SFB 823

Design admissibility and de la Garza phenomenon in multi-factor experiments

Holger Dette, Xin Liu, Rong-Xian Yue

Nr. 8/2020

Discussion Paper

SFB 823

Design admissibility and de la Garza phenomenon in multi-factor experiments

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
Bochum, Germany

Xin Liu
College of Science
Donghua University
Shanghai 201600, China

Rong-Xian Yue
Department of Mathematics,
Shanghai Normal University
Shanghai 200234, China

March 20, 2020

Abstract

The determination of an optimal design for a given regression problem is an intricate optimization problem, especially for models with multivariate predictors. Design admissibility and invariance are main tools to reduce the complexity of the optimization problem and have been successfully applied for models with univariate predictors. In particular several authors have developed sufficient conditions for the existence of saturated designs in univariate models, where the number of support points of the optimal design equals the number of parameters. These results generalize the celebrated de la Garza phenomenon (de la Garza, 1954) which states that for a polynomial regression model of degree $p-1$ any optimal design can be based on at most p points.

This paper provides - for the first time - extensions of these results for models with a multivariate predictor. In particular we study a geometric characterization of the support points of an optimal design to provide sufficient conditions for the occurrence of the de la Garza phenomenon in models with multivariate predictors and characterize properties of admissible designs in terms of admissibility of designs in *conditional* univariate regression models.

MSC 2010 subject classifications: Primary 62K05; secondary 62J05

Key words and phrases: admissibility, dual problem, conditional model, multi-factor experiment, optimal design

1 Introduction

It is well known that an appropriate choice of an experimental design can improve the quality of statistical analysis substantially, and therefore the problem of constructing optimal designs for regression models has found considerable attention in the literature (see, for example, the monographs of Pukelsheim, 2006; Randall et al., 2007). However, the determination of an optimal design often results in an intricate optimization problem that is difficult to handle, in particular for models used for experiments with multivariate predictors.

A useful strategy is to simplify the problem by identifying subclasses of relatively simple designs, which must contain the optimal design. A prominent example of such a class is the class of admissible designs consisting of the designs with an information matrix, that cannot be improved by an information matrix of another design with respect to the Loewner ordering. In decision theoretic terms the set of admissible designs therefore forms a *complete class*, in the sense that the information matrix of any inadmissible design may be improved by the information matrix of an admissible design. It is well known that optimal designs with respect to the most of the commonly used optimality criteria must be admissible (see Pukelsheim, 2006, Chapter 10.10) and consequently in these cases the determination of optimal designs can be restricted to the class of admissible designs. Along this line, in a series of remarkable papers Yang and Stufken (2009, 2012), Yang (2010), Dette and Melas (2011), Dette and Schorning (2013) and Hu et al. (2015) derived several complete classes of designs for regression models with a univariate predictor. In particular it is demonstrated that the celebrated de la Garza phenomenon (de la Garza, 1954), which states that for a polynomial regression model of degree $p-1$ any optimal design can be based on at most p points, appears in a broad class of regression models with a univariate predictor.

While these methods provide a very powerful tool for the determination of optimal designs, its application is limited to single-factor experiments since the key tools to prove these results are not available for functions of several variables. For example, the characterizations developed in Dette and Melas (2011) and Dette and Schorning (2013) are based the theory of Chebyshev systems (see Karlin and Studden, 1966), which requires regression functions with a univariate argument. Consequently, for regression models with multivariate predictor optimal design problems, including investigations of admissibility, have been mostly treated on a case-by-case analysis using various techniques. For example, Heiligers (1992) investigated admissible experimental designs in a multiple polynomial regression model. Yang et al. (2011) derived a class of admissible designs for the commonly used multi-factor logistic and probit models. Huang et al. (2020) characterized an essentially complete class with respect to Schur ordering for binary response models with multiple nonnegative explanatory variables. Moreover, for several specific models with a multivariate predictor optimal designs with respect to various criteria have been determined. Exemplarily, we mention Graßhoff et al. (2007), who studied locally D-optimal designs for generalized linear models using a canonical transformation, Biedermann et al. (2011), who showed that in additive partially nonlinear models D-optimal designs can be found as the products of the corresponding D-optimal designs in one dimension, Dette and Grigoriev (2014), who studied E-optimal designs for second order response surface models, Grigoriev et al. (2018), who discussed locally D-optimal designs for the Cobb-Douglas model, Kabera et al. (2018), who investigated D-optimal designs for the two-variable binary logistic regression model with interaction, and Castro et al. (2019), who used the moment-sum-of-squares hierarchy of semidefinite programming problems to solve approximate optimal design problems for multivariate polynomial regression on a compact space.

In the present paper we study these problems from a more general point of view. In particular we develop a geometric characterization for the support points of an optimal design which can be used to derive sufficient conditions for the occurrence of the de la Garza phenomenon in regression models with a multivariate predictor. Our strategy is to handle the design problem by considering the dual optimization problem. Moreover, in contrast to the previous literature, which considers characterizations in terms of the explanatory variable, our approach uses the *induced design space* of the regression model under consideration. Moreover, we also provide a necessary condition for a class of designs to be admissible in terms of the admissibility of the designs in the corresponding *conditional models*.

In Section 2 we develop sufficient conditions for the occurrence of the de la Garza phenomenon based on the geometric characterization of the support points of an optimal design. Section 3 introduces the concept of conditional models and designs, which are used to investigate design admissibility for models with multivariate predictors. In Section 4, we illustrate the potential of our approach in three examples considering various nonlinear models with a multivariate predictor. Finally all proofs of our technical results are deferred to Section 5.

2 Optimal designs and a geometric characterization

We begin stating the optimal design problem as considered, for example, in Pukelsheim (2006). Throughout this paper let $\mathrm{Sym}(s)$ denote the set of all real symmetric $s \times s$ matrices, $\mathrm{NND}(s) \subset \mathrm{Sym}(s)$ the set of all nonnegative definite matrices and $\mathrm{PD}(s) \subset \mathrm{NND}(s)$ the set of positive definite matrices. We consider the common linear regression model

$$y = \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\theta} + \varepsilon, \tag{2.1}$$

where $\boldsymbol{x} = (x_1, \cdots, x_q)^\top$ is a q-dimensional vector of predictors which varies in the design space $\mathcal{X} \subset \mathbb{R}^q$, $\boldsymbol{f}(\boldsymbol{x})$ is a k-dimensional vector of known linearly independent regression functions, $\boldsymbol{\theta} \in \mathbb{R}^k$ denotes the vector of unknown parameters, and ε is a random variable with mean 0 and constant variance $\sigma^2 > 0$. We assume that the experimenter can take n independent observations of the form $y_i = \boldsymbol{f}^\top(\boldsymbol{x}_i)\boldsymbol{\theta} + \varepsilon_i$ $(i = 1, \dots, n)$ at experimental conditions $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$.

Following Kiefer (1974) we define a (approximate) design for model (2.1) as a probability measure ξ on the design space $\mathcal{X}$ with finite support and the information matrix of the design ξ in model (2.1) by

$$\boldsymbol{M}(\xi) = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^\top(\boldsymbol{x})\xi(d\boldsymbol{x}). \tag{2.2}$$

If the design ξ has masses $\xi_1, \dots, \xi_m$ at m support points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$, and n observations can be taken, the quantities $\xi_\ell n$ are rounded to non-negative integers, say n_ℓ, such that $\sum_{\ell=1}^m n_\ell = n$ and the experimenter takes n_ℓ observations at each $\boldsymbol{x}_\ell$ $(\ell = 1 \dots, m)$. In this case the covariance matrix of the least squares estimator $\sqrt{n}\hat{\boldsymbol{\theta}}$ for the parameter $\boldsymbol{\theta}$ in model (2.1) converges to the matrix $\sigma^2\boldsymbol{M}^{-1}(\xi)$. which is used to measure the accuracy of the estimator $\hat{\boldsymbol{\theta}}$.

We use the notation Ξ for the set of all approximate designs on the design space $\mathcal{X}$ and $\mathcal{M}(\Xi) = \{\boldsymbol{M}(\xi) \mid \xi \in \Xi\}$ for the set of all information matrices. An optimal design ξ^* maximizes an appropriate function, say ϕ, of the information matrix $M(\xi)$, where ϕ : NND$(s)\to$ $\mathbb{R}$ is a positively homogeneous, super-additive, nonnegative, non-constant and upper semi-continuous function. Throughout this paper we call a function with these properties *optimality criterion* or *information function*. The most prominent optimality criteria are the matrix means defined by

$$\phi_p(\boldsymbol{C}) = \begin{cases} \left(\frac{1}{s}\text{trace}(\boldsymbol{C}^p)\right)^{1/p} & \text{for } p \in (-\infty, 1] \setminus \{0\} \\ (\det(\boldsymbol{C}))^{1/s} & \text{for } p = 0 \\ \lambda_{\min}(\boldsymbol{C}) & \text{for } p = -\infty \end{cases}, \tag{2.3}$$

which include the classical A-, D- and E-optimality criteria as special cases $p = -1$, $p = 0$ and $p = -\infty$, respectively (here we define $\phi_p(\boldsymbol{C}) = 0$ if $\boldsymbol{C} \in$ NND$(s)\setminus$PD(s)).

Given an optimality criterion ϕ on NND(k) the *design problem* then reads as follows

$$\max_{\boldsymbol{M} \in \mathcal{M}(\Xi)} \phi(\boldsymbol{M}), \tag{2.4}$$

where, in a second step, one has to identify a design ξ^* corresponding to a maximizer $\boldsymbol{M}^*$ of (2.4). Any design with this property is called *ϕ-optimal design*. As pointed out in the introduction, an important problem in optimal design theory is to identify sufficient conditions on the regression model (2.1) such that (approximate) optimal designs are saturated, which means that the number of support points of the design coincides with the dimension of the parameter. This property is called de la Garza phenomenon referring to the famous result of de la Garza (1954), which shows that the G-optimal design in a polynomial regression of degree $p-1$ on a compact interval has p support points. While this problem has found considerable attention for models with one-dimensional predictors (see the references mentioned in the introduction), there are - to our best knowledge - no general results available which characterize saturated designs in models with a multivariate predictor.

We begin with a geometric characterization of the support points of a ϕ-optimal design, which can be used to derive sufficient conditions for the occurrence of the de la Garza phenomenon in models with multivariate predictors. For this purpose we define for a matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \in \mathbb{R}^{k\times k}$ a linear transformation $\boldsymbol{h}_{\boldsymbol{Z}} : \mathbb{R}^k \to \mathbb{R}^k$ by

$$\boldsymbol{h}_{\boldsymbol{Z}}(\boldsymbol{x}) := (h_{\boldsymbol{Z}1}(\boldsymbol{x}), \ldots, h_{\boldsymbol{Z}k}(\boldsymbol{x}))^\top := \boldsymbol{Z}^\top \boldsymbol{f}(\boldsymbol{x}) = ({\boldsymbol{z}_1}^\top \boldsymbol{f}(\boldsymbol{x}), \ldots, {\boldsymbol{z}_k}^\top \boldsymbol{f}(\boldsymbol{x}))^\top \tag{2.5}$$

and consider the corresponding point

$$P_{\boldsymbol{Z}}(\boldsymbol{x}) = (h^2_{\boldsymbol{Z}1}(\boldsymbol{x}), \ldots, h^2_{\boldsymbol{Z}k}(\boldsymbol{x}))^\top \in \mathbb{R}^k \ . \tag{2.6}$$

Theorem 2.1. *Let $\xi^* = \{(\boldsymbol{x}_i^*, w_i^*)\}_{i=1}^n$ be a ϕ-optimal design for the regression model (2.1). There exists an orthogonal matrix, say $\boldsymbol{Z}^* = (\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_k^*) \in \mathbb{R}^{k\times k}$ with a linear transformation $\boldsymbol{h}_{\boldsymbol{Z}^*}$ of the form* (2.6)*, such that the vectors $P_{\boldsymbol{Z}^*}(\boldsymbol{x}_1^*), \ldots, P_{\boldsymbol{Z}^*}(\boldsymbol{x}_n^*)$ define at most k different supporting hyperplanes of the k-dimensional polytope*

$$\mathcal{P}_{\boldsymbol{Z}^*} := \big\{\lambda = (\lambda_1, \ldots, \lambda_k)^\top : \lambda_i \geq 0, \ \ \forall i = 1, \ldots, k, \ P^\top_{\boldsymbol{Z}^*}(\boldsymbol{x})\lambda \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\big\}. \tag{2.7}$$

Moreover, if $\boldsymbol{f}(\boldsymbol{x}_i^*)$ and $\boldsymbol{f}(\boldsymbol{x}_j^*)$ are two vectors corresponding to the same supporting hyperplane they have the same length.

Example 2.1. To illustrate the result given by Theorem 2.1, we consider a linear regression in two variables with no intercept, that is $\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2)^\top$, where $\boldsymbol{x} = (x_1, x_2) \in [0, 1]^2$. If

$$\boldsymbol{Z} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

is a 2×2 orthogonal matrix, then the vector $\boldsymbol{h}_{\boldsymbol{Z}}$ in (2.6) is given by

$$\boldsymbol{h}_{\boldsymbol{Z}}(\boldsymbol{x}) = (x_1 \cos t - x_2 \sin t, x_1 \sin t + x_2 \cos t)^\top,$$

and it is easy to see that the polytope

$$\mathcal{P}_{\boldsymbol{Z}} = \big\{(\lambda_1, \lambda_2) : \lambda_1 \geq 0, \lambda_2 \geq 0, (x_1 \cos t - x_2 \sin t)^2 \lambda_1 + (x_1 \sin t + x_2 \cos t)^2 \lambda_2 \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\big\} \tag{2.8}$$

is determined by at most three half planes, which are defined by

$$\begin{aligned}\lambda_1 \cos^2 t + \lambda_2 \sin^2 t &\le 1, \\ \lambda_1 \sin^2 t + \lambda_2 \cos^2 t &\le 1, \\ \lambda_1(1 - \sin 2t) + \lambda_2(1 + \sin 2t) &\le 1,\end{aligned}$$

and correspond to the points $(1, 0), (0, 1)$ and $(1, 1)$, respectively. Therefore, the support points of any ϕ-optimal design are contained in the set $\{(1, 0), (0, 1), (1, 1)\}$, and the corresponding weights can now be found by a straightforward calculation.

In fact ϕ_p–optimal designs were determined by Pukelsheim (2006), Section 8.6, who showed that the ϕ_p-optimal design for $p \in [-\infty, 1)$ is given by

$$\xi_p^* = \left\{ \begin{array}{ccc} (1,1) & (1,0) & (0,1) \\ w(p) & (1-w(p))/2 & (1-w(p))/2 \end{array} \right\},$$

where $w(p) = 1 - 4/(3 + 3^{1/(1-p)})$ if $p > -\infty$, and $w(-\infty) = 0$.

For example, the D-optimal design ξ_D^*, i.e., the ϕ_p-optimal design with $p = 0$, has masses $1/3, 1/3$ and $1/3$ at the points $(1, 1)$, $(1, 0)$ and $(0, 1)$. The information matrix of ξ_D^* is given by

$$\boldsymbol{M}(\xi_D^*) = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

The corresponding polytope is obtained for the choice $t = \pi/4$ and given by

$$\mathcal{P}_{\boldsymbol{Z}^*} = \left\{ (\lambda_1, \lambda_2) : \lambda_1 \ge 0,\ \lambda_2 \ge 0,\ \frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 \le 1,\ 2\lambda_2 \le 1 \right\}. \tag{2.9}$$

The two support points $(1, 0)$ and $(0, 1)$ correspond to the same hyperplane defined by the equation $\frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 = 1$ since the equalities

$$\big(\cos(\pi/4)x_1 - \sin(\pi/4)x_2\big)^2 = \frac{1}{2} \quad \text{and} \quad \big(\sin(\pi/4)x_1 + \cos(\pi/4)x_2\big)^2 = \frac{1}{2}$$

hold for $(x_1, x_2) = (1, 0)$ and $(x_1, x_2) = (0, 1)$. The third support point $(1, 1)$ corresponds to the other hyperplane $2\lambda_1 = 1$ because we have

$$\big(\cos(\pi/4)x_1 - \sin(\pi/4)x_2\big)^2 = 0 \quad \text{and} \quad \big(\sin(\pi/4)x_1 + \cos(\pi/4)x_2\big)^2 = 2$$

for $(x_1, x_2) = (1, 1)$. Similarly, we consider the E-optimal design ξ_E^*, i.e., the ϕ_p-optimal design with $p = -\infty$ has equal masses at the points $(0, 1)$ and $(1, 0)$ corresponding to the same hyperplane $\frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 = 1$ of the polytope (2.9) and the information matrix is given by

$$\boldsymbol{M}(\xi_E^*) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

As a direct application of Theorem 2.1, we obtain sufficient conditions for the occurrence of the de la Garza phenomenon in the linear regression model (2.1), and the following corollary gives a bound of the number of support points of an optimal design.

Theorem 2.2. *Either one of the following conditions is sufficient for the existence of a ϕ-optimal design with k support points in the regression model (2.1).*

(a) There are no different support points of a design corresponding to the same supporting hyperplane of the polytope as defined in (2.7).

(b) There are no vectors of the same length in the induced design space

$$\mathcal{F} = \{\boldsymbol{f}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}.$$

Corollary 2.1. *If there are at most N vectors of the same length in the induced design space $\mathcal{F}$, then there exists a ϕ-optimal design for the regression model (2.1) with at most Nk support points.*

Example 2.2. As a first application of Theorem 2.2, we consider the de la Garza phenomenon for the weighted heteroscedastic polynomial regression model on a compact interval, say $[0, 1]$. Optimal design problems in this model have found considerable attention in the literature (see, for example, Fang, 2002; Chang, 2005; Dette et al., 2005; Chang and Lin, 2006; Chang and Jiang, 2007; Sekido, 2012, among others). To be precise let $\mathcal{X} = [0, 1]$ (or any other compact interval on the non-negative line) and consider the vector of regression functions

$$\boldsymbol{f}(x)\boldsymbol{\theta} = \sqrt{\lambda(x)}(1, x, \dots, x^d)^\top, \tag{2.10}$$

where λ is a positive function on the interval $[0, 1]$, which is called *efficiency function* in the literature. It is well known that the information matrix corresponding to this vector of regression functions is proportional to the information matrix in a heteroscedastic polynomial regression model on the interval $[0, 1]$, that is

$$\mathbb{E}[y(x)] = \theta_0 + \theta_1 x + \cdots + \theta_d x^d \ , \quad \mathrm{Var}(y(x)) = \frac{\sigma^2}{\lambda(x)}$$

(see Fedorov, 1972). If the function $x \to \lambda(x)\|\boldsymbol{f}(x)\|^2 = 1 + x^2 + \cdots + x^{2d}$ is injective on the interval $[0, 1]$, it follows from Theorem 2.2 that there exists a ϕ-optimal design $\xi^* \in \Xi$ supported at at most $d + 1$ points. This situations occurs in particular, if the function λ is increasing, because the function $x \to \|\boldsymbol{f}(x)\|^2 = 1 + x^2 + \cdots + x^{2d}$ is a strictly increasing function on the interval $[0, 1]$. For the special case $\lambda(x) \equiv 1$ we obtain the celebrated de la Garza phenomenon (see de la Garza, 1954).

3 Admissibility

In this section, we study the relation between admissibility of a design ξ in the model (2.1) and the admissibility of a corresponding "conditional design" of ξ in a "conditional model" of (2.1), which will be defined below. Throughout this paper we call a design ξ_1 admissible if there **does not** exist any design ξ_2 such that $\boldsymbol{M}(\xi_1) \neq \boldsymbol{M}(\xi_2)$ and $\boldsymbol{M}(\xi_2) \geq \boldsymbol{M}(\xi_1)$, that is the matrix $\boldsymbol{M}(\xi_2) - \boldsymbol{M}(\xi_1)$ is nonnegative definite. For the sake of simplicity, all results in this section are presented for models with a two-dimensional predictor, but the generalization to the q-dimensional case with $q \geq 3$ is straightforward with some additional notation.

To be precise, consider the linear model (2.1) with a two-dimensional predictor $\boldsymbol{x} = (x_1, x_2)$ and define the function

$$\mu(\boldsymbol{x}) = \mu(x_1, x_2) = \sum_{j=1}^{p} f_j(x_1, x_2)\theta_j = \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\theta}, \tag{3.1}$$

as the expected response at experimental condition $\boldsymbol{x} = (x_1, x_2) \in \mathcal{X}$. Let $t : \mathcal{X} \to \mathbb{R}$ denote a real-valued function on $\mathcal{X}$ with range $\mathcal{T} = \{t(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$. The *conditional model given* t is defined on the design space $\widetilde{\mathcal{X}}(t) = \{\boldsymbol{x} \in \mathcal{X} : t(\boldsymbol{x}) = t\}$ (the preimage of the the set $\{t\}$) and given by

$$\widetilde{\mu}_t(\boldsymbol{x}) = \sum_{j=1}^{p_t} \widetilde{f}_{jt}(\boldsymbol{x})\theta_{jt} = \widetilde{\boldsymbol{f}}_t^\top(\boldsymbol{x})\widetilde{\boldsymbol{\theta}}_t, \quad \boldsymbol{x} \in \widetilde{\mathcal{X}}(t), \tag{3.2}$$

where $\left\{\widetilde{f}_{1t}(\boldsymbol{x}), \widetilde{f}_{2t}(\boldsymbol{x}), \ldots, \widetilde{f}_{p_t,t}(\boldsymbol{x})\right\}$ is a set of linearly independent regression functions which spans $\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_p(\boldsymbol{x})\}$ under the condition that $t(\boldsymbol{x}) = t$, and $\widetilde{\boldsymbol{\theta}}_t$ is a $p_t \times 1$ vector of parameters which may depend on t.

In the following we are particularly interested in two cases corresponding to the projections on the margins. To be precise assume that $\mathcal{X} \subset \mathbb{R}^2$ and define $t_1(\boldsymbol{x}) = x_1$, then for fixed x_1 the set $\widetilde{\mathcal{X}}(x_1)$ can be identified with the set $\mathcal{X}_2 := \{x_2 : (x_1, x_2) \in \mathcal{X}\}$ and we obtain the *conditional model for the second factor* x_2 on $\mathcal{X}_2$. Moreover, if the vector $\widetilde{\boldsymbol{f}}_{x_1}(\boldsymbol{x})$ in the conditional model (3.2) is independent of x_1, we use the notation $\widetilde{\boldsymbol{f}}_2(x_2) := \widetilde{\boldsymbol{f}}_{x_1}(\boldsymbol{x})$ and the resulting model

$$\widetilde{\mu}_2(x_2) = \widetilde{\boldsymbol{f}}_2^\top(x_2)\widetilde{\boldsymbol{\theta}}_2, \quad x_2 \in \mathcal{X}_2, \tag{3.3}$$

is called *the marginal model for the second factor.* One can similarly define the conditional model and the marginal model for the first factor.

Example 3.1. To illustrate these ideas we consider the linear model

$$\mu(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 \tag{3.4}$$

on the design space $\mathcal{X} = [0,1]^2$. Consider the mapping $t(\boldsymbol{x}) = x_1 + x_2$ from the square $[0,1]^2$ onto the interval $[0,2]$. For every $t \in [0,2]$ there are only three independent components among the regression functions $\{1, x_1, x_2, x_1x_2\}$ because of the constraint $x_1 + x_2 = t$. Replacing x_2 with $t - x_1$ the conditional model can be expressed in the form (3.2) with $\widetilde{\boldsymbol{f}}_t^\top(\boldsymbol{x}) = (1, x_1, x_1^2)^\top$, where the conditional design space $\widetilde{\mathcal{X}}(t)$ can be identified with the interval $[0,t]$ if $t \in (0,1]$ and with the interval $\widetilde{\mathcal{X}}(t) = [t-1, 1]$ if $t \in (1,2]$.
Moreover, the marginal model for the i-th factor corresponds to the vector of regression functions $\widetilde{\boldsymbol{f}}_i(x_i) = (1, x_i)^\top$ and the marginal design space is given by $\mathcal{X}_i = [0,1], i = 1, 2$.

For a design ξ on the design space $\mathcal{X}$ we define

$$\xi_t(t) = \int_{\widetilde{\mathcal{X}}(t)} \xi(d\boldsymbol{x})$$

as the marginal design of ξ on the design space $\mathcal{T}$, then, if $\xi_t(t) > 0$, the design ξ induces a conditional design $\xi_{\boldsymbol{x}|t}$ on the design region $\widetilde{\mathcal{X}}(t)$ of the conditional model, which is defined by

$$\xi_{\boldsymbol{x}|t}(\cdot) = \frac{1}{\xi_t(t)} \xi(\boldsymbol{x}) \ .$$

In addition, we define

$$\boldsymbol{M}_t(\xi_{\boldsymbol{x}|t}) = \int_{\widetilde{\mathcal{X}}(t)} \widetilde{\boldsymbol{f}}_t(\boldsymbol{x}) \widetilde{\boldsymbol{f}}_t^\top(\boldsymbol{x}) \xi_{\boldsymbol{x}|t}(d\boldsymbol{x})$$

as the information matrix of the design $\xi_{\boldsymbol{x}|t}$ in the conditional model (3.2) and denote by Ξ_t the set of all approximate designs on the design space $\widetilde{\mathcal{X}}(t)$. The following result is proved in the Appendix.

Theorem 3.1. *A necessary condition for the admissibility of a design $\xi \in \Xi$ in the class Ξ for the regression model* (3.1) *is that the conditional design $\xi_{\boldsymbol{x}|t}$ induced by ξ is admissible in the class Ξ_t in the conditional model (3.2) for every $t \in \mathcal{T}$ with $\xi_t(t) > 0$.*

Furthermore, the following theorem gives a complete subclass and a bound of the number of support points of an optimal design.

Corollary 3.1. *Assume that the design region is of the form $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ for some sets $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}$ and suppose that the marginal models exist for both factors. Define Ξ_i^A as the class of admissible designs for the the i-th marginal model ($i = 1, 2$) and denote by Ξ^C the*

subclass of designs on $\mathcal{X}$, where the i-th marginal design belong to Ξ_i^A $(i = 1, 2)$. Then the class of all admissible designs for the model (3.1) is a subset of Ξ^C.
Moreover, if the admissible designs in Ξ_i^A are based on at most p_i points, $i = 1, 2$, then the designs in Ξ^C are based on at most $p_1 p_2$ points.

4 Some applications

In this section we illustrate several applications of the results in Section 2 and 3 in the determination of locally optimal designs for nonlinear models with a multivariate predictor. To be precise we consider the common nonlinear regression models with q factors

$$\mathbb{E}[y(\boldsymbol{x})] = \eta(\boldsymbol{x}, \boldsymbol{\theta}), \quad \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^q, \tag{4.1}$$

where $y(\boldsymbol{x})$ is a normal distributed random variable with constant variance, say $\sigma^2 > 0$ and observations at different experimental conditions are assumed to be independent. We further assume that the (non-linear) regression function $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is continuously differentiable with respect to the parameter $\boldsymbol{\theta}$ and define

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \nabla \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta}) = \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_k} \right)^\top, \tag{4.2}$$

as the gradient of η with respect to the parameter $\boldsymbol{\theta}$. The information matrix of a design ξ for model (4.1) is given by

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \boldsymbol{f}^\top(\boldsymbol{x}, \boldsymbol{\theta}) \xi(d\boldsymbol{x}). \tag{4.3}$$

If n observations are taken according to an approximate design (applying an appropriate rounding procedure) it is well know, that under standard assumptions, the covariance matrix of the maximum likelihood estimate of the parameter $\boldsymbol{\theta}$ is approximately given by the matrix $\sigma^2/n\boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})$ and a locally optimal design maximizes an information function of the matrix $\boldsymbol{M}(\xi, \boldsymbol{\theta})$. Consequently, the results of the previous sections can be used to characterize properties of admissible designs for locally optimal design problems, where the vector of regression function is given by the gradient $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$ defined in (4.2). We illustrate this in a few examples.

4.1 Exponential regression

Dette et al. (2006a,b) studied optimal designs for the exponential regression model

$$\boldsymbol{\eta}(x, \boldsymbol{\theta}) = \sum_{l=1}^{L} a_l \exp(-\lambda_l x), \quad x \in \mathcal{X} = [0, \infty), \tag{4.4}$$

where the vector of parameters is given by $\boldsymbol{\theta} = (a_1, \ldots, a_L, \lambda_1, \ldots, \lambda_L)^\top$ with $a_l \neq 0$, $l = 1, \ldots, L$, and $0 < \lambda_1 < \cdots < \lambda_L$. For $L = 1$ or 2, Dette et al. (2006b) showed that there exists a locally D-optimal design based on $2L$ points. Moreover, for $L \geq 3$ they defined $\widetilde{\boldsymbol{\lambda}} = (\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_L)^\top$ as the vector with components satisfying $0 < \widetilde{\lambda}_1 < \cdots < \widetilde{\lambda}_L$ and $\widetilde{\lambda}_{i+1} = (\widetilde{\lambda}_i + \widetilde{\lambda}_{i+2})/2$, $i = 1, \ldots, L-2$, and showed that for any vector $\widetilde{\boldsymbol{\lambda}}$ of this type the existence of a neighbourhood $\mathcal{U}$ of $\widetilde{\boldsymbol{\lambda}}$, such that for all vectors $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)^\top \in \mathcal{U}$, there exists a locally D-optimal design for the parameter $\boldsymbol{\theta}$ which is supported on $2L$ points. Moreover, they pointed out that numerical results indicate that the set of parameter vectors $\boldsymbol{\lambda}$ for which the locally D-optimal design is minimally supported is usually very large. Yang and Stufken (2012) established similar conclusions for optimal designs with respect to other criteria in the cases $L = 2$ (here the condition $\lambda_1/\lambda_2 < 61.98$ is sufficient for the existence of locally optimal design supported at 4 points) and $L = 3$ (here the conditions $2\lambda_2 = \lambda_1 + \lambda_3$ and $\lambda_2/\lambda_1 < 23.72$ imply the existence of an optimal design supported at 6 points).

We now extend these results in a non-trivial manner using the methodology developed in Section 2.

Theorem 4.1. *Consider the exponential regression model* (4.4) *on the interval* $[0, \infty)$*, where* $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_L$ *and* $a_l \neq 0$ $l = 1, \ldots, L$*. If the parameters satisfy* $\lambda_i \geq \frac{|a_i|}{2}$ *for all* $i = 1, \ldots, L$*, any* ϕ*-optimal design is supported at at most* $2L$ *points. Moreover, for all* $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_L$ *and* $a_1, \ldots, a_L \neq 0$ *there exists a locally* D*-optimal design supported at* $2L$ *points.*

4.2 Exponential regression models with two factors

Rodríguez et al. (2015) considered the maximin optimal design problem for the two-factor exponential growth model

$$\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \exp(-\theta_1 x_1) + \exp(-\theta_2 x_2), \tag{4.5}$$

$(\theta_j \geq 1,\ j = 1, 2)$ on the square $\mathcal{X} = [0, 1]^2$, which have numerous applications in biological and agricultural sciences. In this model the gradient of the function $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})$ in (4.5) is given by

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = (1, -x_1 \exp(-\theta_1 x_1), -x_2 \exp(-\theta_2 x_2))^\top \tag{4.6}$$

and the two vectors of regression functions corresponding to the marginal models of (4.6) are obtained as

$$\widetilde{\boldsymbol{f}}_i(x_i, \boldsymbol{\theta}) = (1, x_i \exp(-\theta_i x_i))^\top, \quad i = 1, 2. \tag{4.7}$$

The admissible designs for the marginal models are supported at the points $\{0, 1/\theta_i\}$ ($i = 1, 2$), and it now follows from Theorem 3.1 and Corollary 3.1 that the admissible designs for model (4.5) are contained in the class of all designs supported at most 4 points from the set $\{(0,0), (0, 1/\theta_2),\ (1/\theta_1, 0), (1/\theta_1, 1/\theta_2)\}$. For example, a straightforward optimization shows that the locally D-optimal design puts masses 1/4 at all four points.

Similarly, admissible designs can be determined for the two-factor exponential model

$$\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 \exp(\theta_1 x_1 + \theta_2 x_2), \tag{4.8}$$

where $\theta_j > 0$, $j = 1, 2, 3$ and the design space is given by $\mathcal{X} = [0, b_1] \times [0, b_2]$. Grigoriev et al. (2018) investigated the locally D-optimal designs for this model by means of a general equivalence theorem and showed that locally D-optimal designs are supported at at most 4 points.

The gradient of the function $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})$ in model (4.8) is given by

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \exp(\theta_1 x_1 + \theta_2 x_2)\big(1, \theta_0 x_1, \theta_0 x_2\big)^\top. \tag{4.9}$$

For given $\boldsymbol{\theta}$ let $t = \theta_1 x_1 + \theta_2 x_2$ and define the matrix

$$\boldsymbol{C}(t, \boldsymbol{\theta}) = \begin{pmatrix} \exp(t) & 0 & 0 \\ 0 & \theta_0 \exp(t) & 0 \\ 0 & 0 & \theta_0 \exp(t) \end{pmatrix},$$

then the vector of regression functions corresponding to the conditional model of (4.9) is given by

$$\widetilde{\boldsymbol{f}}_t(\boldsymbol{x}) = (1, x_1, x_2)^\top \tag{4.10}$$

and the design space for the conditional model is given by $\widetilde{\mathcal{X}}(t) = \{\boldsymbol{x} \in \mathcal{X} : \theta_1 x_1 + \theta_2 x_2 = t\}$. For every $t \in \mathcal{T} = \{t = \theta_1 x_1 + \theta_2 x_2 : \boldsymbol{x} \in \mathcal{X}\}$, it is easy to see that the admissible design for the conditional model (4.10) on the design region $\widetilde{\mathcal{X}}(t)$ is supported at the two end points of the line segment $\ell : \theta_1 x_1 + \theta_2 x_2 = t$, $\boldsymbol{x} \in \mathcal{X}$. Therefore, by Theorem 3.1, the admissible designs for the model (4.8) are supported on the boundary of the design region $\mathcal{X}$.

Moreover, the two marginal models of (4.9) exist with corresponding vectors of regression functions given by

$$\widetilde{\boldsymbol{f}}_i(x_i, \boldsymbol{\theta}) = (\exp(\theta_i x_i), x_i \exp(\theta_i x_i))^\top, \quad i = 1, 2. \tag{4.11}$$

The admissible designs for the i-th marginal model are supported at two points, one of which is b_i ($i = 1, 2$) (see Yang and Stufken, 2009, Theorem 5). It now follows from Theorem 3.1 and Corollary 3.1 that the admissible designs for model (4.8) are contained in the class of all designs supported at the 4 points, (b_1, b_2), (b_1, x_2), (x_1, b_2), (x_1, x_2), where $x_i \in [0, b_i)$ and the point (x_1, x_2) is a boundary point of $\mathcal{X} = [0, b_1] \times [0, b_2]$.

4.3 Mixture of exponentials and polynomials

Rodríguez et al. (2015) considered the maximin optimal design problem for the model

$$\boldsymbol{\eta}(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \exp(-\theta_3 x_2), \tag{4.12}$$

which was used in Langseth et al. (2012) for approximating the potentials associated with general hybrid Bayesian networks. The parameter θ_3 is assumed to be positive, design space is given by $\mathcal{X} = [-1,1] \times [0,2]$ and the gradient of the function $\boldsymbol{\eta}(\boldsymbol{x},\boldsymbol{\theta})$ in (4.12) is obtained as

$$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta}) = \big(1, x_1, x_1^3, -x_2 \exp(-\theta_3 x_2)\big)^\top \ . \tag{4.13}$$

The marginal models of (4.13) exist with corresponding vectors of regression functions given by

$$\widetilde{\boldsymbol{f}}_1(x_1,\boldsymbol{\theta}) \ = \ (1, x_1, x_1^3)^\top, \quad \mathcal{X}_1 = [-1,1] \tag{4.14}$$

$$\widetilde{\boldsymbol{f}}_2(x_2,\boldsymbol{\theta}) \ = \ (1, x_2 \exp(-\theta_3 x_2))^\top, \quad \mathcal{X}_2 = [0,2]. \tag{4.15}$$

The admissible designs for the marginal model (4.14) are supported at at most 4 points including end points -1 and 1 (see Yang, 2010, Theorem 8). In addition, it follows from Corollary 2.1 that the other two support points, say u^* and v^*, satisfy the condition $\|\widetilde{\boldsymbol{f}}_1(u,\boldsymbol{\theta})\| = \|\widetilde{\boldsymbol{f}}_1(v,\boldsymbol{\theta})\|$, which implies $u^* = -v^*$. For the marginal model (4.15) the admissible designs are supported at the points $\{0, x_2^*\}$, where $x_2^* = \min\{1/\theta_3, 2\}$. It now follows from Corollary 3.1 that the admissible designs for model (4.12) are contained in the class of all designs supported at the 8 points, $(\pm 1, 0)$, $(\pm 1, x_2^*)$, $(\pm u^*, 0)$, $(\pm u^*, x_2^*)$, where $u^* \in [0,1)$. For example, the locally D-optimal design for model (4.12) is equally supported at $(\pm 1, 0), (\pm\sqrt{3}, 0), (\pm 1, x_2^*), (\pm\sqrt{3}, x_2^*)$.

5 Appendix: proofs

5.1 Preliminaries

In this section we recall some general results from optimal design theory which will be used in subsequent proofs. For more details the reader is referred to the monograph of Pukelsheim (2006).

The polar function $\phi^\infty : \mathrm{NND}(s) \to [0;\infty)$ of an information function ϕ: $\mathrm{PD}(s) \to (0,\infty)$ is defined by

$$\phi^\infty(\boldsymbol{D}) = \inf_{\boldsymbol{C} \in \mathrm{PD}(s)} \frac{\mathrm{trace}(\boldsymbol{C}\boldsymbol{D})}{\phi(\boldsymbol{C})} \ . \tag{5.1}$$

For every information function function ϕ the corresponding polar function ϕ^∞ is isotonic relative to the Loewner ordering. Define

$$\mathcal{N} = \{\boldsymbol{N} \in \text{NND}(k) : \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{N}\boldsymbol{f}(\boldsymbol{x}) \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\}, \tag{5.2}$$

then a duality relation of the optimal design problem can be established [see Pukelsheim (2006), Theorem 7.12], that is

$$\max_{\boldsymbol{M} \in \mathcal{M}(\Xi)} \phi(\boldsymbol{M}) = \min_{N \in \mathcal{N}} \frac{1}{\phi^\infty(\boldsymbol{N})}. \tag{5.3}$$

In particular, an information matrix $\boldsymbol{M} \in \mathcal{M}(\Xi)$ is optimal for $\boldsymbol{\theta}$ in $\mathcal{M}(\Xi)$ if and only if there exists a matrix $\boldsymbol{N} \in \mathcal{N}$ such that

$$\phi(\boldsymbol{M}) = \frac{1}{\phi^\infty(\boldsymbol{N})}, \tag{5.4}$$

and two matrices $\boldsymbol{M} \in \mathcal{M}(\Xi)$ and $\boldsymbol{N} \in \mathcal{N}$ satisfy (5.4) if and only if the conditions

$$\text{trace}(\boldsymbol{M}\boldsymbol{N}) = 1, \tag{5.5}$$

$$\phi(\boldsymbol{M})\phi^\infty(\boldsymbol{N}) = \text{trace}(\boldsymbol{M}\boldsymbol{N}) \tag{5.6}$$

hold. An application of this result yields the famous general equivalence theorem in optimal design theory.

Theorem 5.1 (Pukelsheim (2006), Theorem 7.17)**.** *A positive definite information matrix $\boldsymbol{M}^* \in \mathcal{M}(\Xi)$ is ϕ-optimal for $\boldsymbol{\theta}$ in $\mathcal{M}(\Xi)$ if and only if there exists a nonnegative definite $k \times k$ matrix $\boldsymbol{N} \in \mathcal{N}$ that solves the polarity equation*

$$\phi(\boldsymbol{M}^*)\phi^\infty(\boldsymbol{N}) = \mathit{trace}(\boldsymbol{M}^*\boldsymbol{N}) = 1$$

and that satisfies the normality inequality

$$\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{N}\boldsymbol{f}(\boldsymbol{x}) \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}. \tag{5.7}$$

Moreover, if $\boldsymbol{M}^*$ is optimal for $\boldsymbol{\theta}$ in Ξ, there is equality for any support point $\boldsymbol{x}_i$ of any ϕ-optimal design $\xi \in \Xi$, that is any design with $\boldsymbol{M}^* = \boldsymbol{M}(\xi)$.

5.2 Proof of Theorem 2.1

The matrix $\boldsymbol{N}$ in NND(k) has an eigenvalue decomposition

$$\boldsymbol{N} = \boldsymbol{Z}_{\boldsymbol{N}}\boldsymbol{\Lambda}_{\boldsymbol{N}}\boldsymbol{Z}_{\boldsymbol{N}}^\top,$$

where $\boldsymbol{\Lambda}_{\boldsymbol{N}} = \text{diag}\ (\lambda_{\boldsymbol{N}1}, \ldots, \lambda_{\boldsymbol{N}k})$ is a diagonal matrix, the eigenvalues $\lambda_{N1}, \ldots, \lambda_{Nk}$ of $\boldsymbol{N}$ are counted with their respective multiplicities, and $\boldsymbol{Z}_{\boldsymbol{N}} = (\boldsymbol{z}_{\boldsymbol{N}1}, \ldots, \boldsymbol{z}_{\boldsymbol{N}k})$ is an orthogonal matrix with eigenvectors corresponding to the eigenvalues. Denote by $\mathcal{S}_{\boldsymbol{Z}}$ the subset of NND(k) consisting of matrices which permit eigenvalue decomposition with the same orthogonal matrix $\boldsymbol{Z}$, i.e.,

$$\mathcal{S}_{\boldsymbol{Z}} = \{\boldsymbol{N} : \boldsymbol{N} = \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{Z}^\top, \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_k), \lambda_i \geq 0, i = 1, \ldots, k\}.$$

Then we can express NND(k) as

$$\text{NND}(k) = \bigcup_{\boldsymbol{Z} \in O(k)} \mathcal{S}_{\boldsymbol{Z}},$$

where $O(k)$ is the set of all $k \times k$ orthogonal matrices.

Furthermore, let

$$\boldsymbol{h}_{\boldsymbol{Z}}(\boldsymbol{x}) = \boldsymbol{Z}^\top \boldsymbol{f}(\boldsymbol{x}) = (\boldsymbol{z}_1^\top \boldsymbol{f}(\boldsymbol{x}), \ldots, \boldsymbol{z}_k^\top \boldsymbol{f}(\boldsymbol{x}))^\top = (h_{\boldsymbol{Z}1}(\boldsymbol{x}), \ldots, h_{\boldsymbol{Z}k}(\boldsymbol{x}))^\top,$$

then the set $\mathcal{N}$ in (5.2) can be represented as

$$\begin{aligned}
\mathcal{N} &= \textstyle\bigcup_{\boldsymbol{Z} \in O(k)} \{\boldsymbol{N} \in \mathcal{S}_{\boldsymbol{Z}} : \boldsymbol{f}^\top(\boldsymbol{x}) \boldsymbol{N} \boldsymbol{f}(\boldsymbol{x}) \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\} \\
&= \textstyle\bigcup_{\boldsymbol{Z} \in O(k)} \{\boldsymbol{N} \in \mathcal{S}_{\boldsymbol{Z}} : \boldsymbol{h}_{\boldsymbol{Z}}^\top(\boldsymbol{x}) \boldsymbol{\Lambda} \boldsymbol{h}_{\boldsymbol{Z}}(\boldsymbol{x}) \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\} \\
&= \textstyle\bigcup_{\boldsymbol{Z} \in O(k)} \{\boldsymbol{N} \in \mathcal{S}_{\boldsymbol{Z}} : \boldsymbol{h}_{\boldsymbol{Z}1}^2(\boldsymbol{x})\lambda_1 + \cdots + \boldsymbol{h}_{\boldsymbol{Z}k}^2(\boldsymbol{x})\lambda_k \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\} \\
&\widehat{=} \textstyle\bigcup_{\boldsymbol{Z} \in O(k)} \mathcal{N}_{\boldsymbol{Z}},
\end{aligned}$$

where the last line defines the set $\mathcal{N}_{\boldsymbol{Z}}$ in an obvious manner. The optimal solution of the dual problem must occur on some subset, say $\mathcal{N}_{\boldsymbol{Z}^*}$. Moreover, the dual problem on any subset $\mathcal{N}_{\boldsymbol{Z}}$ can be viewed as an extremum problem of a multivariate function defined on the convex polytope (2.7), that is

$$\{(\lambda_1, \ldots, \lambda_k) : \lambda_i \geq 0, \ i = 1, \ldots, k, \ \boldsymbol{h}_{\boldsymbol{Z}1}^2(\boldsymbol{x})\lambda_1 + \cdots + \boldsymbol{h}_{\boldsymbol{Z}k}^2(\boldsymbol{x})\lambda_k \leq 1 \ \ \forall \boldsymbol{x} \in \mathcal{X}\}.$$

Note that the polar function ϕ^∞ is isotonic, hence the optimal solution must be attained at a boundary point of the polytope. Let $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_k^*)$ be the vector corresponding to the extremum, then there are at most k effective constraints, say

$$\ell_i : c_{i1}\lambda_1 + \cdots + c_{ik}\lambda_k = 1, \quad i = 1, \ldots, k,$$

which must be satisfied by $\boldsymbol{\lambda}^*$. Corresponding to $\boldsymbol{\lambda}^*$ we define $\boldsymbol{N}^* = \boldsymbol{Z}^* \boldsymbol{\Lambda}^* \boldsymbol{Z}^{*\top}$ with $\boldsymbol{\Lambda}^* = \text{diag}\ (\lambda_1^*, \ldots, \lambda_k^*)$, then $\boldsymbol{N}^*$ is the optimal matrix of the dual problem.

By Theorem 5.1, we have

$$1 = \boldsymbol{f}^\top(\boldsymbol{x}^*)\boldsymbol{N}^*\boldsymbol{f}(\boldsymbol{x}^*) = h^2_{\boldsymbol{Z}^*1}(\boldsymbol{x}^*)\lambda_1^* + \cdots + h^2_{\boldsymbol{Z}^*k}(\boldsymbol{x}^*)\lambda_k^*$$

for any support point $\boldsymbol{x}^*$ of a ϕ-optimal design ξ^*, which implies that

$$(h^2_{\boldsymbol{Z}^*1}(\boldsymbol{x}^*), \ldots, h^2_{\boldsymbol{Z}^*k}(\boldsymbol{x}^*)) \in \{\boldsymbol{c}_i = (c_{i1}, \ldots, c_{ik})^\top, i = 1, \ldots, k\},$$

and $h^2_{\boldsymbol{Z}^*1}(\boldsymbol{x}^*)\lambda_1 + \cdots + h^2_{\boldsymbol{Z}^*k}(\boldsymbol{x}^*)\lambda_k = 1$ is a supporting hyperplane of the polytope (2.7) with $\boldsymbol{Z} = \boldsymbol{Z}^*$. Consequently, the support points of ξ^* can be divided into k sets, say $\{\boldsymbol{x}_{ij}, j = 1, \ldots, m_i\}$, according to the vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$. Moreover, the support points in the i-th set satisfy

$$(h^2_{\boldsymbol{Z}^*1}(\boldsymbol{x}_{ij}), \ldots, h^2_{\boldsymbol{Z}^*k}(\boldsymbol{x}_{ij})) = \boldsymbol{c}_i, \quad j = 1, \ldots, m_i,$$

which yields

$$(\boldsymbol{z}_l^{*\top}\boldsymbol{f}(\boldsymbol{x}_{ij}))^2 = c_{il}, \quad l = 1, \ldots, k;\ j = 1, \ldots, m_i.$$

Therefore, the vectors $\boldsymbol{f}(\boldsymbol{x}_{ij}), j = 1, \ldots, m_i$ share the same length since

$$\|\boldsymbol{f}(\boldsymbol{x}_{ij})\|^2 = \|\boldsymbol{Z}^{*\top}\boldsymbol{f}(\boldsymbol{x}_{ij})\|^2 = \sum_{l=1}^{k}(\boldsymbol{z}_l^{*\top}\boldsymbol{f}(\boldsymbol{x}_{ij}))^2 = \|\boldsymbol{c}_i\|^2,$$

which completes the proof of Theorem 2.1.

5.3 Proof of Theorem 3.1

For fixed t, there exists a full column-rank matrix, say $\boldsymbol{C}(t)$, such that

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{C}(t)\widetilde{\boldsymbol{f}}_t(\boldsymbol{x})$$

on the design space $\widetilde{\mathcal{X}}(t)$, since the elements of the vector $\widetilde{\boldsymbol{f}}_t(\boldsymbol{x})$ are linearly independent. Suppose there exists some $t_* \in \mathcal{T}$ with $\xi_t(t_*) > 0$ such that the conditional design $\xi_{\boldsymbol{x}|t_*}$ is inadmissible for the conditional model (3.2). Then there exists a design $\bar{\xi}_{\boldsymbol{x}|t_*}$ in a set of all conditional designs Ξ_{t_*} satisfying

$$\boldsymbol{M}_{t_*}(\bar{\xi}_{\boldsymbol{x}|t_*}) \geq \boldsymbol{M}_{t_*}(\xi_{\boldsymbol{x}|t_*}) \quad \text{and} \quad \boldsymbol{M}_{t_*}(\bar{\xi}_{\boldsymbol{x}|t_*}) \neq \boldsymbol{M}_{t_*}(\xi_{\boldsymbol{x}|t_*}).$$

Let $\bar{\xi}$ be the design obtained by replacing the conditional design $\xi_{\boldsymbol{x}|t_*}$ of ξ with $\bar{\xi}_{\boldsymbol{x}|t_*}$, then we have

$$\begin{aligned}
\boldsymbol{M}(\xi) &= \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^\top(\boldsymbol{x})\xi(d\boldsymbol{x}) \\
&= \int_{\mathcal{T}}\int_{\widetilde{\mathcal{X}}(t)} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^\top(\boldsymbol{x})\xi_{\boldsymbol{x}|t}(d\boldsymbol{x})\xi_t(dt) \\
&= \int_{\mathcal{T}}\int_{\widetilde{\mathcal{X}}(t)} \boldsymbol{C}(t)\widetilde{\boldsymbol{f}}_t(\boldsymbol{x})\widetilde{\boldsymbol{f}}_t^\top(\boldsymbol{x})\boldsymbol{C}^\top(t)\xi_{\boldsymbol{x}|t}(d\boldsymbol{x})\xi_t(dt) \\
&= \int_{\mathcal{T}} \boldsymbol{C}(t)\left[\int_{\widetilde{\mathcal{X}}(t)} \widetilde{\boldsymbol{f}}_t(\boldsymbol{x})\widetilde{\boldsymbol{f}}_t^\top(\boldsymbol{x})\xi_{\boldsymbol{x}|t}(d\boldsymbol{x})\right]\boldsymbol{C}^\top(t)\xi_t(dt) \\
&= \int_{\mathcal{T}} \boldsymbol{C}(t)\boldsymbol{M}_t(\xi_{\boldsymbol{x}|t})\boldsymbol{C}^\top(t)\xi_t(dt) \\
&\lneq \int_{\mathcal{T}} \boldsymbol{C}(t)\boldsymbol{M}_t(\bar{\xi}_{\boldsymbol{x}|t})\boldsymbol{C}^\top(t)\xi_t(dt) \\
&= \int_{\mathcal{T}} \boldsymbol{C}(t)\boldsymbol{M}_t(\bar{\xi}_{\boldsymbol{x}|t})\boldsymbol{C}^\top(t)\bar{\xi}_t(dt) \\
&= \boldsymbol{M}(\bar{\xi}).
\end{aligned}$$

Therefore, the design ξ would be inadmissible in the class Ξ for the model (3.1), and this contradiction completes the proof of Theorem 3.1.

5.4 Proof of Theorem 4.1

The fist part follows directly from an application of Theorem 2.2(b). To be precise, note that the gradient of the function $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})$ with respect to the parameter $\boldsymbol{\theta}$ in model (4.4) is given by

$$\boldsymbol{f}(x, \boldsymbol{\theta}) = (\exp(-\lambda_1 x), -a_1 x \exp(-\lambda_1 x), \ldots, \exp(-\lambda_L x), -a_L x \exp(-\lambda_L x))^\top \tag{5.8}$$

It is easy to see that the function $x \to \|\boldsymbol{f}(x, \boldsymbol{\theta})\|^2 = \sum_{i=1}^L \exp(-2\lambda_i x)(1 + a_i^2 x^2)$ is a strictly decreasing function on the interval $[0, \infty)$ if $\lambda_i \geq \frac{|a_i|}{2}$ for all $i = 1, \ldots, L$, since in this case the derivative of function $\|\boldsymbol{f}(x, \boldsymbol{\theta})\|^2$ is non-positive.

For the statement regarding the D-optimality criterion recall that the parameter vector in (5.8) is given by $\boldsymbol{\theta} = (a_1, \ldots, a_L, \lambda_1, \ldots, \lambda_L)^\top$. Let $c > 0$ be any constant and note that the vector of regression functions satisfies

$$\boldsymbol{f}(x, \boldsymbol{\theta}) \;=\; \boldsymbol{Q}\boldsymbol{g}(x, \boldsymbol{\theta}),$$

where

$$\boldsymbol{g}(x, \boldsymbol{\theta}) = (\exp(-\lambda_1 x), cx \exp(-\lambda_1 x), \ldots, \exp(-\lambda_L x), cx \exp(-\lambda_L x))^\top.$$

and the matrix $\boldsymbol{Q}$ is given by $\boldsymbol{Q} = \text{diag}(1, -a_1/c, \ldots, 1, -a_L/c)$. Observing the relation

$$\int_{\mathcal{X}} \boldsymbol{f}(x, \boldsymbol{\theta}) \boldsymbol{f}^\top(x, \boldsymbol{\theta}) \xi(d\boldsymbol{x}) = \boldsymbol{Q} \int_{\mathcal{X}} \boldsymbol{g}(x, \boldsymbol{\theta}) \boldsymbol{g}^\top(x, \boldsymbol{\theta}) \xi(d\boldsymbol{x}) \boldsymbol{Q}^\top$$

it is easy to see that a design is D-optimal for the regression model (2.1) with vector $\boldsymbol{f}$ if and only if it is D-optimal for the regression model (2.1) with vector $\boldsymbol{g}$. However, from the first part of the proof this design is the locally D-optimal design if $\lambda_i \geq c/2$ for all $i = 1, \ldots, L$. As the constant $c > 0$ can be chosen arbitrarily, the assertion of Theorem 4.1 follows.

Acknowledgements Dr. Liu and Dr. Yue were partially supported by the National Natural Science Foundation of China under Grants 11871143, 11971318. The work of Dr. Dette 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

Biedermann, S., Dette, H., and Woods, D. C. (2011). Optimal design for additive partially nonlinear models. *Biometrika*, 98:449–458.

Castro, Y. D., Gamboa, F., Henrion, D., Hess, R., and Lasserre, J.-B. (2019). Approximate optimal designs for multivariate polynomial regression. *Ann. Statist.*, 47:127–155.

Chang, F. and Jiang, B. (2007). An algebraic construction of minimally-supported D-optimal designs for weighted polynomial regression. *Statistica Sinica*, 17:1005–1021.

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

Chang, F.-C. and Lin, H.-M. (2006). On minimally-supported D-optimal designs for polynomial regression with log-concave weight function. *Metrika*, 65:227–233.

de la Garza, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Statist.*, 25:123–130.

Dette, H. and Grigoriev, Y. (2014). E -optimal designs for second-order response surface models. *Ann. Statist.*, 42(4):1635–1656.

Dette, H., Haines, L., and Imhof, L. (2005). Bayesian and maximin optimal designs for heteroscedastic regression models. *Canadian Journal of Statistics*, 33(2):221–241.

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

Dette, H., Melas, V. B., and Pepelyshev, A. (2006a). Local c- and e-optimal designs for exponential regression models. *Annals of the Institute of Statistical Mathematics*, 58:407–426.

Dette, H., Melas, V. B., and Wong, W. K. (2006b). Locally D-optimal designs for exponential regression. *Statistica Sinica*, 16:789–803.

Dette, H. and Schorning, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *Ann. Statist.*, 41(3):1260–1267.

Fang, Z. (2002). D-optimal designs for polynomial regression models through origin. *Statistics & Probability Letters*, 57:343–351.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Graßhoff, U., Großmann, H., Holling, H., and Schwabe, R. (2007). Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor. *J. Statist. Plann. Inference*, 137:3882–3893.

Grigoriev, Y. D., Melas, V. B., and Shpilev, P. V. (2018). Excess of locally D-optimal designs for Cobb-Douglas model. *Statist. Papers*, 59(4):1425–1439.

Heiligers, B. (1992). Admissible experimental designs in multiple polynomial regression. *J. Statist. Plann. Inference*, 31:219–233.

Hu, L., Yang, M., and Stufken, J. (2015). Saturated locally optimal designs under differentiable optimality criteria. *Ann. Statist.*, 43(1):30–56.

Huang, S.-H., Huang, M.-N. L., and Lin, C.-W. (2020). Optimal designs for binary response models with multiple nonnegative variables. *J. Statist. Plann. Inference*, 206:75–83.

Kabera, M., Haines, L., and Ndlovu, P. (2018). D-optimal designs for the two-variable binary logistic regression model with interaction. *J. Statist. Plann. Inference*, 193:136–150.

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

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

Langseth, H., Nielsen, T. D., Rumí, R., and Salmerón, A. (2012). Mixtures of truncated basis functions. *Internat. J. Approx. Reason.*, 53:212–227.

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

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

Rodríguez, C., Ortiz, I., and Martínez, I. (2015). Locally and maximin optimal designs for multi-factor nonlinear models. *Statistics*, 49:1157–1168.

Sekido, H. (2012). An algorithm for calculating D-optimal designs for polynomial regression through a fixed point. *Journal of Statistical Planning and Inference*, 142(4):935–943.

Yang, M. (2010). On the de la Garza phenomenon. *Ann. Statist.*, 38(4):2499–2524.

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

Yang, M. and Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Ann. Statist.*, 40:1665–1685.

Yang, M., Zhang, B., and Huang, S.-G. (2011). Optimal designs for generalized linear models with multiple design variables. *Statistica Sinica*, 21:1415–1430.