On the number of support points of maximin and Bayesian D-optimal designs in nonlinear regression models

Dietrich Braess
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: dietrich.braess@rub.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 3214 559

December 20, 2004

Abstract

We consider maximin and Bayesian D-optimal designs for nonlinear regression models. The maximin criterion requires the specification of a region for the nonlinear parameters in the model, while the Bayesian optimality criterion assumes that a prior distribution for these parameters is available. It was observed empirically by many authors that an increase of uncertainty in the prior information (i.e. a larger range for the parameter space in the maximin criterion or a larger variance of the prior distribution in the Bayesian criterion) yields a larger number of support points of the corresponding optimal designs. In this paper we present a rigorous proof of this phenomenon and show that in many nonlinear regression models the number of support points of Bayesian- and maximin D-optimal designs can become arbitrarily large if less prior information is available. Our results also explain why maximin D-optimal designs are usually supported at more different points than Bayesian D-optimal designs.

AMS Subject Classification: 62K05

Keywords and Phrases: Bayesian optimal design, maximin optimal design, nonlinear regression

1 Optimal designs for nonlinear regression models

Consider the common nonlinear regression model

$$E[Y \mid x] = \eta(x, \theta), \qquad (1.1)$$

where $\theta \in \mathbb{R}^m$ is the unknown parameter, x denotes the explanatory variable, which varies in a compact space, say $\mathcal{X}$, and η is a known regression function. We assume that observations at different experimental conditions are independent with constant variance, say $\sigma^2 > 0$. Under the additional assumption of normally distributed errors the Fisher information matrix for the parameter θ at the point x is given by

$$(1.2)\qquad I(x,\theta)=\frac{1}{\sigma^2}\Big(\frac{\partial\eta}{\partial\theta}(x,\theta)\Big)\Big(\frac{\partial\eta}{\partial\theta}(x,\theta)\Big)^T\in\mathbb{R}^{m\times m}.$$

Here we assume that for each $x\in\mathcal{X}$ the regression function is continuously differentiable with respect to θ.

An approximate design ξ for this model is a probability measure on the design space $\mathcal{X}$ with finite support $x_1,\ldots,x_n$ and weights $w_1,\ldots,w_n$ representing the relative proportions of total observations taken at the corresponding design points [see e.g. Kiefer (1974)]. The information matrix of a design ξ is defined by

$$(1.3)\qquad M(\xi,\theta)=\frac{1}{\sigma^2}\int_{\mathcal{X}}I(x,\theta)d\xi(x),$$

and its inverse divided by the sample size is an approximation of the covariance matrix of the least squares estimate for the parameter θ. A local optimal design maximizes an appropriate function of the information matrix [see Silvey (1980) or Pukelsheim (1993)]. There are numerous optimality criteria which can be used to discriminate among competing designs, and we restrict ourselves to the famous D-optimality criterion

$$(1.4)\qquad \log|M(\xi,\theta)|\,.$$

A local D-optimal design, i.e. a design maximizing the function (1.4) for fixed θ, minimizes the volume of the confidence ellipsoid for the unknown parameter $\theta\in\mathbb{R}^m$ [see Pukelsheim (1993)] and it does not depend on the variance σ^2. Therefore, without loss of generality we assume that $\sigma^2=1$ in the following discussion. In general, a local D-optimal design depends on the unknown parameter θ. As a consequence local optimality criteria have been criticized by numerous authors because the resulting optimal designs can be highly inefficient within the true model setting if the unknown parameters are misspecified. A more robust approach to this problem is, in some sense, to quantify the uncertainty in those parameters and to incorporate this additional information into the formulation of suitable optimality criteria. This has been achieved in practice by the introduction of the concepts of Bayesian and maximin optimality.

Mathematically preliminary knowledge of the experimenter can be modeled as follows. Assume that $\theta\in\Theta$ where $\Theta\subset\mathbb{R}^m$ denotes a set with induced Borel field, and let π denote a prior distribution on Θ. A design is called *Bayesian D-optimal* (with respect to the prior π) if it maximizes the function

$$(1.5)\qquad \int_\Theta\log|M(\xi,\theta)|\pi(d\theta).$$

Here and throughout this paper we assume that the corresponding integrals exist. Bayesian D-optimal designs have been studied by several authors [see e.g. Chaloner and Larntz (1989), Pronzato and Walter (1985), Mukhopadhyay and Haines (1995), Dette and Neugebauer (1996, 1997) among many others].

In some circumstances it may be difficult for the experimenter to specify a prior distribution on the parameter space Θ. Therefore many authors propose *standardized maximin D-optimal designs* for the construction of efficient and robust designs, i.e. designs which maximize

$$(1.6)\qquad \min\Big\{\frac{|M(\xi,\theta)|}{|M(\xi[\theta],\theta)|}\mid\theta\in\Theta\Big\}$$

in the class of all approximate designs. Here $\xi[\theta]$ denotes the local D-optimal design maximizing the criterion (1.4) for fixed θ [see e.g. Müller (1995), Dette (1997), Imhof (2001) or Dette and Biedermann (2003)]. The criterion (1.6) does not compare the quantities $|M(\xi, \theta)|$ directly but with respect to the values, which could be obtained if θ and as a consequence the local D-optimal design would be known. The main reason for considering efficiencies instead of the non-standardized quantities is that the values $|M(\xi, \theta)|$ are usually of rather different size and for this reason often not comparable; see Dette (1997).

Bayesian- and standardized maximin D-optimal designs can only be given explicitly in rare circumstances. Usually the optimization of the function (1.5) or (1.6) is performed in the class of all minimally supported designs (i.e. in the class of all designs with m support points) and their optimality within the class of all designs is checked by an application of equivalence theorems [see e.g. Chaloner (1993), Mukhopadhyay and Haines (1995), Dette and Neugebauer (1996, 1997) for some results on Bayesian D-optimal designs and Haines (1995), Imhof (2001), Dette and Biedermann (2003) for results on standardized maximin D-optimal designs]. In these examples minimally supported designs are only optimal with respect to the chosen criterion if the minimum in the maximin criterion is taken over a sufficiently "narrow parameter space" Θ or the prior in Bayesian criterion puts "most of its mass at a small" subset of Θ; cf. the literature cited above. On the other hand, if less information about the unknown parameter is available, it was observed empirically that the number of support points of Bayesian or standardized maximin optimal designs exceeds the number of parameters in the nonlinear regression model. Usually the number increases if less knowledge about θ is incorporated in the optimality criteria [see Chaloner and Larntz (1989), Dette and Biedermann (2003) among others].

In the present paper we will verify that for a broad class of nonlinear regression models the number of support points of the Bayesian- and standardized maximin D-optimal design can become arbitrarily large if less prior information regarding the unknown parameters is available. More precisely, we establish sufficient conditions on the regression models such that an increasing uncertainty about the nonlinear parameters leads to an arbitrary large number of support points of Bayesian and standardized maximin D-optimal designs. We also demonstrate by several examples that these conditions are satisfied in commonly used models; in fact, we did not find any example, where this is not the case. While the idea gives rise to a general tool, specific details and technical difficulties have to be dealt with for the specific models involved. For the sake of a transparent presentation we therefore consider in Section 2 and 3 nonlinear regression models with one or two parameters, where only one parameter enters nonlinearly in the model. In this case the main idea of the construction of optimal designs with a large number of support points becomes most transparent, but it can be easily transferred to models with more parameters as indicated in Section 4 and 5. Finally, some conclusions are given in Section 5, while all technical details are deferred to an appendix.

This paper provides a rigorous proof of some phenomena, which were observed a long time empirically in the literature. Our results yield a better understanding of the structure of optimal designs with respect to Bayesian and maximin optimality criteria and also explain why maximin D-optimal designs are usually supported at more different points than Bayesian D-optimal designs.

2 Standardized maximin D-optimal designs

2.1 The general principle

Throughout the paper we assume that the local D-optimal design depends only on one component of the parameter $\theta \in \mathbb{R}^m$. The general situation can be obtained by a reduction to this case, which is briefly indicated in Section 5. We denote this component by β and the corresponding design by $\xi[\beta]$. Consequently, we reflect only this dependence in our notation (1.2) and (1.3), and the optimality criteria in (1.5) and (1.6) are represented by the functions

$$\Psi(\xi) = \int_{\mathcal{B}} \log |M(\xi, \beta)| \pi(d\beta), \tag{2.1}$$

$$\Phi(\xi) = \min\left\{ \frac{|M(\xi, \beta)|}{|M(\xi[\beta], \beta)|} \mid \beta \in \mathcal{B} \right\}, \tag{2.2}$$

respectively. Here $M(\xi, \beta)$ is the information matrix (1.3) in the nonlinear regression model (with $\sigma^2 = 1$), $\mathcal{B} = [\beta_{\min}, \beta_{\max}]$ represents our prior knowledge about the location of the unknown parameter and π denotes a prior distribution on $\mathcal{B}$. [If the local D-optimal design does not depend on the parameter β, it is also standardized maximin and Bayesian D-optimal, and therefore this case is of no interest in the following discussion.]

Moreover, if the underlying regression model is of size $m \times m$, we assume that the local optimal design $\xi[\beta]$ (depending only on the parameter β by the previous assumption) is minimally supported, i.e. $\#\, supp\, \xi[\beta] = m$. In this case the local D-optimal design has equal masses at its support points [see Silvey (1980)]. This situation is very typical for nonlinear regression models involving only one nonlinear parameter [see e.g. Rasch (1990), He, Studden and Sun (1996), Dette and Neugebauer (1996, 1997)], and several examples will be given below. Let ξ denote a design on $\mathcal{X}$ with masses w_k at support points $x_k \quad (k = 1, \ldots, n)$, then the information matrix of ξ is given by

$$M(\xi, \beta) = \sum_{k=1}^{n} w_k I(x_k, \beta).$$

Throughout this paper,

$$Q(\beta, \tilde{\beta}) = \frac{|M(\xi[\tilde{\beta}], \beta)|}{|M(\xi[\beta], \beta)|} \tag{2.3}$$

quantifies the loss of information if β is the "true" unknown parameter, but the experimenter uses the local D-optimal design for a (wrong) guess of the unknown parameter, say $\tilde{\beta}$.

We will derive sufficient conditions such that the number of support points of the optimal designs with respect to the optimality criteria (2.1) and (2.2) exceeds any given number if the amount of prior information in the optimality criterion is decreased. In order to cover a very broad class of regression models, we write these assumptions in a very general form, but we will immediately add a more transparent one that is sufficient for many commonly used models. Our first Definition quantifies the loss of efficiency caused by an application of a local D-optimal design based on a misspecified parameter.

Definition 2.1 Let $\ell : \mathcal{B} \to \mathbb{R}$ be a continuously differentiable function with $\ell'(\beta) > 0$ for all $\beta \in \mathcal{B}$. The function Q defined in (2.3) is called *uniformly decreasing with respect to* ℓ if the following two conditions hold

(i) For all $\beta, \tilde{\beta} \in \mathcal{B}$ the inequality

$$Q(\beta, \tilde{\beta}) \leq \varphi(\ell(\beta) - \ell(\tilde{\beta})) \tag{2.4}$$

is satisfied, where φ is a real-valued function whose decay (i.e. $\varphi(z) \to 0$) for $z \to \infty$ will be sufficiently fast as specified later for each case under consideration.

(ii) There is a positive constant $\lambda > 0$ such that

$$Q(\beta, \tilde{\beta}) \geq \frac{1}{2} \quad \text{whenever } \ |\ell(\beta) - \ell(\tilde{\beta})| \leq \lambda. \tag{2.5}$$

In many models of actual interest, the transformation ℓ is either the identity or the logarithm. In the latter case we avoid formulas with the transformation. We say that Q is *uniformly decreasing on a logarithmic scale* if

$$Q(\beta, \tilde{\beta}) \leq \psi(\frac{\beta}{\tilde{\beta}}) = \varphi(\log \frac{\beta}{\tilde{\beta}}), \tag{2.6}$$

for some function ψ, where the functions ψ and φ in both definitions are related by $\psi(e^z) = \varphi(z)$. In this context, condition (ii) is rewritten as

$$Q(\beta, \tilde{\beta}) \geq \frac{1}{2} \quad \text{whenever } e^{-\lambda} \leq \frac{\beta}{\tilde{\beta}} \leq e^{\lambda}. \tag{2.7}$$

We now discuss the one- and two-dimensional case separately.

2.2 The case $m = 1$

The following result shows for regression models with one parameter that the number of support points of the standardized maximin D-optimal design can become arbitrarily large under general assumptions. We will demonstrate below that many commonly used models have this property.

In order to derive results of this type a final assumption is required. This property guarantees that points in the design space $\mathcal{X}$ which are not in the support of any local D-optimal design $\xi[\beta]$ can be disregarded for the construction of the standardized maximin D-optimal design; cf. the discussion of the first example. In the case $m = 1$ it reads

$$\text{for any } x \in \mathcal{X} \text{ there exists a local } D\text{-optimal design } \xi[\tilde{\beta}],\ \tilde{\beta} \in \mathcal{B}, \ \text{ such that} \\ |I(x, \beta)| \leq |M(\xi[\tilde{\beta}], \beta)| \quad \forall\, \beta \in \mathcal{B}. \tag{2.8}$$

Theorem 2.2 *Let $m = 1$. Assume that Q is uniformly decreasing with respect to ℓ in the sense of Definition 2.1, where*

$$\varphi(z) \leq c_1 |z|^{-\gamma} \quad \text{with } c_1 > 0,\ \gamma > 1, \tag{2.9}$$

and that (2.8) is satisfied. Assume that $N \in \mathbb{N}$ is given. If $\ell(\beta_{\max}) - \ell(\beta_{\min})$ is sufficiently large, then the standardized maximin D-optimal design with respect to the interval $\mathcal{B} = [\beta_{\min}, \beta_{\max}]$ is supported at more than N points.

Corollary 2.3 *Let $m = 1$. Assume that Q is uniformly decreasing on a logarithmic scale, where*

$$\psi(z) \leq c_1 |\log z|^{-\gamma} \quad \text{holds with } c_1 > 0,\ \gamma > 1, \tag{2.10}$$

and that (2.8), (2.7) are satisfied. Assume that $N \in \mathbb{N}$ is given. If $\beta_{\max}/\beta_{\min}$ is sufficiently large, then the standardized maximin D-optimal design with respect to the interval $\mathcal{B} = [\beta_{\min}, \beta_{\max}]$ is supported at more than N points.

Example 2.4 Consider the one-dimensional exponential growth model

$$\eta(x, \beta) = e^{-\beta x}, \quad \beta \in [1, B],\ x \in [0, 1] \tag{2.11}$$

with Fisher information of the parameter β

$$I(x, \beta) = x^2 e^{-2\beta x}. \tag{2.12}$$

The local D-optimal design is a one-point design supported at the point $x[\beta] = 1/\beta$, and from

$$M(\xi[\beta], \beta) = I(x[\beta], \beta) = (e\beta)^{-2} \tag{2.13}$$

it is easy to see that the function Q in (2.3) is given by

$$Q(\beta, \tilde{\beta}) = \left(\frac{\beta}{\tilde{\beta}} e^{1-\beta/\tilde{\beta}} \right)^2.$$

Thus, we have $Q(\beta, \tilde{\beta}) = \psi(\frac{\beta}{\tilde{\beta}})$, where

$$\psi(z) = (z e^{1-z})^2 \leq \begin{cases} e^2 z^2 & \text{if } z \leq 1, \\ 3\, z^{-2} & \text{if } z > 1. \end{cases}$$

Thus (2.6) holds and $\psi(z) \leq e^2 e^{-2|\log z|}$ decays much faster than required in (2.10). Moreover $\psi(z) \geq \frac{1}{2}$ if $\frac{1}{2} \leq z \leq 2$, which proves property (ii) in Definition 2.1 with $\ell(\beta) = \log \beta$. Finally, we verify property (2.8). Consider a point, say x_0. If $1 \geq x_0 \geq 1/B$, we have $\delta_{x_0} = \xi[1/x_0]$ for the Dirac measure at the point x_0, and there is in fact equality in (2.8) with $\tilde{\beta} = 1/x_0$. On the other hand if $x_0 < 1/B$, then the Dirac measure δ_{x_0} is not local D-optimal for any $\beta \in [1, B]$ and $x_0 \beta < \frac{\beta}{B} \leq 1$ for all $\beta \in [1, B]$. Since the function $z \mapsto z^2 e^{-2z}$ is increasing on the interval $[0, 1]$, it follows that

$$\beta^2 I(x_0, \beta) = \beta^2 M(\delta_{x_0}, \beta) \leq \beta^2 M(\delta_{1/B}, \beta) = \beta^2 M(\xi[B], \beta) \quad \text{for all } \beta \in [1, B],$$

which shows (2.8). Similarly, if $x_0 > 1$ we obtain from the fact that the function $z \to z^2 e^{-2\beta z}$ is decreasing for $z \geq 1$ (note that $\beta \geq 1$)

$$I(x_0, \beta) = x_0^2 e^{-2\beta x_0} \leq e^{-2\beta} = I(1, \beta) = M(\xi[1], \beta)$$

for all $\beta \in [1, B]$. Therefore the assumptions of Corollary 2.3 are satisfied, and the number of support points of the standardized maximin D-optimal design for the regression model (2.11) becomes arbitrarily large with increasing parameter $B \to \infty$.

B	standardized maximin D-optimal design					
7	0.185	0.906				
	0.567	0.433				
10	0.142	0.771				
	0.553	0.447				
20	0.066	0.298	0.919			
	0.441	0.259	0.300			
30	0.047	0.233	0.836			
	0.426	0.266	0.309			
40	0.037	0.193	0.772			
	0.414	0.272	0.314			
50	0.028	0.131	0.374	0.972		
	0.379	0.221	0.170	0.230		
70	0.020	0.099	0.309	0.913		
	0.363	0.220	0.181	0.237		
100	0.014	0.064	0.156	0.287	0.838	
	0.336	0.193	0.093	0.137	0.241	
200	0.007	0.034	0.101	0.250	0.326	0.856
	0.306	0.182	0.147	0.089	0.066	0.210

Table 2.1: *Standardized maximin D-optimal designs for the exponential regression model (2.11) on the interval $[0,1]$ with respect to various parameter spaces $[1,B]$. First row: support points; second row: weights.*

In Table 2.1 we show some numerical results illustrating this fact. We have calculated standardized maximin D-optimal designs for the regression model (2.11) using Matlab for various parameter spaces $\mathcal{B} = [1,B]$. The optimality of the calculated designs was checked by the the equivalence theorem of Wong (1992). By this result a design ξ^* is standardized maximin D-optimal for the one-dimensional exponential growth model (2.11) if and only if there exists a distribution π^* such that the inequality

$$\int_{\mathcal{B}} \frac{x^2 e^{-2\beta x}}{M(\xi[\beta],\beta)} \pi^*(d\beta) \leq 1 \tag{2.14}$$

holds for all $x \in [0,1]$. The distribution π^* is called least favorable distribution or duality measure and depicted in Table 2.2 for the cases considered in Table 2.1. We observe that the number of support points of the standardized maximin D-optimal design increases with the length of the interval $[1,B]$.

Example 2.5 As a further example we consider a simplification of the Bleasdale-Nelder model for describing the dependence of plant yield Y on plant density x [see Ratkowsky (1983), p. 60], that is

$$E[Y \mid x] = (1+\beta x)^{-1}, \tag{2.15}$$

where the explanatory variable varies again in the interval $[0,1]$ and $\beta \in [1, \beta_{\max}]$, i.e. $B = \beta_{\max}$.

B	least favourable distribution					
7	1 0.504	2.9 0.105	7 0.391			
10	1 0.454	3.5 0.227	10 0.319			
20	1 0.396	4.1 0.265	20 0.339			
30	1 0.355	3.8 0.209	9.3 0.167	30 0.269		
40	1 0.337	3.9 0.214	12.0 0.215	40 0.234		
50	1 0.330	4.3 0.227	14.5 0.216	50 0.227		
70	1 0.310	3.9 0.154	7.7 0.131	21 0.192	70 0.213	
100	1 0.289	4 0.183	11.0 0.152	29.8 0.172	100 0.204	
200	1 0.261	4.1 0.159	10.2 0.120	24 0.126	59.3 0.141	200 0.193

Table 2.2: *Least favourable distributions π^* in the equivalence theorem (2.14) corresponding to the standardized maximin D-optimal designs in Table 2.1. First row: support points, second row: weights*

The Fisher information of the parameter β at the point x is given by

$$I(x, \beta) = \frac{x^2}{(1+\beta x)^4} .$$

The local D-optimal design is again a one-point design concentrating its mass at the point $x[\beta] = 1/\beta$ with $M(\xi[\beta], \beta) = (\beta 2^2)^{-2}$. A straightforward calculation shows that

$$Q(\beta, \tilde{\beta}) = \psi\Big(\frac{\beta}{\tilde{\beta}}\Big), \quad \text{where} \quad \psi(z) = \frac{2^4 z^2}{(1+z)^4} .$$

Therefore, Q is uniformly decreasing in the logarithmic scale, where the function ψ satisfies (2.10), and $\psi(z) \geq 1/2$ if $1/3 \leq z \leq 3$. Finally, the remaining property (2.8) can be shown exactly in the same way as in Example 2.4. It follows that the number of support points of the standardized maximin D-optimal design for the regression model (2.15) becomes arbitrarily large with increasing $B = \beta_{\max}$.

Remark 2.6 On a first glance the results of Theorem 2.2, Corollary 2.3 and the examples above are surprising because it was never observed in numerical studies that the number of support points of the standardized maximin D-optimal design exceeds the number of parameters substantially; to our knowledge the numerical results of Example 2.4 are the first ones in this direction. On the other hand, it follows from the proof of Theorem 2.2 in the Appendix that the construction of a design with more than N support points outperforming a given design requires a large parameter space in the maximin D-optimality criterion. Thus in practice standardized maximin D-optimal designs in a one-parametric nonlinear regression model with a large number of support points will only be observed if a large parameter space is involved [cf. Example 2.4].

Example 2.7 Our last one-dimensional example

$$E[Y \mid x] = \frac{1}{1+e^{x+\beta}}, \quad x \in [0,B],\ \beta \in [-B,0], \tag{2.16}$$

arises from the logistic regression model and does not refer to the logarithmic scale. The Fisher information of the parameter β at the point x is given by

$$I(x,\beta) = \frac{e^{x+\beta}}{(1+e^{x+\beta})^2}\,.$$

The local D-optimal design is a one-point design concentrating its mass at the point $x[\beta] = -\beta$ with $M(\xi[\beta],\beta) = 1/4$. Hence,

$$Q(\beta,\tilde\beta) = \frac{4\,e^{\beta-\tilde\beta}}{(1+e^{\beta-\tilde\beta})^2} = \varphi(\beta-\tilde\beta),$$

where the function φ is defined by

$$\varphi(z) = \frac{4\,e^{z}}{(1+e^{z})^2} \le 4e^{-|z|}.$$

Moreover, $Q(\beta,\tilde\beta) \ge \frac{1}{2}$ if $|\beta-\tilde\beta| \le 1$. Therefore we observe a uniform decay on the natural scale [that is $\ell(\beta)=\beta$ in Definition 2.1], the remaining condition (2.8) can be checked by similar arguments as given in the previous examples and Theorem 2.2 applies. If B is sufficiently large the number of support points of the standardized maximin D-optimal design for the regression model (2.16) exceeds any given bound $N \in \mathbb{N}$.

2.3 Two-dimensional cases

In this section we deal with models with two parameters $\theta = (\alpha,\beta)^T$, where only one of them, say β, appears in the local D-optimal design. As a consequence, the Fisher information is of dimension 2×2 and can be represented as

$$I(x,\beta) = f(x,\beta)f^T(x,\beta), \tag{2.17}$$

where $f(x,\beta) = (f_1(x,\beta), f_2(x,\beta))^T \in \mathbb{R}^2$ denotes the gradient of the response function with respect to the parameter θ. Again, the second parameter α is assumed to be 1 without loss of generality since it has no influence on the solution of the D-optimal design problem. Let ξ denote a design with masses w_k at the points x_k $(k=1,\dots,n)$. By the Cauchy-Binet formula the determinant of the information matrix of the design ξ can be represented as

$$|M(\xi,\beta)| = \sum_{i<j} w_i w_j \, I_2(x_i,x_j,\beta), \tag{2.18}$$

where

$$I_2(x_i,x_j,\beta) = \begin{vmatrix} f_1(x_i,\beta) & f_1(x_j,\beta) \\ f_2(x_i,\beta) & f_2(x_j,\beta) \end{vmatrix}^2 . \tag{2.19}$$

Note that the function I_2 is symmetric, i.e. $I_2(x_1,x_2,\beta) = I_2(x_2,x_1,\beta)$. We will show in the following that the properties of the function I_2 determine the behaviour of the standardized maximin D-optimal design.

For a motivation of the following definition assume for a moment that $\mathcal{X} \subset \mathbb{R}$. We call the function I_2 *increasing* (with respect to the second argument) if the inequality

$$I_2(x_1,x_2,\beta) < I_2(x_1,y_2,\beta) \tag{2.20}$$

holds for all $x_1 < x_2 < y_2$ in the experimental domain $\mathcal{X}$. In a nonlinear regression model with increasing function I_2 on the design space $\mathcal{X} = [x_{\min}, x_{\max}]$ the right boundary point, say $\bar{x} = x_{\max}$, is always a support point of the local D-optimal design. Specifically, the inequality

$$I_2(x_1,x_2,\beta) \le I_2(x_1,\bar{x},\beta) \le I_2(x_1,\bar{x},\beta) + I_2(x_2,\bar{x},\beta) \tag{2.21}$$

holds for all $x_1 < x_2$. Similarly, the function I_2 is called *decreasing* (with respect to the first argument) if the opposite inequality is satisfied in the first argument. In this case $\bar{x} = x_{\min}$ belongs always to the support of the local D-optimal design and an analogous inequality holds. In fact, there are regression models like in Example 2.12 below which do not satisfy (2.20) or (2.21), but nevertheless there exists a common support point of all local D-optimal designs and a generalization of the inequality (2.21) is found.

Definition 2.8 A nonlinear regression model with $m = 2$ parameters is called *reducible* if there exists a common support point, say $\bar{x}$, of all local D-optimal designs and a positive constant c such that

$$I_2(x_1,x_2,\beta) \le c[I_2(\bar{x},x_1,\beta) + I_2(\bar{x},x_2,\beta)] \tag{2.22}$$

holds for all x_1, $x_2 \in \mathcal{X}$.

Note that Definition 2.8 does not require the design space $\mathcal{X}$ to be one-dimensional (although we used this assumption for its motivation). Finally, the two dimensional analogue of assumption (2.8) is given by

$$\begin{aligned} &\text{for any } x \in \mathcal{X} \text{ there exists a local } D\text{-optimal design } \xi[\tilde{\beta}], \tilde{\beta} \in \mathcal{B}, \text{ such that} \\ &\qquad |I_2(x,\bar{x},\beta)| \le 4|M(\xi[\tilde{\beta}],\beta)| \quad \forall \beta \in \mathcal{B}. \end{aligned} \tag{2.23}$$

Theorem 2.9 *Let $m = 2$. Assume that the nonlinear regression model is reducible, that (2.23) is satisfied and that Q is uniformly decreasing with respect to ℓ in the sense of Definition 2.1, where the function φ satisfies (2.9). Assume that $N \in \mathbb{N}$ is given. If $\ell(\beta_{\max}) - \ell(\beta_{\min})$ is sufficiently large, then the standardized maximin D-optimal design with respect to the interval $\mathcal{B} = [\beta_{\min}, \beta_{\max}]$ is supported at more than N points.*

The specialization to the logarithmic scale is clear from Corollary 2.3 and therefore omitted. The crucial point in the proof of Theorem 2.9 is that in a reducible nonlinear regression model the double sums (2.18) for the evaluation of the determinant of the information matrix can be estimated by a simple sum, that is

$$
\begin{aligned}
|M(\xi,\beta)| &= \sum_{i<j} w_i w_j \, I_2(x_i,x_j,\beta) = \frac{1}{2}\sum_{i,j} w_i w_j \, I_2(x_i,x_j,\beta) \\
&\le \frac{c}{2}\sum_{i,j} w_i w_j [I_2(\bar{x},x_i,\beta) + I_2(\bar{x},x_j,\beta)] \\
&= c\sum_k w_k I_2(\bar{x},x_k,\beta). \qquad (2.24)
\end{aligned}
$$

Thus the proof of Theorem 2.9 will be obtained by some modifications of the proof of Theorem 2.2. The details are given in the Appendix.

Example 2.10 Consider the Michaelis-Menten model

$$E(Y|x) = \frac{\alpha x}{\beta + x}\,, \; x \in [0,1]\,, \; \beta > 0, \qquad (2.25)$$

which is widely used to describe numerous physical and biological phenomena [see e. g. Cressie and Keightley (1979) or Cornish-Browden (1979)]. Here the Fisher information for the parameter $\theta = (\alpha, \beta)$ is given by

$$I(x,\beta) = \frac{x^2}{(\beta+x)^2}\begin{pmatrix} 1 & -\frac{\alpha}{\beta+x} \\ -\frac{\alpha}{\beta+x} & \frac{\alpha^2}{(\beta+x)^2} \end{pmatrix}. \qquad (2.26)$$

The quotients $|M(\xi,\beta)|/|M(\xi[\beta],\beta)|$ and consequently the local D-optimal design do not depend on the parameter α. Therefore we assume without loss of generality $\alpha = 1$, which justifies the use of the notation $I(x,\beta)$ in (2.26). A straightforward calculation yields

$$I_2(x_1,x_2,\beta) = \frac{x_1^2}{(x_1+\beta)^4}\frac{x_2^2}{(x_2+\beta)^4}(x_1-x_2)^2, \qquad (2.27)$$

Obviously I_2 is increasing with respect to its second argument, and Theorem 2.9 is applicable. The local D-optimal designs are well known to have equal masses at the points

$$x_1 = x[\beta] = \frac{\beta}{2\beta+1}, \quad x_2 = \bar{x} = 1 \qquad (2.28)$$

[see e.g. Rasch (1990)], and

$$|M(\xi[\beta],\beta)| = \frac{1}{64\beta^2(\beta+1)^6} \qquad (2.29)$$

[see Dette and Biedermann (2003)]. We now consider the standardized maximin D-optimal design problem for the parameter space $\mathcal{B} = [\beta_{\min}, 1] = [B^{-1}, 1]$ and derive an estimate for the function $Q(\beta, \tilde{\beta})$. Since $\bar{x} = 1$, we can restrict ourselves to

$$I_2(x,1,\beta) = \frac{x^2(1-x)^2}{(x+\beta)^4(1+\beta)^4}\,. \qquad (2.30)$$

A straightforward calculation yields

$$
\begin{aligned}
I_2(x[\tilde\beta],1,\beta) &= \Big[4(\beta+1)\frac{\beta\,\frac{\tilde\beta}{2\tilde\beta+1}}{(\frac{\tilde\beta}{2\tilde\beta+1}+\beta)^2}(1-\frac{\tilde\beta}{2\tilde\beta+1})\Big]^2 4\,|M(\xi[\beta],\beta)| \\
&= \Big[2\Big\{(\beta+1)(\tilde\beta+1)\frac{(\beta+\tilde\beta)^2}{(\beta+\tilde\beta+2\beta\tilde\beta)^2}\Big\}\frac{2\beta/\tilde\beta}{(1+\beta/\tilde\beta)^2}\Big]^2 4\,|M(\xi[\beta],\beta)|. \qquad (2.31)
\end{aligned}
$$

The expression within the braces does not exceed $1+\max\{\beta,\tilde\beta\}$. Hence,

$$
I_2(x[\tilde\beta],1,\beta) \;\le\; 16\Big[\frac{4\beta/\tilde\beta}{(1+\beta/\tilde\beta)^2}\Big]^2|M(\xi[\beta],\beta)| = 4\psi\Big(\frac{\beta}{\tilde\beta}\Big)\,|M(\xi[\beta],\beta)|,
$$

where the function ψ is defined by

$$
\psi(z) := \Big[\frac{8z}{(1+z)^2}\Big]^2 . \qquad (2.32)
$$

We have for any design with support in the interval $[B^{-1}(2B^{-1}+1),1/3]$

$$
|M(\xi,\beta)| = \sum_{i<j} w_i w_j I_2(x_i,x_j,\beta) \le \sum_{i=1}^{n} w_i I_2(x_i,1,\beta).
$$

This implies for the design $\xi[\tilde\beta]$

$$
Q(\beta,\tilde\beta) = \frac{|M(\xi[\tilde\beta],\beta)|}{|M(\xi[\beta],\beta)|} \le 4\psi\Big(\frac{\beta}{\tilde\beta}\Big),
$$

and consequently Q decreases in the logarithmic scale, where the function

$$
\psi(z) \le \begin{cases} 4z^2 & \text{if} \quad z\le 1 \\ 4z^{-2} & \text{if} \quad z>1 \end{cases}
$$

decays faster than required in condition (2.10). For estimating Q from below we note that the expression within the braces of (2.31) is not smaller than $4/5$ if $2/3 \le \frac{\beta}{\tilde\beta} \le 3/2$, and it follows that

$$
Q(\beta,\tilde\beta) \ge \frac{1}{2} \quad \text{whenever} \quad \frac{2}{3} \le \frac{\beta}{\tilde\beta} \le \frac{3}{2}.
$$

Assumption (2.23) is obviously satisfied, if $x \in [B^{-1}/(2B^{-1}+1),1/3]$. In this case x belongs to the support of the local D-optimal design $\xi[\tilde\beta]$ with $\tilde\beta = x/(1-2x)$. If $x \ge \frac{1}{3}$ or $x \le B^{-1}/(2B^{-1}+1)$, the inequality (2.23) is satisfied for the local D-optimal design $\xi[1]$ and $\xi[B^{-1}]$, respectively. Now by Theorem 2.9 the number of support points of the standardized maximin D-optimal design for the Michaelis-Menten model becomes arbitrarily large, if the minimum in the optimality criterion (2.2) is taken over the range $\mathcal{B} = [B^{-1},1]$ with $B\to\infty$.

Example 2.11 Consider the exponential regression model

$$E(Y|x) = \alpha e^{-\beta x}, \ x \in [0,1], \ \beta \in [1, \beta_{\max}], \tag{2.33}$$

which has applications in pharmacokinetics [see e.g. Landaw and DiStefano (1984)]. By the same argument as in the previous example we may assume $\alpha = 1$ and obtain for the information matrix of the parameter $\theta = (\alpha, \beta)$

$$I(x,\beta) = \begin{pmatrix} e^{-2\beta x} & -xe^{-2\beta x} \\ -xe^{-2\beta x} & x^2 e^{-2\beta x} \end{pmatrix} = e^{-2\beta x} \begin{pmatrix} 1 & -x \\ -x & x^2 \end{pmatrix}.$$

The local D-optimal design can be found in Dette and Neugebauer (1997) and has equal masses at the points

$$x_1 = \bar{x} = 0, \quad x_2 = \frac{1}{\beta},$$

with corresponding determinant

$$|M(\xi[\beta], \beta)| = \frac{1}{4\,(e\beta)^2}\,.$$

The function I_2 is given by

$$I_2(x_1, x_2, \beta) = (x_1 - x_2)^2 e^{-2\beta(x_1+x_2)},$$

which is obviously decreasing with respect to its first argument. Because of

$$I_2(0, 1/\tilde{\beta}, \beta) = \tilde{\beta}^{-2} e^{-2\beta/\tilde{\beta}}$$

we obtain $Q(\beta, \tilde{\beta}) = 4\psi(\beta/\tilde{\beta})$, where ψ is the same function as in Examples 2.4 and 2.10. The assumptions of Theorem 2.9 can be established by the same arguments as given in Example 2.4. Consequently, the number of support points of the standardized maximin D-optimal design in the exponential growth model (2.33) becomes arbitrarily large if the parameter space $\mathcal{B} = [1, \beta_{\max}]$ is increased (that is $\beta_{\max} \to \infty$).

Example 2.12 Consider the exponential growth model

$$E(Y|x) = \alpha + e^{-\beta x}, \ x \in [0,1], \ \beta \in [1, \beta_{\max}], \tag{2.34}$$

which is used for analyzing the growth of crops [see Liebig (1988) or Krug and Liebig (1988)]. In this model the Fisher information for the parameter $\theta = (\alpha, \beta)$ is given by

$$I(x,\beta) = \begin{pmatrix} 1 & -xe^{-\beta x} \\ -xe^{-\beta x} & x^2 e^{-2\beta x} \end{pmatrix},$$

and the function I_2 defined in (2.19) is obtained as

$$I_2(x_1, x_2, \beta) = (x_1 e^{-\beta x_1} - x_2 e^{-\beta x_2})^2.$$

This function is not monotone with respect to the first or second argument. However, it is easy to see that $I_2(x_1, x_2, \beta)$ can always be increased by using $x_1 = 0$ and it follows from Dette and Neugebauer (1997) that the local D-optimal design puts equal masses at the points

$$x_1 = \bar{x} = 0, \quad x_2 = \frac{1}{\beta}$$

with corresponding determinant

$$|M(\xi[\beta],\beta)| = \frac{1}{4\,(e\beta)^2}$$

[see also Han and Chaloner (2003)]. From $I_2(0,1/\tilde{\beta},\beta) = \tilde{\beta}^{-2}e^{-2\beta/\tilde{\beta}}$ we obtain $Q(\beta,\tilde{\beta}) = \psi(\beta/\tilde{\beta})$, where the function $\psi(z) = z^2 e^{2(1-z)}$ was introduced in Example 2.4. This shows that (2.6) and (2.10) are fullfilled. Obviously,

$$I_2(x_1,x_2,\beta) \le (x_1 e^{-x_1})^2 + (x_2 e^{-x_2})^2 \le I_2(0,x_1,\beta) + I_2(0,x_2,\beta),$$

i.e., the model is reducible. We conclude as in the previous examples that the last assumption (2.23) of Theorem 2.9 is also satisfied. Therefore, the number of support points of the standardized maximin D-optimal design in the exponential growth model (2.34) becomes arbitrarily large if the parameter space $\mathcal{B} = [1,\beta_{\max}]$ is increased ($\beta_{\max} \to \infty$).

3 Bayesian D-optimal designs

In the present section we consider similar problems for the Bayesian D-optimality criterion defined in (2.1). When Bayesian D-optimal designs are considered, it does not make a difference whether the information matrix or its standardized analogue is considered. The difference between the criterion

$$\Psi_{st}(\xi) = \int_{\mathcal{B}} \log \frac{|M(\xi,\beta)|}{|M(\xi[\beta],\beta)|}\,\pi(d\beta), \tag{3.1}$$

and the function defined in (2.1) is a constant that does not depend on the design ξ. We begin our investigations with the one-dimensional case corresponding to the situation considered in Section 2.2.

Theorem 3.1 *Let $m = 1$, assume that (2.8) holds and that Q is uniformly decreasing with respect to ℓ in the sense of Definition 2.1, where the function φ satisfies*

$$\varphi(z) \le c_1 e^{-|z|^\gamma} \tag{3.2}$$

for some positive constants c_1, γ, and the prior distribution and the function ℓ in Definition 2.1 satisfy for all Borel set $B \subset \mathcal{B} = [\beta_{\min},\beta_{\max}]$

$$\int_B \frac{c_3}{\ell(\mathcal{B})}\ell(d\beta) \le \int_B \pi(d\beta) \tag{3.3}$$

for some positive constant c_3. Assume that $N \in \mathbb{N}$ is given. If $\ell(\beta_{\max}) - \ell(\beta_{\min})$ is sufficiently large, then the Bayesian D-optimal design with respect to the prior π on the interval $\mathcal{B}$ is supported at more than N points.

Note that increasing the interval $\mathcal{B}$ in the optimality criterion (2.1) such that condition (3.3) is satisfied also changes the prior π on $\mathcal{B}$. A typical example is the uniform distribution on the set $\mathcal{B}$, which obviously changes with $\mathcal{B}$. For this prior the assumption (3.3) is obviously satisfied if

$\ell(\beta) = \beta$ or $\ell(\beta) = \log \beta$. It can easily be shown that the functions φ or ψ, resp., in Examples 2.4, 2.5 and 2.7 also satisfy the stronger assumptions in Theorem 3.1. As a consequence it follows that the number of support points of Bayesian D-optimal designs in these examples is also unbounded if the support of the prior distribution is increased such that (3.3) holds. This includes the important case of the non-informative uniform prior in the Bayesian optimality criterion.

Example 3.2 Consider the exponential regression model (2.11) of Example 2.4. It follows by the preceding discussion that the number of support points of the Bayesian D-optimal design with respect to a uniform prior on the interval $[1, B]$ becomes arbitrarily large with increasing $B \to \infty$. In Table 3.1 we show the Bayesian D-optimal designs corresponding to the situation considered in Table 2.1. We observe that the standardized maximin D-optimal designs have remarkably more support points than the Bayesian D-optimal designs with respect to the uniform prior. The standardized maximin D-optimal design for the parameter space $\mathcal{B} = [1, 30]$ has three support points [see the third row in Table 2.1], while the corresponding Bayesian D-optimal design has only one support point. The other cases are similar, where in the extreme case $B = 200$ the standardized maximin D-optimal design has 6 support points while the Bayesian D-optimal design has only two points concentrating most of it mass at the smaller support point [see the last rows in Table 2.1 and 3.1].

B	Bayesian D-optimal design	
7	0.250 1.000	
10	0.182 1.000	
20	0.095 1.000	
30	0.065 1.000	
40	0.048 0.981	0.354 0.019
50	0.038 0.973	0.318 0.027
70	0.027 0.966	0.266 0.034
100	0.019 0.962	0.215 0.038
200	0.010 0.959	0.134 0.041

Table 3.1. *Bayesian D-optimal designs with respect to a uniform distribution on the interval* $[1, B]$ *in the exponential regression model (2.11). First row: support points; second row: weights.*

Now we turn to the two-dimensional models.

Theorem 3.3 *Let $m = 2$, assume that the nonlinear regression model is reducible, that (2.23) holds and that the function Q is uniformly decreasing with respect to ℓ in the sense of Definition 2.1, where the function φ satisfies*

$$\varphi(z) \leq c_1 e^{-|z|^\gamma} \quad \text{with } c_1,\ \gamma > 0, \tag{3.4}$$

and the prior π and the function ℓ in Definition 2.1 satisfy (3.3). Assume that $N \in \mathbb{N}$ is given. If $\ell(\beta_{\max}) - \ell(\beta_{\min})$ is sufficiently large, then the Bayesian D-optimal design with respect to the prior π on the interval $\mathcal{B} = [\beta_{\min}, \beta_{\max}]$ is supported at more than N points.

Remark 3.4. Note that Theorem 3.1 and 3.3 require a stronger decay of the function φ than Theorem 2.2 and 2.9. As a consequence it follows from the proofs of these results in the Appendix that the number of support points of Bayesian D-optimal designs usually increases more slowly with the length of the parameter space compared to the maximin case. We have illustrated this fact in Example 3.2, where we compare the standardized maximin and the Bayesian D-optimal design with respect to the uniform distribtion in model (2.11).

When we analyzed the two-dimensional examples in Section 2.3, we established already the faster decay of Q that is required for the Bayesian setting. Consequently, in all models considered in Section 2.3 the Bayesian D-optimal designs with respect to priors satisfying (3.3) are supported at an arbitrarily large number of support points if the difference $\ell(\beta_{\max}) - \ell(\beta_{\min})$ is sufficiently large. This includes the important case of a uniform prior in the Bayesian optimality criterion.

4 A remark on models with more than two parameters

Although our main results are stated for models with one or two parameters (i.e. $m = 1, 2$), the arguments given in the preceding sections can be extended to models with $m \geq 3$ parameters. However, the technical difficulties increase substantially. The approach presented in Section 2 and 3 is a general one, but specific details and technical difficulties have to be dealt with for the specific models involved. To indicate how this can be done, we consider as an additional example the exponential growth model with 3 parameters

$$E(Y|x) = \alpha_1 + \alpha_2 e^{-\beta x},\ x \in [0, 1],\ \beta \in [1, \beta_{\max}]. \tag{4.1}$$

Let ξ denote a design with masses w_k at the points x_k ($k = 1, \ldots, n$); then by the Cauchy Binet formula the determinant of the information matrix of the design ξ can be represented as

$$|M(\xi, \beta)| \;=\; \sum_{i<j<\ell} w_i w_j w_\ell\, I_3(x_i, x_j, x_\ell, \beta), \tag{4.2}$$

where

$$I_3(x_i, x_j, x_\ell, \beta), \;=\; \begin{vmatrix} f_1(x_i,\theta) & f_1(x_j,\theta) & f_1(x_\ell,\theta) \\ f_2(x_i,\theta) & f_2(x_j,\theta) & f_2(x_\ell,\theta) \\ f_3(x_i,\theta) & f_3(x_j,\theta) & f_3(x_\ell,\theta) \end{vmatrix}^2 \tag{4.3}$$

is the analogue of the function I_2 introduced in (2.19) and $(f_1(x,\theta), f_2(x,\theta), f_3(x,\theta))^T$ denotes the gradient of the response function with respect to the parameter $\theta = (\alpha_1, \alpha_2, \beta)^T$. In particular, we obtain for the model (4.1)

$$\begin{aligned} I_3(x_1,x_2,x_3,\beta) &= [x_1 e^{-\beta x_1}(e^{-\beta x_3} - e^{-\beta x_2}) + x_2 e^{-\beta x_2}(e^{-\beta x_1} - e^{-\beta x_3}) + x_3 e^{-\beta x_3}(e^{-\beta x_2} - e^{-\beta x_1})]^2 \\ &= H^2(x_1,x_2,x_3), \end{aligned}$$

where the last line defines the function $H(x_1,x_2,x_3)$. Han and Chaloner (2003) showed that local D-optimal designs for the exponential regression model (4.1) have three support points and that $\bar{x} = 0$ and $\hat{x} = 1$ are in fact support points. An alternative derivation of the lastnamed fact will be given below. For the points $\bar{x} = 0$ and $\hat{x} = 1$ this expression reduces to

$$I_3(0,x,1,\beta) = [xe^{-\beta x}(1-e^{-\beta}) - e^{-\beta}(1-e^{-\beta x})]^2 = H^2(0,x,1,\beta). \tag{4.4}$$

Note that sums of cyclic products as $a(b-c) + b(c-a) + c(a-b)$ vanish. Therefore we obtain the very useful representation

$$\begin{aligned} H(x_1,x_2,x_3,\beta) &= \frac{e^{-\beta x_3} - e^{-\beta x_2}}{1-e^{-\beta}} \left[x_1 e^{-\beta x_1}(1-e^{-\beta}) - e^{-\beta}(1-e^{-\beta x_1})\right] \\ &+ \frac{e^{-\beta x_1} - e^{-\beta x_3}}{1-e^{-\beta}} \left[x_2 e^{-\beta x_2}(1-e^{-\beta}) - e^{-\beta}(1-e^{-\beta x_2})\right] \\ &+ \frac{e^{-\beta x_2} - e^{-\beta x_1}}{1-e^{-\beta}} \left[x_3 e^{-\beta x_3}(1-e^{-\beta}) - e^{-\beta}(1-e^{-\beta x_3})\right] \\ &= \sum_{k=1}^{3} a_k H(0,x_k,1,\beta) \end{aligned} \tag{4.5}$$

with coefficients a_k satisfying $|a_k| \le 1$. If the support points are ordered, i.e.

$$0 \le x_1 < x_2 < x_3 \le 1, \tag{4.6}$$

then a_1, $a_3 < 0$ and $a_2 > 0$. The estimation of the quantity I_3 heavily depends on the knowledge of the signs of the function H. Obviously, $H(0,x,1,\beta)$ vanishes at $x = 0$ and $x = 1$. Moreover, the derivative at $x = 0$ is positive. Since exponential sums of the form $a_1 + (a_2 + a_3 x)e^{-\beta x}$ have at most two real zeros [see Karlin and Studden (1996)], it follows that $H(0,x,1,\beta) > 0$ for $0 < x < 1$. For the same reason, the function $x \mapsto H(x,x_2,x_3,\beta)$ vanishes only at $x = x_2$ and $x = x_3$. Hence, $\operatorname{sign} H(x_1,x_2,x_3,\beta) = \operatorname{sign} H(0,x_2,x_3,\beta)$ if the ordering (4.6) holds. Similarly, it follows that $\operatorname{sign} H(0,x_2,x_3,\beta) = \operatorname{sign} H(0,x_2,1,\beta) = +1$. Recalling the statement on the coefficients a_1, a_2 and a_3 in (4.5) we see that the summands with $k = 1$ and $k = 3$ diminish the sum, and it follows that

$$I_3(x_1,x_2,x_3,\beta) \le I_3(0,x_2,1,\beta).$$

This does not only yield the cited result regarding the location of the smallest and largest support point of local D-optimal designs, but additionally provides the bound

$$I_3(x_1,x_2,x_3,\beta) \le \sum_{k=1}^{3} I_3(0,x_k,1,\beta)$$

which holds for support points x_1, x_2, x_3 in any order. Therefore the exponential regression model (4.1) is reducible in an obvious way.

The model can now be treated exactly in the same way as the reducible two-dimensional models although the technical details are more involved. For large β we get from (4.4)

$$I_3(0,x,1,\beta) \le \frac{2}{3}x^2 e^{-2\beta x}, \qquad |M(\xi[\beta],\beta)| = \frac{1}{3^3}\sup_x I_3(0,x,1,\beta) \ge \frac{1}{27}\frac{1}{3e^2\beta^2}\,.$$

Although there is no simple representation for $|M(\xi[\beta],\beta)|$, the estimates above are comparable to the corresponding equations (2.12) and (2.13) for the one-dimensional exponential model. The estimates differ only by constants. Thus by some changes of the arguments we also conclude that the number of support points of the standardized maximin D-optimal design in the model (4.1) is unbounded for sufficiently large $\beta_{\max}$.

5 Conclusions

A common tool for the construction of efficient designs in nonlinear regression models are Bayesian or maximin criteria. Both optimality criteria require prior information regarding the parameters which enter in the model nonlinearly. It was observed numerically by many authors that the number of support points of Bayesian and maximin D-optimal designs is increasing with the amount of uncertainty about the location of the nonlinear parameters. In this paper we have established sufficient conditions for the nonlinear regression models under which the number of support points of Bayesian and maximin D-optimal designs can become arbitrarily large if the prior information regarding the unknown nonlinear parameters in the optimality criterion is diminished. These conditions apply to many of the commonly used regression models (in fact we did not find any model, where these conditions were not satisfied).

For the sake of brevity we restricted our investigations to one- and two parametric regression models, where at most one parameter appears nonlinearly in the model. However, our approach is a general one and can also be applied to regression models with more nonlinear parameters, where some of technicalities have to be adapted to the specific model under consideration. For example, consider a nonlinear regression model with two parameters, say $\theta = (\theta_1, \theta_2)$, such that the local D-optimal design depends on both components of θ. Assume that the minimum in the optimality criterion (1.6) is taken over a rectangular parameter space, say $\Theta = [a,b] \times [c,d]$. If one of these intervals degenerates to a point and the length of the other interval is increased (with respect to an appropriate scaling function ℓ), the results of Section 2 show that the number of the support points of the standardized maximin D-optimal design becomes arbitrarily large provided that the sufficient conditions in Theorem 2.9 are satisfied. A similar result is also available for the Bayesian D-optimality criterion by combining this argument with the results of Section 3.

Similarly, models with more than two parameters can be treated, but the technical difficulties increase substantially. We have indicated in Section 4, how this can be done for a model with three parameters, where one parameter appears nonlinearly in the regression model. The extension to models with a larger number of parameters follows essentially the arguments presented in this paper with an additional amount of notation and specific details have to be dealt with for the specific model under consideration. We did not present results in this direction here because the technical details usually depend on the specific model and become too difficult to be presented in a concise paper [see our example in Section 4].

Nevertheless, our results make a general statement on the structure of optimal designs with respect to the standardized maximin and Bayesian D-optimality criterion, which is important for a better understanding of these sophisticated optimality criteria. In all examples that we have investigated we were able to prove that the number of support points of the standardized maximin and Bayesian D-optimal designs exceeds any given bound if the knowledge about the underlying parameter space, which is incorporated in the optimality criteria, is diminished. This gives a rigorous proof of a phenomenon which was conjectured in many nonlinear regression models for a long time in the literature.

A Proofs

Proof of Theorem 2.2. The proof consists of two steps. Set $B = \ell(\beta_{\max}) - \ell(\beta_{\min})$. At first we show that for an arbitrary design, say ξ_N, with N support points it follows that

$$\Phi(\xi_N) = \min\Big\{ \frac{|M(\xi_N,\beta)|}{|M(\xi[\beta],\beta)|} \mid \beta \in [\beta_{\min},\beta_{\max}] \Big\} \leq d_1(N+1)B^{-\gamma}, \tag{A.1}$$

where d_1 is a positive constant not depending on B. Secondly, we show that there exists a design ξ_n on $\mathcal{X}$ such that

$$\Phi(\xi_n) \geq \frac{d_2}{B} \tag{A.2}$$

for some positive constant d_2 not depending on B. Since $\gamma > 1$, given N, we have

$$d_1(N+1)B^{-\gamma} < \frac{d_2}{B}$$

if B is sufficiently large, and the optimal design is supported at more than N points in this case.

To verify the estimate (A.1) let $\xi_N = \sum_{k=1}^N w_k \delta_{x_k}$ denote any design with mass w_k at the point x_k $(k = 1,\dots,N)$. Here δ_{x_k} denotes the Dirac measure at the point x_k. Then

$$M(\xi_N,\beta) = \sum_{k=1}^N w_k M(\delta_{x_k},\beta)$$

(here we use the condition $m = 1$). By assumption (2.8) there exist real numbers $\beta_{\min} \leq \beta_1 < \ldots < \beta_N \leq \beta_{\max}$ such that the inequality

$$M(\xi_N,\beta) \leq \sum_{k=1}^N w_k M(\xi[\beta_k],\beta) = M(\xi[\beta],\beta) \sum_{k=1}^N w_k\, Q(\beta,\beta_k) \tag{A.3}$$

holds for all $\beta \in \mathcal{B}$. For convenience, we put $\beta_0 = \beta_{\min}$, $\beta_{N+1} = \beta_{\max}$. Now at least one gap between the numbers $\ell(\beta_k)$ must be large. Specifically, there exists an index $j \in \{0,\dots,N\}$ such that

$$\ell(\beta_{j+1}) - \ell(\beta_j) \geq \frac{\ell(\beta_{N+1}) - \ell(\beta_0)}{N+1} = \frac{B}{N+1}. \tag{A.4}$$

We consider the inequality at the point $\bar\beta$ defined by $\ell(\bar\beta)=\frac{1}{2}[\ell(\beta_j)+\ell(\beta_{j+1})]$ and derive from (A.4)

$$(\text{A.5})\qquad |\ell(\bar\beta)-\ell(\beta_k)|\ge\frac{1}{2}(\ell(\beta_{j+1})-\ell(\beta_j))\ge\frac{B}{2(N+1)}\qquad\forall\, k\in\{0,1,2,\dots,N+1\}.$$

We now use the inequality (A.3), the definition of Q in (2.3) and obtain from assumption (2.4), (2.9) and (A.5)

$$\begin{aligned}M(\xi_N,\bar\beta) &\le \sum_{k=1}^N w_k Q(\bar\beta,\beta_k)M(\xi[\bar\beta],\bar\beta)\le\sum_{k=1}^N w_k\,\varphi(\ell(\bar\beta)-\ell(\beta_k))M(\xi[\bar\beta],\bar\beta)\\ &\le c_1\left(\frac{B}{2(N+1)}\right)^{-\gamma}M(\xi[\bar\beta],\bar\beta)=\frac{c_1(2N+2)^\gamma}{B^\gamma}M(\xi[\bar\beta],\bar\beta)\end{aligned}$$

for some positive constant c_1. We define the constant $d_1=c_1(2N+2)^\gamma$, and the proof of the upper bound (A.1) is complete.

For a proof of the lower bound (A.2) we choose $n=\lfloor\frac{1}{2}B/\lambda\rfloor$, where λ is the constant defined in Definition 2.1 (ii), and set β_k such that

$$(\text{A.6})\qquad \ell(\beta_k)=\ell(\beta_{\min})+(2k-1)\lambda,\ (k=1,\dots,n).$$

Note that these points are contained in the interval $[\beta_{\min},\beta_{\max}]$. Let δ_{x_k} denote the local D-optimal design for the parameter β_k with corresponding support point x_k, $k=1,\dots,n$, (note that these are one-point designs by assumption) and define $\xi_n=\sum_{k=1}^n\frac{1}{n}\delta_{x_k}$ as the uniform distribution on the points $x_1,\dots,x_n$, then

$$(\text{A.7})\qquad M(\xi_n,\beta)=\frac{1}{n}\sum_{k=1}^n I(x_k,\beta)=\frac{1}{n}\sum_{k=1}^n M(\xi[\beta_k],\beta).$$

Obviously, given $\beta\in[\beta_{\min},\beta_{\max}]$, there exists an index $j=j_\beta$ such that

$$|\ell(\beta)-\ell(\beta_j)|\le\lambda\,.$$

By construction $M(\xi[\beta_j],\beta)=Q(\beta,\beta_j)M(\xi[\beta],\beta)\ge\frac{1}{2}M(\xi[\beta],\beta)$. Since all terms in the sum (A.7) are nonnegative, it follows that for all $\beta\in\mathcal{B}$

$$M(\xi_n,\beta)\ \ge\ \frac{1}{n}M(\xi[\beta_j],\beta)\ge\frac{1}{2n}M(\xi[\beta],\beta)\ge\frac{\lambda}{B}M(\xi[\beta],\beta).$$

Recalling the definition of the standardized maximin criterion in (2.2) we conclude that

$$\Phi(\xi_n)\ge\lambda/B.$$

With the choice $d_2=\lambda$ the proof of the lower bound (A.2) is complete. □

Proof of Theorem 2.9. Let ξ_N denote a design with masses w_k at the points x_k $(k=1,\dots,N)$ and let $\xi[\beta_k]$ denote the design corresponding to the point x_k by the inequality (2.23). Assumption (2.22) admits the reduction (2.24), and we obtain

$$\begin{aligned}|M(\xi_N,\beta)| &\le c\sum_k w_k I_2(\bar x,x_k,\beta)\le 4c\sum_k w_k|M(\xi[\beta_k],\beta)|\\ &= 4c\,|M(\xi[\beta],\beta)|\sum_k w_kQ(\beta,\beta_k),\qquad(\text{A.8})\end{aligned}$$

where we used the definition of Q in (2.3). Note that this inequality corresponds to (A.3) in the proof of Theorem 2.2. By the same arguments as in the proof of Theorem 2.2 it follows that

$$\Phi(\xi_N) \leq d_1 B^{-\gamma}$$

with $d_1 = 4c\, c_1(2N+2)^\gamma$ and $\gamma > 1$.

In order to prove the corresponding lower bound we adapt the construction in the proof of Theorem 2.2 to the two-dimensional setting. Let $n = \lfloor \frac{1}{2} B/\lambda \rfloor$ again and define β_k by (A.6). Denote $x_k = x[\beta_k]$ as the corresponding non-trivial design point of the local D-optimal design $\xi[\beta_k]$, and set

$$\xi_n = \frac{1}{2}\delta_{\bar{x}} + \frac{1}{2n}\sum_{k=1}^{n} \delta_{x_k} = \sum_{k=0}^{n} w_k \delta_{x_k},$$

where the last identity defines the weights w_k $(k = 0, \ldots, n)$, and we put $x_0 = \bar{x}$. Note that all terms in the sum (2.18) are nonnegative. We omitt the terms in which no argument is $\bar{x}$, to obtain

$$\begin{aligned} |M(\xi_n, \beta)| &= \frac{1}{2}\sum_{i,k=0}^{n} w_i w_k I_2(x_i, x_k, \beta) \\ &\geq \sum_{k=1}^{n} \frac{1}{2}\frac{1}{2n} I_2(x_k, \bar{x}, \beta) = \frac{1}{n}\sum_{k=1}^{n} |M(\xi[\beta_k], \beta)| \\ &= \frac{1}{n}|M(\xi[\beta], \beta)| \sum_{k=1}^{n} Q(\beta, \beta_k). \end{aligned}$$

The same argument as presented in the first part of the proof of Theorem 2.2 shows the lower bound

$$\Phi(\xi_n) \geq \frac{\lambda}{B},$$

and the assertion of Theorem 2.9 follows by the same arguments as given in the first part of the proof of Theorem 2.2. $\square$

Proof of Theorem 3.1. For convenience, in a first step we assume that the given transformation ℓ is the identity and define $B = \ell(\mathcal{B}) = \beta_{\max} - \beta_{\min}$. For a given design ξ_N with N support points, we know that inequality (A.3) holds. With assumption (3.3) and the notation $\beta_0 = \beta_{\min}$, $\beta_{N+1} = \beta_{\max}$ and $\Delta_j = \frac{1}{2}(\beta_{j+1} - \beta_j)$ we estimate the contribution of the interval $[\beta_j, \beta_j + \Delta_j]$ to the Bayesian D-optimality criterion as follows

$$\begin{aligned} \int_{\beta_j}^{\beta_j+\Delta_j} \log \frac{|M(\xi_N, \beta)|}{|M(\xi[\beta], \beta)|}\, \pi(d\beta), &\leq \frac{c_3}{B}\int_{\beta_j}^{\beta_j+\Delta_j} \log \sum_{k=1}^{N} w_k \frac{|M(\xi[\beta_k], t)|}{|M(\xi[t], t)|}\, dt \\ &\leq \frac{c_3}{B}\int_{\beta_j}^{\beta_j+\Delta_j} \log \sum_{k=1}^{N} w_k \varphi(t - \beta_k)\, dt \\ &\leq \frac{c_3}{B}\int_{\beta_j}^{\beta_j+\Delta_j} \log\Big(c_1 \exp(-|t - \beta_j|^\gamma)\Big)\, dt \\ &= \frac{c_3}{B}\int_{0}^{\Delta_j} [\log c_1 - z^\gamma]\, dz \\ &= \frac{c_3}{B}\left[\Delta_j \log c_1 - \frac{1}{1+\gamma}\Delta_j^{1+\gamma}\right]. \end{aligned}$$

The same bound is derived for the interval $[\beta_j + \Delta_j, \beta_{j+1}]$. Summing over all intervals of this form we conclude that

$$\Psi_{st}(\xi_N) \le \frac{2c_3}{B} \sum_{j=0}^{N} [\Delta_j \log c_1 - \frac{1}{1+\gamma} \Delta_j^{1+\gamma}].$$

Since $\sum_{j=0}^{N} \Delta_j = \frac{B}{2}$ and the function $z^{1+\gamma}$ is strictly convex, the right hand side attains its maximum if all Δ_j's are equal and we obtain the upper bound

$$\Psi_{st}(\xi_N) \le c_4 \left[\log c_1 - \frac{B^\gamma}{(1+\gamma)} \right] \tag{A.9}$$

for some positive constant c_4. Note that the right hand side of this inequality is dominated by the term with B^γ when $B \to \infty$.

The construction of a better design with respect to the Bayesian optimality criterion follows the arguments given in the proof of Theorem 2.2. Let $\lambda > 0$ be defined by

$$Q(\beta, \tilde{\beta}) \ge \frac{1}{2}, \quad \text{whenever } |\beta - \tilde{\beta}| \le \lambda \, .$$

Set $n = \lfloor \frac{B}{2\lambda} \rfloor$ and $\beta_j = \beta_{\min} + (2j-1)\lambda$ for $k = 1, 2, \ldots, n$. We choose a design ξ with n support points such that

$$M(\xi, \beta) = \frac{1}{n} \sum_{k=1}^{n} M(\xi[\beta_k], \beta)$$

[see the identity in (A.7)]. For any given $\beta \in [\beta_{\min}, \beta_{\max}]$, there exists a β_j with $|\beta - \beta_j| \le \lambda$ satisfying (2.5). Therefore it follows for all $\beta \in \mathcal{B}$ that

$$M(\xi, \beta) \ge \frac{1}{n} M(\xi[\beta_j], \beta) = \frac{1}{n} Q(\beta, \beta_j) M(\xi[\beta], \beta) \ge \frac{1}{2n} M(\xi[\beta], \beta) \, .$$

Hence,

$$\Psi_{st}(\xi_n) \ge \int_{\mathcal{B}} \log \frac{1}{2n} \pi(d\beta) = \log \frac{1}{2n} \ge -\log B + \log \lambda.$$

This value is larger than the upper bound (A.9) if B is sufficiently large. Therefore, a design with N support points cannot be optimal if B is sufficiently large.

We have restricted ourselves to the case $\ell(\beta) = \beta$ for the sake of simplicity. The general case proceeds exactly in the same way. For instance, we have to choose $dt = \ell(d\beta)/\ell(\mathcal{B}) = \ell'(\beta) d\beta / \ell(\mathcal{B})$ in the first integral of the proof, and the boundaries of the intervals have to be adapted. The details are left to the reader. □

Proof of Theorem 3.3. The proof proceeds as in the proof of Theorem 2.9. In essence, (A.8) reduces the two-dimensional problem to the one-dimensional expression and the arguments for Theorem 3.1 can be used. □

Acknowledgements. The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen") is gratefully acknowledged. Parts of the paper were written during the second author's visit at the Institute of Statistics in Louvain-la-Neuve in September 2004 and this author would like to thank the institute for its hospitality. The authors are also grateful to Isolde Gottschlich who typed parts of this paper with considerable technical expertise and to A. Pepelyshev for computational assistance.

References

K. Chaloner (1993). A note on optimal Bayesian design for nonlinear models. J. Statist. Plann. Inf. 37, 229–235.

K. Chaloner, K. Larntz (1989). Optimal Bayesian designs applied to logistic regression experiments. J. Statist. Plan. Inference 21, 191–208.

A. Cornish-Browden (1979). *Fundamentals of Enzyme Kinetics*. Butterworth, London.

N.A.C. Cressie, D.D. Keightley (1979). The underlying structure of a direct linear plot with applications to the analysis of hormone-receptor interactions. Journal of Steroid Biochemistry 11, 1173–1180.

H. Dette (1997). Designing experiments with respect to 'standardized' optimality criteria. J. Roy. Statist. Soc. Ser. B 59, 97–110.

H. Dette and S. Biedermann (2003). Robust and efficient designs for the Michaelis-Menten model. J. Amer. Stat. Assoc., 98, 679–686.

H. Dette, H.-M. Neugebauer (1996). Bayesian optimal one point designs for one parameter nonlinear models. Journal of Statistical Planning and Inference, 52, 1, 17–31.

H. Dette, H.-M. Neugebauer (1997). Bayesian D-optimal designs for exponential regression models. Journal of Statistical Planning and Inference, 60, 2, 331–349.

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

C. Han, K. Chaloner (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.

Z. He, W.J. Studden and D. Sun (1996). Optimal designs for rational models. Ann. Statist. 24, 2128-2142.

L.A. Imhof (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. Ann. Statist. 29, No.2, 561–576.

L.A. Imhof, D. Song, W.K. Wong (2002). Optimal design of experiments with possibly failing trials. Stat. Sinica 12, 1145-1155.

J.C. Kiefer (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849–879.

H. Krug, H.-P. Liebig (1988). Static regression models for planning greenhouse production. Acta Horticulturae 230, 427–433.

E.W. Landaw, J.J. DiStefano 3rd (1984). Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. Am. J. Physiol. 246, 665–677.

H.-P. Liebig (1988). Temperature integration by kohlrabi growth. Acta Horticulturae 230, 371–380.

Ch.H. Müller (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. Journal of Statistical Planning and Inference, 44, 117–132.

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

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

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

D. Rasch (1990). Optimum experimental design in nonlinear regression. Commun. Statist., Theor. Meth. 19, No.12, 4789–4806.

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

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

W.K. Wong (1992). A unified approach to the construction of minimax designs. Biometrika 79, 611–619.