol{\theta})|^{1/2}d\boldsymbol{\theta}$$

holds for all $x \in \mathcal{X}$. Moreover, there is equality for all support points of the optimal design ξ^*.

The next Bayesian optimality criterion with respect to a non-informative prior distribution is motivated by the fact that not all components of the vector $\boldsymbol{\theta}$ are of equal importance. To be precise, we use similar arguments as in Berger and Bernardo (1992) and decompose the parameter $\boldsymbol{\theta}$ into $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ where $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ are k_1 and k_2-dimensional parameters, respectively, and $k = k_1 + k_2$. The information matrix $M(\xi,\boldsymbol{\theta})$ is decomposed in a similar way, that is

$$M(\xi,\boldsymbol{\theta}) = \begin{pmatrix} M_{11}(\xi,\boldsymbol{\theta}) & M_{12}(\xi,\boldsymbol{\theta}) \\ M_{21}(\xi,\boldsymbol{\theta}) & M_{22}(\xi,\boldsymbol{\theta}) \end{pmatrix}, \tag{2.12}$$

where $M_{ij}(\xi,\boldsymbol{\theta}) \in \mathbb{R}^{k_i \times k_j}$ $(i,j = 1,2)$. In the following we assume that $\boldsymbol{\theta_2}$ is a nuisance parameter and that the parameter $\boldsymbol{\theta_1}$ is of primary interest to the experimenter.
This approach results in a criterion where the marginal expected Kullback-Leibler distance between the prior and posterior distribution of the parameter of primary interest $\boldsymbol{\theta_1}$ is maximized with respect to the choice of the experimental design ξ_N, that is

$$\begin{aligned} U_1(\xi_N) &= \int\int \log\Big(\frac{p(\boldsymbol{\theta_1}|\boldsymbol{y},\xi_N)}{p(\boldsymbol{\theta_1}|\xi_N)}\Big)p(\boldsymbol{\theta_1},\boldsymbol{y}|\xi_N)d\boldsymbol{\theta_1}d\boldsymbol{y} \\ &= \int\int \log\Big(\frac{p(\boldsymbol{\theta}|\boldsymbol{y},\xi_N)}{p(\boldsymbol{\theta}|\xi_N)}\Big)p(\boldsymbol{\theta},\boldsymbol{y}|\xi_N)d\boldsymbol{\theta}d\boldsymbol{y} \\ &\quad - \int\int \log\Big(\frac{p(\boldsymbol{\theta_2}|\boldsymbol{\theta_1},\boldsymbol{y},\xi_N)}{p(\boldsymbol{\theta_2}|\boldsymbol{\theta_1},\xi_N)}\Big)p(\boldsymbol{\theta},\boldsymbol{y}|\xi_N)d\boldsymbol{\theta}d\boldsymbol{y}. \end{aligned}$$

Under regularity assumptions it can be shown that the marginal expected Kullback-Leibler distance can be approximated by

$$\begin{aligned} U_1(\xi_N) \approx\ & \frac{1}{2}\int \log\Big(\exp\Big\{\int p(\boldsymbol{\theta_2}|\boldsymbol{\theta_1},\xi_N)\log\Big(\frac{|M(\xi_N,\boldsymbol{\theta})|^{1/2}}{|M_{22}(\xi_N,\boldsymbol{\theta})|^{1/2}}\Big)d\boldsymbol{\theta_2}\Big\}\Big)p(\boldsymbol{\theta_1}|\xi_n)\ d\boldsymbol{\theta_1} \\ & + \frac{k_1}{2}\log\Big(\frac{N}{2\pi e}\Big) - \int \log p(\boldsymbol{\theta_1}|\xi_N)\ p(\boldsymbol{\theta_1}|\xi_N)\ d\boldsymbol{\theta_1}. \end{aligned} \tag{2.13}$$

This follows by similar arguments as in equation (2.2) [see Berger and Bernardo (1992) or Ghosh and Mukerjee (1992)].
Following Berger and Bernardo (1992) we decompose the prior for the parameter $\boldsymbol{\theta}$ (which may depend on the experimental design) as

$$p(\boldsymbol{\theta}|\xi_N) = p(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi_N) p(\boldsymbol{\theta_1}|\xi_N),$$

where $p(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi_N)$ denotes the conditional density of the distribution of $\boldsymbol{\theta_2}$ given $\boldsymbol{\theta_1}$ and $p(\boldsymbol{\theta_1}|\xi_N)$ is the density of the prior distribution for $\boldsymbol{\theta_1}$. More precisely, for the conditional density of $\boldsymbol{\theta_2}$ given $\boldsymbol{\theta_1}$ an analogue of the Jeffreys prior is used, that is

$$p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi_N) = \frac{|M_{22}(\xi_N, \boldsymbol{\theta_1}, \boldsymbol{\theta_2})|^{1/2}}{\int |M_{22}(\xi_N, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2} d\boldsymbol{t_2}}, \tag{2.14}$$

while the density of the prior distribution for $\boldsymbol{\theta_1}$ is given by

$$p^{BB}(\boldsymbol{\theta_1}|\xi_N) = \exp\Big\{ \int p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi_N) \log\Big(\frac{|M(\xi_N, \boldsymbol{\theta})|^{1/2}}{|M_{22}(\xi_N, \boldsymbol{\theta})|^{1/2}}\Big) d\boldsymbol{\theta_2} \Big\} \cdot \alpha \tag{2.15}$$

where

$$\alpha = \Big(\int \exp\Big\{ \int p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi_N) \log\Big(\frac{|M(\xi_N, \boldsymbol{\theta})|^{1/2}}{|M_{22}(\xi_N, \boldsymbol{\theta})|^{1/2}}\Big) d\boldsymbol{\theta_2} \Big\} d\boldsymbol{\theta_1} \Big)^{-1} \tag{2.16}$$

is a normalizing constant. Since this pioneering work on the construction of reference priors, several authors have worked on this subject and we refer to the work of Clarke and Wasserman (1993) and Kass and Wasserman (1996) for a general discussion on this subject. Combining this prior and equation (2.13) yields the following optimality criterion

$$\Phi_{BB}(\xi) = \int \exp\Big\{ \int \frac{|M_{22}(\xi, \boldsymbol{\theta_1}, \boldsymbol{\theta_2})|^{1/2}}{\int |M_{22}(\xi, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2} d\boldsymbol{t_2}} \log\Big(\frac{|M(\xi, \boldsymbol{\theta_1}, \boldsymbol{\theta_2})|^{1/2}}{|M_{22}(\xi, \boldsymbol{\theta_1}, \boldsymbol{\theta_2})|^{1/2}}\Big) d\boldsymbol{\theta_2} \Big\} d\boldsymbol{\theta_1}, \tag{2.17}$$

where we again assume that the integral exists for all designs ξ. Designs maximizing the function Φ_{BB} are called *Bayesian optimal with respect to the Berger-Bernardo prior*. Again this criterion is in general not convex and a necessary condition for optimality will be derived.

Theorem 2.3 *If a design ξ^* is Bayesian-optimal with respect to the Berger-Bernardo prior, then the inequality*

$$\begin{aligned}
d(\xi^*, \delta_x) \;=\; & \int\int \operatorname{tr}\big[M_{22}^{-1}(\xi^*, \boldsymbol{\theta}) M_{22}(\delta_x, \boldsymbol{\theta})\big] \log\Big(\frac{|M(\xi^*, \boldsymbol{\theta})|^{1/2}}{|M_{22}(\xi^*, \boldsymbol{\theta})|^{1/2}}\Big) p^{BB}(\boldsymbol{\theta}, \xi^*) d\boldsymbol{\theta} \\
- \; & \int \Big\{ \int \log\Big(\frac{|M(\xi^*, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2}}{|M_{22}(\xi^*, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2}}\Big) p^{BB}(\boldsymbol{t_2}|\boldsymbol{\theta_1}, \xi^*) d\boldsymbol{t_2} \\
& \times \int \operatorname{tr}(M_{22}^{-1}(\xi^*, \boldsymbol{\theta_1}, \boldsymbol{t_2}) M_{22}(\delta_x, \boldsymbol{\theta_1}, \boldsymbol{t_2})) p^{BB}(\boldsymbol{t_2}|\boldsymbol{\theta_1}, \xi^*) d\boldsymbol{t_2} \Big\} p^{BB}(\boldsymbol{\theta_1}|\xi^*) d\boldsymbol{\theta_1} \\
+ \; & \int\int \big[\operatorname{tr}(M^{-1}(\xi^*, \boldsymbol{\theta}) M(\delta_x, \boldsymbol{\theta})) - \operatorname{tr}(M_{22}^{-1}(\xi^*, \boldsymbol{\theta}) M_{22}(\xi_x, \boldsymbol{\theta}))\big] p^{BB}(\boldsymbol{\theta}|\xi^*) d\boldsymbol{\theta} \leq k_1
\end{aligned}$$

holds for all $x \in \mathcal{X}$, *where* $p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi)$ *and* $p^{BB}(\boldsymbol{\theta_1}|\xi)$ *are defined by* (2.14) *and* (2.15), *respectively, and* $p^{BB}(\boldsymbol{\theta}|\xi) = p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi)p^{BB}(\boldsymbol{\theta_1}|\xi)$. *Moreover, there is equality for all support points of the design* ξ^*.

Proof. The proof follows by a standard argument calculating the directional derivative

$$\frac{\partial}{\partial t}\Phi_{BB}(\xi_t)|_{t=0},$$

where the design ξ_t is defined by $\xi_t = \xi^* + t(\eta - \xi^*)$, η denotes an additional approximative design and $t \in (0, 1)$. Observing the fact

$$\frac{\partial}{\partial t}\log|M(\xi_t, \boldsymbol{\theta})|\Big|_{t=0} = \mathrm{tr}(M^{-1}(\xi^*, \boldsymbol{\theta})(M(\eta, \boldsymbol{\theta}) - M(\xi^*, \theta)))$$

we obtain (recalling the definitions (2.14), (2.15) and (2.16))

$$\begin{aligned}
&\frac{\partial}{\partial t}\Phi_{BB}(\xi_t)\Big|_{t=0} \\
&= \frac{1}{2\alpha}\int p^{BB}(\boldsymbol{\theta_1}|\xi^*)\int\Big[\log(\frac{|M(\xi^*, \boldsymbol{\theta})|^{1/2}}{|M_{22}(\xi^*, \boldsymbol{\theta})|^{1/2}})\Big\{p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi^*)\mathrm{tr}(M_{22}^{-1}(\xi^*, \boldsymbol{\theta})(M_{22}(\eta, \boldsymbol{\theta}) - M_{22}(\xi^*, \boldsymbol{\theta})) \\
&-p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi^*)\int p^{BB}(\boldsymbol{t_2}|\boldsymbol{\theta_1}, \xi^*)\mathrm{tr}(M_{22}^{-1}(\xi^*, \boldsymbol{\theta_1}, \boldsymbol{t_2})(M_{22}(\eta, \boldsymbol{\theta_1}, \boldsymbol{t_2}) - M_{22}(\xi^*, \boldsymbol{\theta_1}, \boldsymbol{t_2}))d\boldsymbol{t_2}\Big\} \\
&+p^{BB}(\boldsymbol{\theta_2}|\boldsymbol{\theta_1}, \xi^*)\Big\{\mathrm{tr}(M^{-1}(\xi^*, \boldsymbol{\theta})(M(\eta, \boldsymbol{\theta}) - M(\xi^*, \boldsymbol{\theta}))) \\
&-\mathrm{tr}(M_{22}^{-1}(\xi^*, \boldsymbol{\theta})(M_{22}(\eta, \boldsymbol{\theta}) - M_{22}(\xi^*, \boldsymbol{\theta}))\Big\}\Big]d\boldsymbol{\theta_2}d\boldsymbol{\theta_1} \\
&= \frac{1}{2\alpha}(d(\xi, \eta) - k_1).
\end{aligned}$$

The assertion now follows by the same arguments as given in (Silvey, 1980, p.19). □

In the following chapters we will discuss optimal designs maximizing the criteria (2.11) and (2.17) in several examples.

3 Canonical moments

In Section 4 we discuss Bayesian optimal designs with respect to non-informative priors for heteroscedastic polynomial regression models. An important tool to derive optimal saturated designs for polynomial models is the theory of canonical moments which was firstly used by Studden (1980, 1982b) to determine D_s-optimal designs for homoscedastic polynomial regression explicitly and will be briefly introduced in this section. Since these seminal papers numerous authors have used this methodology to determine optimal designs in polynomial and trigonometric regression models [see Lau and Studden (1985), Spruill (1990), Dette (1994, 1995), and Zen and Tsai (2004) among many others]. A detailed description of the theory of canonical moments can be found in the monograph of Dette and Studden (1997).

To be precise let $a, b \in \mathbb{R}$ denote two constants such that $a < b$ and introduce by $\mathcal{P}([a,b])$ the set of all probability measures on the interval $[a,b]$. We define for a design $\xi \in \mathcal{P}([a,b])$ its moments by

$$c_i = c_i(\xi) = \int_a^b x^i \xi(dx), \ i = 1, 2 \ldots .$$

Define $\mathcal{M}_n = \{(c_1, \ldots, c_n)^T \mid \xi \in \mathcal{P}([a,b])\}$ as the nth moment space and $\Phi_n(x) = (x, \ldots, x^n)$ as the vector of monomials of order n. Consider for a fixed vector $\boldsymbol{c}_n = (c_1, \ldots, c_n)^T \in \mathcal{M}_n$ the set

$$\mathcal{S}_n(\boldsymbol{c}_n) = \Big\{ \xi \in \mathcal{P}([a,b]) : \ \int_a^b \Phi_n(x) \xi(dx) = \boldsymbol{c}_n \Big\}$$

of all probability measures on the interval $[a,b]$ whose moments up to the order n coincide with $\boldsymbol{c}_n = (c_1, \ldots, c_n)^T$. For $n = 2, 3, \ldots$ and for a given point $(c_1, \ldots, c_{n-1})^T \in \mathcal{M}_{n-1}$ we define $c_n^+ = c_n^+(c_1, \ldots, c_{n-1})$ and $c_n^- = c_n^-(c_1, \ldots, c_{n-1})$ as the largest and smallest value of c_n such that $(c_1, \ldots, c_n)^T \in \partial \mathcal{M}_n$ (here $\partial \mathcal{M}_n$ denotes the boundary of $\mathcal{M}_n$), that is

$$\begin{aligned} c_n^- &= \min\Big\{ \int_a^b x^n \xi(dx) \mid \xi \in S_{n-1}(c_1, \ldots, c_{n-1}) \Big\}, \\ c_n^+ &= \max\Big\{ \int_a^b x^n \xi(dx) \mid \xi \in S_{n-1}(c_1, \ldots, c_{n-1}) \Big\}. \end{aligned}$$

Note that $c_n^- \le c_n \le c_n^+$ and that both inequalities are strict if and only if $(c_1, \ldots, c_{n-1})^T \in \text{int}(\mathcal{M}_{n-1})$ where $\text{int}(\mathcal{M}_{n-1})$ denotes the interior of the set $\mathcal{M}_{n-1}$ [see Dette and Studden (1997)].
For a design ξ on the interval $[a,b]$ with corresponding moment point $\boldsymbol{c}_n = (c_1(\xi), \ldots, c_n(\xi))^T$, such that $\boldsymbol{c}_{n-1} = (c_1(\xi), \ldots, c_{n-1}(\xi))^T$ is in the interior of the moment space $\mathcal{M}_{n-1}$, the canonical moments or canonical coordinates are defined by $p_1 = c_1(\xi)$ and

$$p_i = p_i(\xi) = \frac{c_i(\xi) - c_i^-}{c_i^+ - c_i^-}, \quad i = 2, \ldots, n \ . \tag{3.1}$$

Note that the canonical moments p_i vary independently in the interval $[0,1]$ (whenever they are defined). Moreover, it follows that $p_i \in (0,1)$, $i = 1, \ldots, n-1$ and $p_n \in \{0,1\}$ if and only if $(c_1(\xi), \ldots, c_{n-1}(\xi)) \in \text{int}(\mathcal{M}_{n-1})$ and $(c_1(\xi), \ldots, c_n(\xi))^T \in \partial \mathcal{M}_n$. In this case the canonical moments p_i of order $i > n$ remain undefined.
The main idea of Studden (1980) was to describe designs in terms of their canonical moments, to find a (simple) representation of the optimality criterion by these quantities and to perform optimization on the unit cube. For this purpose optimality criteria have to be expressed explicitly in terms of canonical moments and we recall the following basic facts [for a proof see Studden (1982b,a) and Lau and Studden (1988)].

Theorem 3.1 *Let ξ denote a design on the interval $[a,b]$ with moments $c_1, c_2, \ldots$, canonical moments $p_1, p_2, \ldots$ and $q_0 = 1, q_1 = 1 - p_1, q_2 = 1 - p_2, \ldots$*

(a) Let $H_n(\xi) = (c_{i+j})_{i,j=1,\ldots,n}$ denote the Hankel matrix of the moments of the design ξ. If if $(c_1, \ldots, c_{2n-1})^T \in int(\mathcal{M}_{2n-1})$, then

$$|H(\xi)| = (b-a)^{n(n+1)} \prod_{i=1}^{n} (q_{2i-2} p_{2i-1} q_{2i-1} p_{2i})^{n-i+1}.$$

(b) Let ξ denote a design on the interval $[a,b]$ with $n+1$ support points $x_1, \ldots x_{n+1}$, then

$$\prod_{i=1}^{n+1}(x_i - a) = (b-a)^{n+1} p_{2n+1} \prod_{i=1}^{n} p_{2i-1} q_{2i} \ , \quad \prod_{i=1}^{n+1}(b - x_i) = (b-a)^{n+1} \prod_{i=1}^{2n+1} q_i,$$
$$\sum_{i=1}^{n+1}(x_i - a) = (b-a) \sum_{i=1}^{2n+1} q_{i-1} p_i.$$

The following results are shown in Dette and Studden (1997) and can be used to derive a design corresponding to an "optimal" sequence of canonical moments (i.e. a sequence maximizing a particular optimality criterion).

Theorem 3.2 *Let ξ denote a design on the interval $[a,b]$ with canonical moments $p_1, .p_2, \ldots$.*

(1) If $p_i \in (0,1)$, $i = 1, \ldots, 2n-1$ and $p_{2n} = 0$, then ξ has $m = n$ support points in the interior of the interval (a,b).

(2) If $p_i \in (0,1)$, $i = 1, \ldots, 2n$ and $p_{2n+1} = 0$, then ξ has $m = n+1$ support points, n points in the interior of the interval (a,b) and the point a.

(3) If $p_i \in (0,1)$, $i = 1, \ldots, 2n$ and $p_{2n+1} = 1$, then ξ has $m = n+1$ support points, n points in the interior of the interval (a,b) and the point b.

(4) If $p_i \in (0,1)$, $i = 1, \ldots, 2n-1$ and $p_{2n} = 1$, then ξ has $m = n+1$ support points, $n-1$ points in the interior of the interval (a,b) and the points a and b.

Moreover, the support points $x_1, \ldots, x_m$ are the roots of the polynomial $P_m(x) = W_m(x)$, where the polynomials $W_i(x)$ are defined recursively by

$$W_{i+1} = (x - a - (b-a)(\zeta_{2i} + \zeta_{2i+1}))W_i(x) - (b-a)^2 \zeta_{2i-1}\zeta_{2i} W_{i-1}(x), \tag{3.2}$$

with initial conditions $W_0(x) = 1, W_{-1}(x) = 0$ and we use the notation $\zeta_0 = 0, \zeta_1 = p_1, \zeta_i = (1 - p_{i-1})p_i, i \geq 2$. The weights $\xi(x_1), \ldots, \xi(x_m)$ at the support points $x_1, \ldots, x_m$ are obtained by the formula

$$\xi(x_i) = \frac{P^{(1)}_{m-1}(x_i)}{\frac{\partial}{\partial x} P_m(x)|_{x=x_i}} \ ; \quad i = 1, \ldots, m, \tag{3.3}$$

where $P^{(1)}_i(x) = W_{i+1}(x)$ and the polynomials $W_i(x)$ are defined recursively by (3.2) with initial conditions $W_1(x) = 1, W_0(x) = 0$.

4 Robust designs for heteroscedastic polynomials

We are now in a position to determine Bayesian optimal saturated designs with respect to non-informative priors for the polynomial regression model. To be precise, we assume that the density $p(\boldsymbol{y}|\boldsymbol{\theta}, x)$ of the response Y (at experimental condition x) is governed by a by normal distribution with mean

$$\mu(x, \boldsymbol{\theta}) = \sum_{j=0}^{n} \theta_j x^j \tag{4.1}$$

and variance $\sigma^2(x, \boldsymbol{\theta})$, where the variance and design space are given by

$$\sigma^2(x, \boldsymbol{\theta}) \;=\; \theta_{n+1} \exp(\theta_{n+2} x) \;, \quad \mathcal{X} = [0, b] \;\; (\theta_{n+1} > 0, \theta_{n+2} \geq 0) \tag{4.2}$$

$$\sigma^2(x, \boldsymbol{\theta}) \;=\; (1-x)^{-\theta_{n+1}-1}(1+x)^{-\theta_{n+2}-1} \;, \quad \mathcal{X} = (-1, 1) \;\; (\theta_{n+1}, \theta_{n+2} > 0) \tag{4.3}$$

and $b > 0$ is a constant. We also note that there are several other variance functions, which are usually investigated in the context of polynomial regression [see Karlin and Studden (1966), p. 328, Chang (2005) or Chang et al. (2009)]. For these variance functions similar results to those described in the following section can be obtained, but the details are omitted for the sake of brevity.
Adapting the notation of the previous section we have for the parameter of interest $\boldsymbol{\theta_1} = (\theta_0, \ldots, \theta_n)^T$ and for the nuissance parameters $\boldsymbol{\theta_2} = (\theta_{n+1}, \theta_{n+2})^T$. The Fisher information at a point $x \in \mathcal{X}$ is given by

$$I(x, \boldsymbol{\theta}) = \begin{pmatrix} I_{11}(x, \boldsymbol{\theta}) & I_{12}(x, \boldsymbol{\theta}) \\ I_{21}(x, \boldsymbol{\theta}) & I_{22}(x, \boldsymbol{\theta}) \end{pmatrix} \in \mathbb{R}^{n+3 \times n+3}, \tag{4.4}$$

where $I_{12}(x, \boldsymbol{\theta}) = 0 \in \mathbb{R}^{n+1 \times 2}$ and

$$I_{11}(x, \boldsymbol{\theta}) \;=\; \sigma^{-2}(x, \boldsymbol{\theta})(x^{i+j})_{i,j=0,\ldots,n} \in \mathbb{R}^{n+1 \times n+1}, \tag{4.5}$$

$$I_{22}(x, \boldsymbol{\theta}) \;=\; \frac{1}{2\sigma^2(x, \boldsymbol{\theta})} \Big(\frac{\partial}{\partial \boldsymbol{\theta_2}} \sigma^2(x, \boldsymbol{\theta}) \Big) \Big(\frac{\partial}{\partial \boldsymbol{\theta_2}} \sigma^2(x, \boldsymbol{\theta}) \Big)^T \in \mathbb{R}^{2 \times 2}. \tag{4.6}$$

In the following we call a design optimal m-point design, if it maximizes a particular optimality criterion in the class of all designs supported at m points. Our first result describes the class of all Bayesian-optimal $(n+1)$-point designs for polynomial regression and variance function (4.2) with respect to the Jeffreys and the Berger-Bernardo prior.

Theorem 4.1 *Consider the polynomial regression model* (4.1) *with variance function* (4.2) *and design space* $\mathcal{X} = [0, b]$.

(1) Assume that $(\theta_0, \ldots, \theta_{n+2}) \in \Theta \subset (\mathbb{R}_0^+)^{(n+1)} \times \mathbb{R}^+ \times \mathbb{R}_0^+$, *where* Θ *is a compact set. The canonical moments of the Bayesian optimal* $(n+1)$*-design with respect to the Jeffreys prior*

are given by $(p_1, \ldots, p_{2n-1}, 1)$*, where* $p_1, \ldots, p_{2n-1} \in (0,1)$ *are obtained as a solution of the system of equations*

$$0 = \frac{n-i+1}{2}(\frac{1}{p_{2i-1}} - \frac{1}{q_{2i-1}}) + (p_{2i} - q_{2i-2})(\sum_{j=1}^{2n} q_{j-1}p_j)^{-1} \ (i = 2, ..., n)$$

$$0 = \frac{n-i+1}{2p_{2i}} - \frac{n-i}{2q_{2i}} + (p_{2i+1} - q_{2i-1})(\sum_{j=1}^{2n} q_{j-1}p_j)^{-1} \ (i = 2, ..., n-1)$$

$$0 = \frac{n+1}{2}(\frac{1}{p_1} - \frac{1}{q_1}) + (p_2 - 1)(\sum_{j=1}^{2n} q_{j-1}p_j)^{-1}$$

$$0 = \frac{n+1}{2p_2} - \frac{n-1}{2q_2} + (p_3 - q_1)(\sum_{j=1}^{2n} q_{j-1}p_j)^{-1}.$$

(2) Assume that $(\theta_0, \ldots, \theta_{n+2}) \in \Theta \subset (\mathbb{R}_0^+)^{(n+1)} \times (\mathbb{R}^+)^2$ *is a compact set, denote by* z *the largest root of the* n*th Laguerre polynomial* $L_n^{(1)}(x)$ *and define*

$$\gamma = \frac{\int \theta_{n+2} d\theta_{n+2}}{\int d\theta_{n+2}}.$$

(a) If $b\gamma \geq z$*, then the Bayesian optimal* $(n+1)$*-design with respect to the Berger-Bernardo prior puts equal masses at the roots of the polynomial* $xL_n^{(1)}(x\gamma)$*.*

(b) If $b\gamma < z$*, then the canonical moments of the Bayesian optimal* $(n+1)$*-design with respect to Berger-Bernardo prior are obtained as a solution of the system of equations* $p_{2n} = 1$

$$\frac{n-i+1}{p_{2i-1}} - \frac{n-i+1}{1-p_{2i-1}} - b\gamma(1-p_{2i-2}) + b\gamma p_{2i} = 0 \ (i = 1, ..., n)$$

$$\frac{n-i+1}{p_{2i}} - \frac{n-i}{1-p_{2i}} - b\gamma(1-p_{2i-1}) + b\gamma p_{2i+1} = 0 \ (i = 1, ..., n-1)$$

with $q_0 = 0$*. Moreover, the optimal design has equal masses at its support points.*

Proof. Note that the lower diagonal block of the Fisher information is given by

$$I_{22}(x, \boldsymbol{\theta}) = \frac{1}{2\theta_{n+1}^2} \begin{pmatrix} 1 & \theta_{n+1}x \\ \theta_{n+1}x & \theta_{n+1}^2 x^2 \end{pmatrix}.$$

If ξ denotes a design with $n+1$ support points $x_1, \ldots, x_{n+1}$, then it follows from Theorem 3.1 that

$$\begin{aligned} |M_{11}(\xi, \boldsymbol{\theta})| &= \frac{|H(\xi)|}{(\theta_{n+1})^{n+1}} \exp(-b\theta_{n+2} \sum_{i=1}^{n+1} x_i) \\ &= (\frac{b^n}{\theta_{n+1}})^{n+1} \prod_{j=1}^{n} (q_{2j-2}p_{2j-1}q_{2j-1}p_{2j})^{n-j+1} \exp(-b\theta_{n+2} \sum_{j=1}^{2n+1} q_{j-1}p_j). \end{aligned} \tag{4.7}$$

Moreover, the canonical moments p_1 and p_2 are related to the moments c_1 and c_2 by $c_1 = bp_1$ and $c_2 = b^2(p_1 + q_1p_2)$, respectively [see Dette and Studden (1997)], which yields for the lower right block of the matrix $M(\xi, \boldsymbol{\theta})$ in (2.12)

$$|M_{22}(\xi, \boldsymbol{\theta})| = \frac{1}{4}\theta_{n+1}^{-2} b^2 p_1 p_2 q_1. \tag{4.8}$$

Consequently, Bayesian optimal designs with respect to the Jeffreys and the Bernardo-Berger prior depend only on the parameter θ_{n+2}, and only this dependence will be reflected in the optimality criterion.
For a proof of (1) note that the criterion (2.11) reduces to

$$\begin{aligned}
\Phi_J(\xi) \;=\; & \alpha_1 \int (\frac{\theta_{n+1}^{-\frac{n+3}{2}}}{2} b^{\frac{n(n+1)}{2}+1}) d\theta_{n+1} \int_{\mathbb{R}^+} \Big[\prod_{j=1}^{n} (q_{2j-2}p_{2j-1}q_{2j-1}p_{2j})^{\frac{n-j+1}{2}} \\
& \times \exp\Big(-\frac{b\theta_{n+2}}{2} \sum_{j=1}^{2n+1} q_{j-1}p_j \Big) (p_1p_2q_1)^{\frac{1}{2}} \Big] d\theta_{n+2} \\
=\; & \frac{\alpha_2}{b \sum_{j=1}^{2n+1} q_{j-1}p_j} \prod_{j=1}^{n} (q_{2j-2}p_{2j-1}q_{2j-1}p_{2j})^{\frac{n-j+1}{2}} (p_1p_2q_1)^{\frac{1}{2}}
\end{aligned}$$

with appropriate constants α_1 and α_2. Obviously this expression is maximized if $(p_1, \ldots, p_{2n-1}) \in (0,1)^{2n-1}$ and $q_{2n}p_{2n+1} = 0$, which can be achieved either by $p_{2n+1} = 0$ and $p_{2n} \in (0,1)$ or if $p_{2n} = 1$. Now assume that $p_{2n} \in (0,1)$ then $(p_1, \ldots p_{2n}) \in (0,1)^{2n}$ would be a solution of the system of equations $\frac{\partial}{\partial p_j} \log \Phi_J(\xi) = 0$, $j = 1, \ldots, 2n$. The derivative with respect to the coordinate p_{2n} yields the equation

$$\frac{\partial}{\partial p_{2n}} \log \Phi_J(\xi) = \frac{1}{2p_{2n}} - \frac{q_{2n-1}}{\sum_{j=1}^{2n} q_{j-1}p_j} = 0 \;,$$

which gives

$$q_{2n-1}p_{2n} = \sum_{j=1}^{2n-1} q_{j-1}p_j. \tag{4.9}$$

Inserting this expression in the partial derivative with respect to p_{2n-1} yields

$$\frac{\partial}{\partial p_{2n-1}} \log \Phi_J(\xi) \;=\; \frac{\frac{1}{p_{2n-1}} - \frac{1}{q_{2n-1}}}{2} - \frac{q_{2n-2} - p_{2n}}{\sum_{j=1}^{2n} q_{j-1}p_j} = \frac{\frac{1}{p_{2n-1}} - \frac{1}{q_{2n-1}}}{2} - \frac{q_{2n-2} - p_{2n}}{2p_{2n}q_{2n-1}} = 0 \;,$$

which is equivalent to $p_{2n}q_{2n-1} = q_{2n-2}p_{2n-1}$. Combining this equation with (4.9) gives

$$\sum_{j=1}^{2n-2} q_{j-1}p_j = 0,$$

which is a contradiction to the assumption $p_i \in (0,1)$ $(i = 1, \ldots, 2n)$. Consequently, we have $p_{2n} = 1$ and calculating $\frac{\partial}{\partial p_j} \log \Phi_J(\xi) = 0$ for $j = 1, \ldots, 2n-1$ gives the system of equation stated in part (1) of Theorem 4.1.
We now turn to a proof of part (2). Recall the representation (4.8), which yields for the first ratio of the determinants in criterion (2.17)

$$\frac{|M_{22}(\xi, \boldsymbol{\theta})|^{1/2}}{\int |M_{22}(\xi, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2} d\boldsymbol{t_2}} \;=\; \frac{\theta_{n+1}^{-1}}{\int t_{n+1}^{-1} dt_{n+1} \int dt_{n+2}} = \frac{\theta_{n+1}^{-1}}{\alpha_3},$$

where the last equality defines the constant α_3 in an obvious manner. We introduce the notation

$$\alpha_1 = \int d\boldsymbol{\theta_1} \;; \quad \alpha_2 = \int d\theta_{n+2} \;; \quad \alpha_4 = \int \theta_{n+2} d\theta_{n+2}.$$

Observing the fact that the Fisher information matrix is block diagonal we obtain

$$|M(\xi, \boldsymbol{\theta})|/|M_{22}(\xi, \boldsymbol{\theta})| = |M_{11}(\xi, \boldsymbol{\theta})|,$$

and (4.8) yields for the optimality criterion (2.17)

$$\begin{aligned} \Phi_{BB}(\xi) \;=\; \alpha_1 \exp\Big(\int \Big[\frac{n+1}{2\alpha_3} \theta_{n+1}^{-1} \log(\frac{1}{\theta_{n+1}})] d\boldsymbol{\theta_2} + \frac{1}{2} \big[n(n+1) \log(b) \\ + \sum_{j=1}^{n} (n-j+1) \log(q_{2j-2} p_{2j-1} q_{2j-1} p_{2j}) - \frac{b\alpha_4}{\alpha_2} \sum_{j=1}^{2n+1} q_{j-1} p_j \Big] \Big). \end{aligned}$$

Consequently, the Bayesian optimal $(n+1)$-point design with respect to the Berger-Bernardo prior is obtained by maximizing the expression

$$\sum_{j=1}^{n} (n-j+1) \log(q_{2j-2} p_{2j-1} q_{2j-1} p_{2j}) - \frac{b\alpha_4}{\alpha_2} \sum_{j=1}^{2n+1} q_{j-1} p_j$$

with respect to the canonical moments $(p_1, \ldots, p_{2n+1}) \in [0,1]^{2n+1}$ and identifying the design corresponding to these canonical moments by Theorem 3.2. But this problem has been solved by Dette and Wong (1998), and the assertion follows from Theorem 3.2 in this reference observing that $\gamma = \alpha_4/\alpha_2$. □

Example 4.1 In this example we illustrate the application of Theorem 4.1 by calculating Bayesian optimal designs with respect to non-informative priors in the polynomial regression model (4.1) with variance function (4.2). Recall that only the parameter θ_{n+2} appears in the optimality criterion in a non-trivial way and as a consequence Bayesian optimal designs depend only on prior information regarding this parameter. We assume that $\theta_{n+2} \in [0,4]$. In Table 1 we present Bayesian optimal 4-point designs for the cubic regression model on the interval $\mathcal{X} = [0,1]$ with respect to the Jeffreys prior, the Berger-Bernardo prior and Bayesian D-optimal 4-point design with respect to a uniform distribution. The Bayesian D-optimal design with respect to the uniform prior and Bayesian optimal design with respect to the Bernardo-Berger

prior are similar, where the latter puts less weights at the boundary of the design space. On the other hand the support points of the Bayesian optimal design with respect to the Jeffreys prior in the interior of the design space are larger. In Figure 1 we illustrate the application of Theorem 2.2 and 2.3. We observe that all designs satisfy the necessary condition for optimality.
Corresponding results for a quadratic polynomial regression model are depicted in Table 2, where the design space is now given by the interval $[0, 3]$. Here the right boundary point of the design space is a support point of the Bayesian optimal design with respect to the Jeffreys prior and the Bayesian D-optimal design with respect to the uniform prior. On the other hand the Bayesian optimal design with respect to the Berger-Bernardo prior does not contain the point 3 in its support. We observe from Figure 2 that not all designs satisfy the necessary condition of optimality. Therefore we maximized the criteria for 4-point designs numerically, and the corresponding designs are shown in Table 3. Only the criterion based on the Jeffreys prior yields a 3-point design while the other hand two criteria yield 4-point designs. Moreover, all designs meet the corresponding necessary condition for optimality (these results are not depicted for the sake of brevity).

(2.6) with (2.7)		(2.11)		(2.17)	
$\xi(x_i)$	x_i	$\xi(x_i)$	x_i	$\xi(x_i)$	x_i
0.2760	0	0.2809	0	0.25	0
0.2195	0.2072	0.2170	0.2347	0.25	0.2177
0.2082	0.6606	0.2114	0.7018	0.25	0.6497
0.2963	1	0.2907	1	0.25	1

Table 1: *Bayesian optimal 4-point designs with respect to non-informative priors for the cubic polynomial regression model on the interval* $[0, 1]$ *with variance structure* (4.2)*, where* $\theta_{n+2} \in [0, 4]$*. Left column: Bayesian D-optimal designs with respect to the uniform prior. Middle column: Bayesian optimal designs with respect to the Jeffreys prior. Right column: Bayesian optimal designs with respect to the Bernardo-Berger prior.*

We finally briefly discuss optimal designs with respect the variance function (4.3). In this case we are only able to determine the Bayesian optimal designs with respect to the Berger-Bernardo prior.

Theorem 4.2 *Consider the polynomial regression model* (4.1)*. If the design space and the variance function are given by* $\mathcal{X} = (-1, 1)$ *and by* (4.3)*, respectively, then the Bayesian optimal* $(n + 1)$*-design with respect to the Berger-Bernardo prior puts equal masses at the roots of the* $(n + 1)$*th Jacobi polynomial* $P_{n+1}^{(g_1,g_2)}(x)$*, where the parameters* g_1 *and* g_2 *are given by*

$$g_j = \frac{\int \theta_{n+j} d\boldsymbol{\theta_2}}{\int d\boldsymbol{\theta_2}} \; ; \quad j = 1, 2.$$

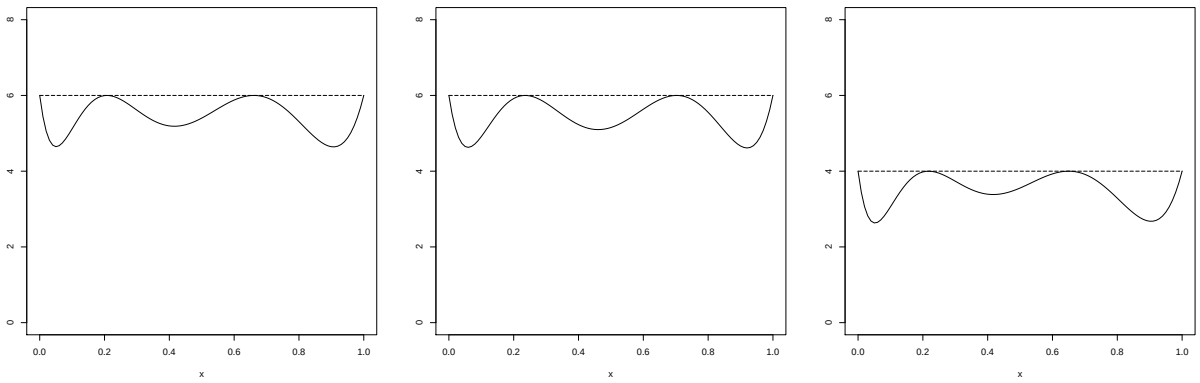


Figure 1: *The necessary condition of optimality for three Bayesian-optimal* 4*-point designs in the cubic polynomial regression model with variance structure* (4.2)*. Left: Bayesian D-optimal design with respect to a uniform prior. Middle: Bayesian optimal design with respect to the Jeffreys prior. Right: Bayesian optimal design with respect to the Berger-Bernardo-prior.*

(2.6) with (2.7)		(2.11)		(2.17)	
$\xi(x_i)$	x_i	$\xi(x_i)$	x_i	$\xi(x_i)$	x_i
0.3356	0	0.3624	0	0.3333	0
0.2686	0.6532	0.2527	1.1859	0.3333	0.6340
0.3958	3	0.3849	3	0.3333	2.36603

Table 2: *Bayesian optimal* 3*-point designs with respect to non-informative priors for the quadratic polynomial regression model on the interval* $[0, 3]$ *with variance structure* (4.2)*, where* $\theta_{n+2} \in [0, 4]$*. Left column: Bayesian D-optimal designs with respect to the uniform prior. Middle column: Bayesian optimal designs with respect to the Jeffreys prior. Right column: Bayesian optimal designs with respect to the Bernardo-Berger prior.*

Proof: Observing the representation (4.6) we obtain for the lower block in the Fisher information matrix $I(x, \boldsymbol{\theta})$ the representation

$$I_{22}(x, \boldsymbol{\theta}) = \begin{pmatrix} \log^2(1-x) & \log(x+1)\log(1-x) \\ \log(x+1)\log(1-x) & \log^2(x+1) \end{pmatrix}.$$

Therefore we have

$$\frac{|M_{22}(\xi, \boldsymbol{\theta})|^{1/2}}{\int |M_{22}(\xi, \boldsymbol{\theta_1}, \boldsymbol{t_2})|^{1/2} d\boldsymbol{t_2}} = \frac{1}{\int d\boldsymbol{\theta_2}} = \frac{1}{\alpha_2},$$

where the last equality defines the constant α_2 in an obvious manner. Consequently, for an $(n+1)$-point design with masses $\xi(x_0), \ldots, \xi(x_n)$ at the points $x_0, \ldots, x_n$ the optimality criterion

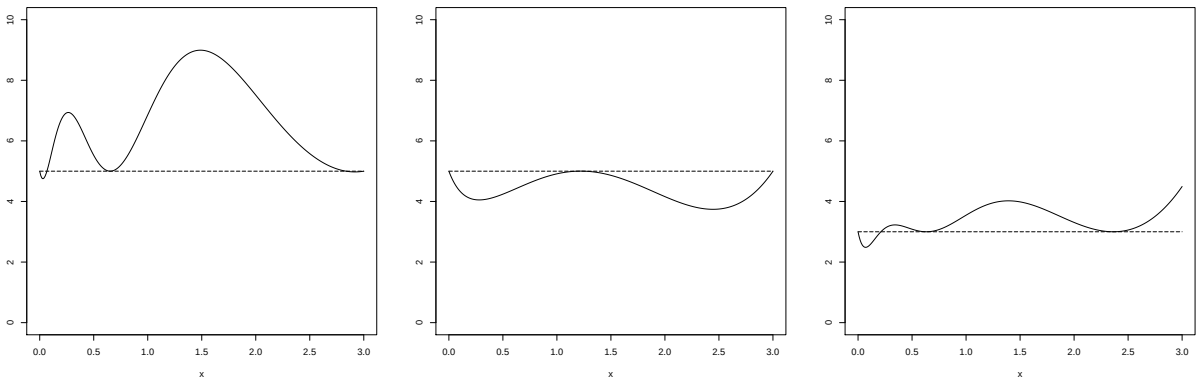


Figure 2: *The necessary condition of optimality for Bayesian optimal* 3*-point designs in the quadratic polynomial regression model with variance structure* (4.2)*. Left: Bayesian D-optimality. Middle: Bayesian optimal design with respect to the Jeffreys prior. Right: Bayesian optimal design with respect to the Berger-Bernardo-prior.*

(2.6) with (2.7)		(2.11)		(2.17)	
$\xi(x_i)$	x_i	$\xi(x_i)$	x_i	$\xi(x_i)$	x_i
0.3209	0	0.3624	0	0.3193	0
0.1931	0.4480	0.2527	1.1859	0.2478	0.4728
0.1601	1.2939	0.3849	3	0.2453	1.4472
0.3259	3			0.1876	3

Table 3: *Bayesian optimal* 3*- or* 4*- point designs with respect to non-informative priors for the quadratic polynomial regression model on the interval* $[0, 3]$ *with variance structure* (4.2)*, where* $\theta_{n+2} \in [0, 4]$*. Left column: Bayesian D-optimal designs with respect to the uniform prior. Middle column: Bayesian optimal designs with respect to the Jeffreys prior. Right column: Bayesian optimal designs with respect to the Bernardo-Berger prior.*

reduces to

$$\begin{aligned}\Phi_{BB}(\xi) &= \alpha_1 \exp\Big(\int \frac{1}{2\alpha_2} \log\Big[\prod_{j=0}^{n} \xi(x_j) \prod_{j=0}^{n} (1-x_j)^{\theta_{n+1}+1}(1+x_j)^{\theta_{n+2}+1} \prod_{\substack{m,\ell=0,\dots,n \\ m<\ell}} (x_m - x_\ell)^2\Big] d\boldsymbol{\theta_2}\Big) \\ &= \alpha_1 \exp\Big(\int \frac{1}{2\alpha_2} \log\Big[\prod_{j=0}^{n} (1-x_j)^{\theta_{n+1}+1}(1+x_j)^{\theta_{n+2}+1} |H_n(\xi)|\Big] d\boldsymbol{\theta_2}\Big),\end{aligned}$$

where $\alpha_1 := \int d\boldsymbol{\theta_1}$ and the matrix $H_n(\xi)$ is the Hankel matrix of the $(n+1)$-point design ξ, that is

$$H_n(\xi) = (c_{i+j}(\xi))_{i,j=0,\dots n} = \prod_{j=0}^{n} \xi(x_j) \prod_{\substack{m,\ell=0,\dots,n \\ m<\ell}} (x_m - x_\ell)^2.$$

Observing Theorem 3.2 it therefore follows that the Bayesian $(n+1)$-point optimal design with respect to the Bernardo-Berger prior can be determined by maximizing the expression

$$\Big(\prod_{i=1}^{2n+1} q_i\Big)^{g_1+1}\Big(p_{2n+1}\prod_{i=1}^{n}(p_{2i-1}q_{2i})\Big)^{g_2+1}\prod_{i=1}^{n}(q_{2i-2}p_{2i-1}q_{2i-1}p_{2i})^{n-i+1}$$

with respect to the canonical moments $p_1,\dots,p_{2n+1}$. Straightforward algebra gives for the corresponding "optimal" canonical moments

$$\begin{aligned} p_{2i-1} &= \frac{g_2+n+1-i}{g_1+g_2+2(n+1-i)}\ ; \quad i=1,\dots,n+1 \\ p_{2i} &= \frac{n+1-i}{g_1+g_2+2(n+1-i)+1}\ ; \quad i=1,\dots,n+1. \end{aligned}$$

The design corresponding to these canonical moments has been determined in Studden (1982a) and puts equal masses at the roots of the $(n+1)$th Jacobi polynomial $P_{n+1}^{(g_1,g_2)}(x)$ [see also Dette and Studden (1997) for an alternative proof], which completes the proof of Theorem 4.2. □

5 Bayesian optimal designs for nonlinear regression

In this section we illustrate the application of the methodology determining Bayesian optimal designs for the EMAX model and a compartment model, which are frequently used in pharmacology. Locally optimal designs for this model have been determined by numerous authors [see Atkinson et al. (1993), Jones et al. (1999), Dette et al. (2008) and Dette et al. (2010)] and we present some Bayesian optimal designs with respect to non-informative priors.
For both models we assume that the response at experimental condition $x \in \mathcal{X}$ is normally distributed with mean $\mu(x,\boldsymbol{\theta})$ and variance $\sigma^2(\boldsymbol{\theta}) = \theta_3 > 0$. Here the variance is considered as a nuissance parameter. For the criterion (2.6) we use a uniform and a functional uniform prior for the parameters $(\theta_0,\theta_1,\theta_2)$ and an arbitrary prior for θ_3. The criteria with respect to the Jeffreys prior and the Berger-Bernardo-prior are equivalent in this case. All designs have been calculated numerically using Maple.
We begin with the EMAX model which describes a dose-response relationship

$$\mu(x,\boldsymbol{\theta}) = \theta_0 + \frac{\theta_1 x}{x+\theta_2},$$

where θ_1 determines the asymptotic maximum effect, θ_2 the dose that gives half of the asymptotic maximum effect and θ_0 describes the effect of placebo.
In Table 4 we display some Bayesian optimal designs with respect to non-informative priors, where the design space is given by the interval $[0,4]$. For the parameters we assume $\theta_0 \geq 0, \theta_1 \in (0,5]$, and $\theta_2 \in [1,6]$, where θ_0 and θ_3 are each from a compact interval.
We observe that the Bayesian-optimal 3-point designs with respect to the Jeffreys prior and the Berger-Bernardo prior and the Bayesian D-optimal design with respect to the functional uniform prior look similar, while the Bayesian D-optimal design with respect to the uniform

(2.6) with (2.7)		(2.11)/(2.17)		(2.6) with (2.8)	
$\xi(x_i)$	x_i	$\xi(x_i)$	x_i	$\xi(x_i)$	x_i
0.333	0	0.333	0	0.333	0
0.333	1.2028	0.333	0.9472	0.333	0.9766
0.333	4	0.333	4	0.333	4

Table 4: *Bayesian optimal* 3*-point designs with respect to non-informative priors for the EMAX model on the interval* $[0, 4]$*. Left column: Bayesian D-optimal design with respect to the uniform prior. Middle column: Bayesian optimal designs with respect to the Jeffreys and Bayesian optimal designs with respect to the Bernardo-Berger prior. Right column: Bayesian D-optimal designs with respect to the functional uniform prior.*

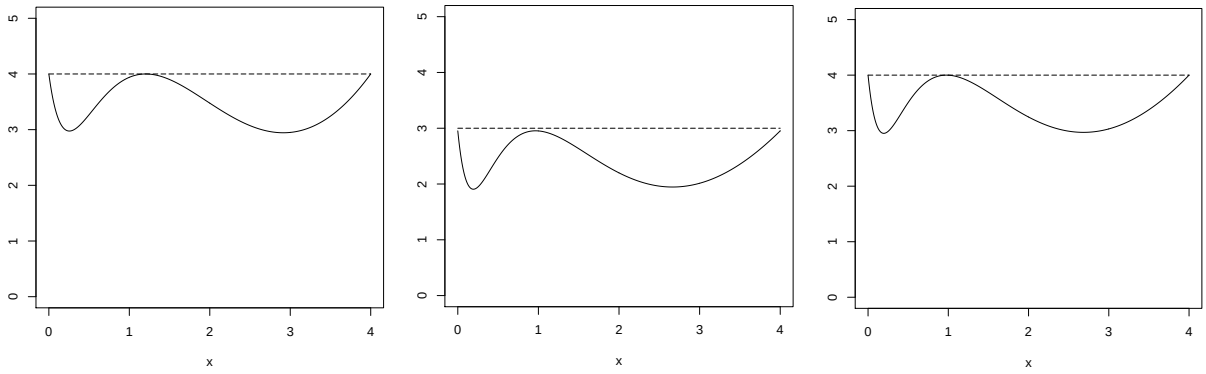


Figure 3: *The necessary condition of optimality for optimal* 3*-point designs in the EMAX model. Left: Bayesian D-optimality design with respect to a uniform prior. Middle: Bayesian optimal design with respect to the Jeffreys prior and the Berger-Bernardo-prior. Right: Bayesian optimal design with respect to a functional uniform prior.*

prior has a larger interior support point. The application of Theorem 2.1 - 2.3 is illustrated in Figure 3. We observe that all designs satisfy the necessary condition for optimality.
We conclude this paper with a brief discussion of Bayesian optimal designs for a compartment model, which is used as a model for the concentration of a substrate over time involving absorption and the elimination of a substrate. Here the mean is given by

$$\mu(x, \boldsymbol{\theta}) = \theta_0(\exp(-\theta_1 x) - \exp(-\theta_2 x)),$$

where θ_1 is the elimination constant and θ_2 the absorption constant. The corresponding optimal designs are displayed in Table 5, where the design space is given by $\mathcal{X} = [0, 20]$ and $\theta_0 > 0, \theta_1 \in [0.05, 0.07]$, and $\theta_2 \in [3.3, 5.3]$ [see Atkinson et al. (1993)]. As before θ_0 and θ_3 are each from a compact interval. All designs presented in this table satisfy the necessary condition of optimality (the corresponding plots are not displayed for the sake of brevity). Interestingly all designs exhibit a very similar structure.

(2.6) with (2.7)		(2.11)/(2.17)		(2.6) with (2.8)	
$\xi(x_i)$	x_i	$\xi(x_i)$	x_i	$\xi(x_i)$	x_i
0.333	0.2286	0.333	0.2321	0.333	0.2343
0.333	1.4106	0.333	1.4310	0.333	1.4420
0.333	18.1145	0.333	18.3185	0.333	18.3132

Table 5: *Bayesian optimal* 3*-point designs with respect to non-informative priors for the compartment model. Left column: Bayesian D-optimal designs with respect to the uniform prior. Middle column: Bayesian optimal designs with respect to the Jeffreys and the Bernardo-Berger prior. Right column: Bayesian D-optimal designs with respect to the functional uniform prior.*

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, Teilprojekt C2) of the German Research Foundation (DFG).

References

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49(2):325–337.

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics*, volume 4, pages 35–60.

Bornkamp, B. (2012). Functional uniform priors for nonlinear modeling. *Biometrics*, 68(3):893–901.

Braess, D. and Dette, H. (2007). On the number of support points of maximin and Bayesian D-optimal designs in nonlinear regression models. *Annals of Statistics*, 35:772–792.

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.

Chang, F.-C. (2005). D-optimal designs for weighted polynomial regression - a functional-algebraic approach. *Statistica Sinica*, 15:153–163.

Chang, F.-C., Chang, H.-C., and Wang, S.-S. (2009). d-optimal designs for polynomial regression with exponential weight function. *Metrika*, 70(3):339–354.

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

Clarke, B. and Wasserman, L. (1993). Noninformative priors and nuisance parameters. *Journal of the American Statistical Association*, 88(424):1427–1432.

Dette, H. (1994). Discrimination designs for polynomial regression on a compact interval. *Annals of Statistics*, 22:890–904.

Dette, H. (1995). Optimal designs for identifying the degree of a polynomial regression. *Annals of Statistics*, 23:1248–1267.

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., 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. and Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Annals of Statistics*, 39(2):1266–1281.

Dette, H., Melas, V. B., and Shpilev, P. (2013). Robust T-optimal discriminating designs. *Annals of Statistics*, 41:1693–1715.

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 Studden, W. J. (1997). *Canonical Moments with Applications in Statistics, Probability and Analysis*. Wiley and Sons, New York.

Dette, H. and Wong, W. K. (1998). Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models. *Biometrika*, 85(4):869–882.

Fang, X. and Hedayat, A. S. (2008). Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics*, 36:428–444.

Firth, D. and Hinde, J. P. (1997a). On Bayesian d-optimum design criteria and the equivalence theorem in non-linear models. *Journal of the Royal Statistical Society, Ser. B*, 59(4):793–797.

Firth, D. and Hinde, J. P. (1997b). Parameter neutral optimum design for non-linear models. *Journal of the Royal Statistical Society, Ser. B*, 59(4):799–811.

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors (with discussion). In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 4, pages 195,210.

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.

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.

He, Z., Studden, W. J., and Sun, D. (1996). Optimal designs for rational models. *Annals of Statistics*, 24:2128–2142.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A*, 186:453–461.

Jones, B., Wang, J., Jarvis, P., and Byrom, W. (1999). Design of cross-over trials for pharmacokinetic studies. *Journal of Statistical Planning and Inference*, 78(1–2):307 – 316.

Karlin, S. and Studden, W. J. (1966). *Tchebysheff Systems: With Application in Analysis and Statistics*. Wiley, New York.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1369.

Khuri, A., Mukherjee, B., Sinha, B., and Ghosh, M. (2006). Design issues for generalized linear

models. *Statistical Science*, 21(3):376–399.

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

Lau, T.-S. and Studden, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *Annals of Statistics*, 13:383–394.

Lau, T. S. and Studden, W. J. (1988). On an extremal problem of Fejère. *Journal of Approximation Theory*, 53:184–194.

Polson, N. G. (1992). On the expected amount of information from a nonlinear model. *Journal of the Royal Statistical Society, Ser. B*, 54:889–895.

Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. *Mathematical Biosciences*, 75:103–120.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker, New York.

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

Spruill, M. C. (1990). Good designs for testing the degree of a polynomial mean. *Sankhya, Ser. B*, 52(1):67–74.

Studden, W. J. (1980). d_s-optimal designs for polynomial regression using continued fractions. *Annals of Statistics*, 8(5):1132–1141.

Studden, W. J. (1982a). Optimal designs for weighted polynomial regression using canonical moments. Statistical decision theory and related topics III, Proc. 3rd Purdue Symp., West Lafayette/Indiana 1981, Vol. 2, 335-350 (1982).

Studden, W. J. (1982b). Some robust-type D-optimal designs in polynomial regression. *Journal of the American Statistical Association*, 77(380):916–921.

Yang, M. (2010). On the de la Garza Phenomenon. *Annals of Statistics*, 38(4):2499–2524.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Annals of Statistics*, 37:518–541.

Zen, M.-M. and Tsai, M.-H. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation in Fourier regression models. *Journal of Statistical Planning and Inference*, 124:475–487.