DESIGN ADMISSIBILITY AND DE LA GARZA PHENOMENON IN MULTI-FACTOR EXPERIMENTS

BY HOLGER DETTE¹, XIN LIU²,* AND RONG-XIAN YUE³,†

¹Ruhr-Universität Bochum holger.dette@rub.de
²Donghua University liuxin@dhu.edu.cn
³Shanghai Normal University yue2@shnu.edu.cn

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 minimally supported 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 $k - 1$ any optimal design can be based on $k$ 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.

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, Donev and Atkinson, 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 for the information matrix of any inadmissible design there exists a design in the class of admissible designs with a larger (or equal) information matrix with respect to the Loewner ordering. It is well known that optimal designs with respect to most of the commonly used optimality criteria must be admissible (see Pukelsheim, 2006, Chapter 10.10) and consequently in these cases the optimal designs can be found in 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, Yang and Stufken (2015) derived several complete classes of designs for regression models with a univariate

¹MSC 2010 subject classifications: Primary 62K05; secondary 62J05
²Keywords and phrases: admissibility, dual problem, conditional model, multi-factor experiment, optimal design
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 $k - 1$ any optimal design can be based on at most $k$ 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 on the theory of Chebyshev systems (see Karlin and Studden, 1966), which requires regression functions with a univariate argument. Consequently, for regression models with a 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, Zhang and Huang (2011) derived a class of admissible designs for the commonly used multi-factor logistic and probit models. Huang, Huang and Lin (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, Dette and Woods (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, Melas and Shpilev (2018), who discussed locally $D$-optimal designs for the Cobb-Douglas model, Kabera, Haines and Ndlovu (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. 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. Our general strategy is to handle the design problem by considering the dual optimization problem.

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 the Appendix.

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 $\text{Sym}(k)$ denote the set of all real symmetric $k \times k$ matrices, $\text{NND}(k) \subset \text{Sym}(k)$ the set of all nonnegative definite matrices and $\text{PD}(k) \subset \text{NND}(k)$ the set of positive definite matrices. We consider the common linear regression model

\[ y = f^\top(x)\theta + \varepsilon, \]
where \( x = (x_1, \ldots, x_q)^\top \) is a \( q \)-dimensional vector of predictors which varies in a compact design space \( \mathcal{X} \subset \mathbb{R}^d \), \( f(x) \) is a \( k \)-dimensional vector of known linearly independent, continuous regression functions, \( \theta \in \mathbb{R}^k \) denotes the vector of unknown parameters, and \( \varepsilon \) 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 \( y_i = f^\top(x_i)\theta + \varepsilon_i \) \((i = 1, \ldots, n)\) at experimental conditions \( x_1, \ldots, x_n \).

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

\[
M(\xi) = \int_{\mathcal{X}} f(x)f^\top(x)\xi(dx) \in \mathbb{R}^{k \times k}.
\]

If the design \( \xi \) has masses \( w_1, \ldots, w_m \) at \( m \) support points \( x_1, \ldots, x_m \), and \( n \) observations can be taken, the quantities \( w_\ell \) \((\ell = 1, \ldots, m)\) are rounded to non-negative integers, say \( n_\ell \), such that \( \sum_{\ell=1}^m n_\ell = n \) and the experimenter takes \( n_\ell \) observations at each \( x_\ell \) \((\ell = 1, \ldots, m)\). In this case the covariance matrix of the least squares estimator \( \sqrt{n}\theta \) for the parameter \( \theta \) in model (2.1) converges to the matrix \( \sigma^2 M^{-1}(\xi) \), which is used to measure the accuracy of the estimator \( \hat{\theta} \).

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

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

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

Given an optimality criterion \( \phi \) on \( \text{NND}(k) \) the design problem then reads as follows

\[
\max_{\xi \in \Xi} \phi(M(\xi)).
\]

A design \( \xi^* \) corresponding to the maximum in (2.4) is called a \( \phi \)-optimal design. As the design space is compact and the vector of regression functions is continuous, it follows that the set of information matrices \( M(\Xi) \) is compact as well. According to the existence theorem in Section 7.13 of Pukelsheim (2006), there exists a \( \phi \)-optimal design. Moreover, by Carathéodory's theorem, we can always restrict our consideration to designs with at most \( k(k+1)/2 \) support points. 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 minimally supported, which means that the number of support points of the design coincides with the dimension of the parameter. This property is called the 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 \( k - 1 \) on a compact interval has \( k \) 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 minimally supported designs in models with a multivariate predictor.
We begin with a geometric characterization of the support points of a \( \phi \)-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 \( Z = (z_1, \ldots, z_k) \in \mathbb{R}^{k \times k} \) a linear transformation \( h_Z : \mathbb{R}^k \to \mathbb{R}^k \) by
\[
(2.5) \quad h_Z(x) := (h_{Z1}(x), \ldots, h_{Zk}(x))^\top := Z^\top f(x) = (z_1^\top f(x), \ldots, z_k^\top f(x))^\top
\]
and consider the corresponding point
\[
(2.6) \quad P_Z(x) = (h_{Z1}^2(x), \ldots, h_{Zk}^2(x))^\top \in \mathbb{R}^k.
\]
Let
\[
(2.7) \quad \mathcal{P}_Z := \{ \lambda \in \mathbb{R}_2^k : P_Z(x) \lambda \leq 1 \ \forall x \in \mathcal{X}\}
\]
denote the convex polytope associated with the point \( P_Z(x) \), where
\[
(2.8) \quad \mathbb{R}_2^k = \{ \lambda = (\lambda_1, \ldots, \lambda_k)^\top : \lambda_1 \geq \ldots \geq \lambda_k \geq 0 \} \subset \mathbb{R}^k
\]
denotes the subset of all \( k \)-dimensional vectors with nonnegative ordered components. Note that multiplication of a column of the matrix \( Z \) by \(-1\) does not change the polytope. Similarly, interchanging two columns \( z_\ell \) and \( z_{\ell+1} \) of \( Z \) corresponding to equal values \( \lambda_\ell \) and \( \lambda_{\ell+1} \) does not change the structure of \( \mathcal{P}_Z \). In the following we use the convention that a vector \( e \in \mathbb{R}^k \) defines a supporting hyperplane of a closed convex set \( C \subset \mathbb{R}^k \) if there exists a point \( h^* \in C \) such that \( e^\top h^* = 1 \) and \( e^\top h \leq 1 \) for all \( h \in C \).

**Theorem 2.1.** Let \( \xi^* = \{(x_i^*, w_i^*)\}_{i=1}^m \) be a \( \phi \)-optimal design for the regression model (2.1). There exists an orthogonal matrix, say \( Z^* = (z_1^*, \ldots, z_k^*) \in \mathbb{R}^{k \times k} \) with a linear transformation \( h_{Z^*} \) of the form (2.5), such that the vectors \( P_{Z^*}(x_1^*), \ldots, P_{Z^*}(x_m^*) \) define a set of supporting hyperplanes of the \( k \)-dimensional polytope \( \mathcal{P}_{Z^*} \) defined by (2.7) for the matrix \( Z^* \), and the rank of the \((k+1) \times m\) matrix
\[
(2.9) \quad P_m = (P_Z(x_1^*) \ P_Z(x_2^*) \ldots P_Z(x_m^*))
\]
is at most \( k \). Moreover, if \( f(x_1^*) \) and \( f(x_2^*) \) are two vectors corresponding to the same supporting hyperplane they have the same length.

**Remark 2.1.** The proof of Theorem 2.1 is based on a duality result, which relates the optimal design problem to the minimax problem
\[
(2.10) \quad \min_{N \in \mathcal{N}} \max_{C \in \text{PD}(\phi)} \dfrac{\phi(C)}{\text{trace}(CN)},
\]
where the set \( \mathcal{N} \) is defined by
\[
(2.11) \quad \mathcal{N} = \{ N \in \text{NND}(k) : f^\top(x)Nf(x) \leq 1 \ \forall x \in \mathcal{X} \},
\]
(see the Appendix for more details). In particular, the spectral decomposition of a solution of (2.10), say \( N^* \), defines the orthogonal matrix \( Z^* \) in Theorem 2.1.

**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 \( f(x) = (x_1, x_2)^\top \), where \( x = (x_1, x_2)^\top \in [0, 1]^2 \). If
\[
(2.12) \quad Z = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}
\]
is a $2 \times 2$ orthogonal matrix, then the vector $h_Z$ in (2.5) is given by

$$h_Z(x) = (x_1 \cos t - x_2 \sin t, x_1 \sin t + x_2 \cos t)^T,$$

and it is easy to see that the polytope (2.13)

$$P_Z = \{ (\lambda_1, \lambda_2)^T \in \mathbb{R}^2_+ : (x_1 \cos t - x_2 \sin t)^2 \lambda_1 + (x_1 \sin t + x_2 \cos t)^2 \lambda_2 \leq 1 \ \forall x \in [0,1]^2 \}$$

(which depends on the matrix $Z$ defined in (2.12)) is determined by at most three half planes (besides the half plane $-\lambda_1 + \lambda_2 \leq 0$), which are defined by

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

(2.14)

and are obtained by setting $(x_1, x_2)^T$ in (2.13) to be the points $(1,0)^T, (0,1)^T$ and $(1,1)^T$, respectively. Two polytopes for the choice $t = \pi/8$ and $t = 15/8\pi$ are depicted in the left and right panel of Figure 1, respectively (note that these do not correspond to an optimal design).

Consequently, the support points of any $\phi$-optimal design are contained in the set $$\{(1,0)^T, (0,1)^T, (1,1)^T\},$$ and the corresponding weights can now be found by a straightforward calculation. In fact $\phi_p$-optimal designs were determined by Pukelsheim (2006), Section 8.6, who showed that the $\phi_p$-optimal design for $p \in (-\infty, 1)$ is given by

$$\xi_p^* = \left\{ \frac{1}{w(p)} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\
\end{pmatrix} \right\},$$

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

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

$$M(\xi_D^*) = \begin{pmatrix} 2/3 & 1/3 & 1/3 \\ 1/3 & 2/3 & 1/3 \end{pmatrix}.$$
define three supporting hyperplanes of the polytope \(P_Z\), which are obtained using the support points \((1,0)^\top\), \((0,1)^\top\) and \((1,1)^\top\) of the \(D\)-optimal design in (2.13). Let \(\lambda^* = (\lambda_1^*, \lambda_2^*)^\top\) be the vector of ordered eigenvalues in the spectral decomposition of the solution \(N^*\) of the dual problem (2.10) of the \(D\)-optimal design problem, then it follows from the proof of Theorem 2.1 that these supporting hyperplanes intersect at the point \(\lambda^*\). This leads to \(\lambda_1^* = 3/2, \lambda_2^* = 1/2\), \(t = \pi/4\) or \(\lambda_1^* = 1/2, \lambda_2^* = 3/2\), \(t = 3\pi/4\) (corresponding to the same matrix \(N^*\)). Because the eigenvalues are ordered, the corresponding polytope is obtained for the choice \(t = \pi/4\), that is
\[
P_Z^* = \left\{ (\lambda_1, \lambda_2)^\top : \lambda_1 \geq \lambda_2 \geq 0, \frac{1}{2} \lambda_1 + \frac{1}{2} \lambda_2 \leq 1, 2 \lambda_2 \leq 1 \right\}. \tag{2.15} \]

This polytope is depicted in Figure 2. The two support points \((1,0)^\top\) and \((0,1)^\top\) correspond to the same hyperplane defined by the equation \(\frac{1}{2} \lambda_1 + \frac{1}{2} \lambda_2 = 1\), since the equalities
\[
(\cos(\pi/4)x_1 - \sin(\pi/4)x_2)^2 = \frac{1}{2} \quad \text{and} \quad (\sin(\pi/4)x_1 + \cos(\pi/4)x_2)^2 = \frac{1}{2}
\]
hold for \((x_1, x_2)^\top = (1,0)^\top\) and \((x_1, x_2)^\top = (0, 1)^\top\). The third support point \((1,1)^\top\) corresponds to the other hyperplane \(2 \lambda_2 = 1\), because we have
\[
(\cos(\pi/4)x_1 - \sin(\pi/4)x_2)^2 = 0 \quad \text{and} \quad (\sin(\pi/4)x_1 + \cos(\pi/4)x_2)^2 = 2
\]
for \((x_1, x_2)^\top = (1,1)^\top\). Similarly, the \(\phi_p\)-optimal design with \(p = -\infty\), say \(\xi^*_E\), has equal masses at the points \((0,1)^\top\) and \((1,0)^\top\) corresponding to the same hyperplane \(\frac{1}{2} \lambda_1 + \frac{1}{2} \lambda_2 = 1\) of the polytope (2.15) and the information matrix is given by
\[
M(\xi^*_E) = \begin{pmatrix}
\frac{1}{2} & 0 \\
0 & \frac{1}{2}
\end{pmatrix}.
\]

The corresponding polytope coincides with polytope for the \(D\)-optimal design depicted in Figure 2.

As a direct application of Theorem 2.1, we obtain a sufficient condition for the occurrence of the de la Garza phenomenon in the linear regression model (2.1). Note that by the assumptions stated in Section 2 the existence of a \(\phi\)-optimal design is guaranteed.

**Corollary 2.1.** Each of the following conditions is sufficient for the existence of a \(\phi\)-optimal design with \(k\) support points in the regression model (2.1).
(a) For every orthogonal matrix $Z$ there do not exist $m \geq k + 1$ points in $X$, say $x_1, \ldots, x_m$, such that the vectors $P_Z(x_1^*), \ldots, P_Z(x_m^*)$ define supporting hyperplanes of the polytope $P_Z$.

(b) For every orthogonal matrix $Z$ for which there exist $k + 1$ points $x_1, \ldots, x_{k+1} \in X$, such that the vectors $P_Z(x_1^*), \ldots, P_Z(x_{k+1}^*)$ define supporting hyperplanes of the polytope $P_Z$, the rank of the $(k + 1) \times (k + 1)$ matrix $P_{k+1}$ defined in (2.9) is $k + 1$.

Example 2.2. As an application of Corollary 2.1, we consider the de la Garza phenomenon for a bivariate polynomial regression model on a triangular region. To be precise, let $X = \{x = (x_1, x_2) \in [-1, 1]^2 : x_1 + x_2 \geq 0\}$ and consider the 5-dimensional vector of regression functions

$$f(x) = (1, x_1, x_2, x_1^2, x_1 x_2)^	op.$$ 

For every orthogonal matrix $Z$ of order 5, let $N = [n_{ij}]_{1 \leq i, j \leq 5}$ be a matrix in NND(5) with an eigenvalue decomposition

$$N = Z \Lambda Z^\top,$$

where $\Lambda = \text{diag} (\lambda_1, \ldots, \lambda_5)$ is a diagonal matrix with diagonal elements $\lambda_1 \geq \ldots \geq \lambda_5 \geq 0$. From the proof of Theorem 2.1 it follows that if the vector $P_Z(x^*)$ defines a supporting hyperplane of the polytope $P_Z$ at the point $x^* \in X$, then the inequality

$$d_N(x) = f^\top(x) N f(x) = P_Z^\top(x) \Lambda = h_{Z1}^2(x) \lambda_1 + \cdots + h_{Z5}^2(x) \lambda_5 \leq 1$$

holds for all $x \in X$ and there is equality for $x = x^*$. For fixed $x_2$ or fixed $x_1$ this function has the form

$$d_N(x_1, x_2) = a_1(x_2) x_1^4 + a_2(x_2) x_1^3 + a_3(x_2) x_1^2 + a_4(x_2) x_1 + a_5(x_2),$$

respectively, where

$$a_1(x_2) = n_{41}, \quad a_2(x_2) = 2n_{25} + 2n_{45} x_2,$$

$$a_3(x_2) = 2n_{14} + n_{22} + 2(n_{25} + n_{34}) x_2 + n_{55} x_2^2,$$

$$a_4(x_2) = 2n_{12} + 2(n_{15} + n_{23}) x_2 + 2n_{35} x_2^2,$$

$$a_5(x_2) = n_{11} + 2n_{13} x_2 + n_{33} x_2^2,$$

$$b_1(x_1) = n_{33} + 2n_{35} x_1 + n_{55} x_1^2,$$

$$b_2(x_1) = 2n_{23} + 2(n_{15} + n_{23}) x_1 + 2(n_{25} + n_{34}) x_1^2 + 2n_{45} x_1^3,$$

$$b_3(x_1) = n_{11} + 2n_{12} x_1 + (2n_{14} + n_{22}) x_1^2 + 2n_{25} x_1^3 + n_{44} x_1^4.$$

Note that $a_1(x_2) \geq 0$, $b_1(x_1) \geq 0$ because the matrix $N$ is nonnegative definite. It is easy to see that this function has at most 5 maxima at the points

$$(2.17) \quad \{(−1, 1)^\top, (1, −1)^\top, (1, 1)^\top, (u, v)^\top, (v, −v)^\top\}, \quad u, v \in (−1, 1).$$

Indeed, for any fixed $x_2^* \in [−1, 1]$ the function $d_N(x_1^*, x_2)$ has at most two maxima at the points $x_2 = 1$ and $x_2 = −x_2^*$, which means $d_N(x_1, x_2)$ can only attain maxima on $x_2 = 1$ and $x_1 + x_2 = 0$. If $x_2 = 1$ the function $d_N(x_1, 1)$ has at most three maxima at the points $x_1 = ±1$ and at some point $u \in (−1, 1)$ on the interval $[−1, 1]$. Similarly, for the line segment $\ell = \{(x_1, x_2)^\top \in X : x_1 + x_2 = 0\}$ we have

$$d_N(x_1, −x_1) = c_1 x_1^4 + c_2 x_1^3 + c_3 x_1^2 + c_4 x_1 + c_5,$$
where
\[ c_1 = n_{44} - 2n_{45} + n_{55}, \quad c_2 = 2n_{35} - 2n_{34}, \]
\[ c_3 = 2n_{14} - 2n_{15} + n_{22} - 2n_{23} + n_{33}, \quad c_4 = 2n_{12} - 2n_{13}, \quad c_5 = n_{11}. \]

As \( c_1 \geq 0 \), it follows that the function \( d_N(x_1, -x_1) \) has at most three maxima on the line segment \( \ell \) and two of them have to be located at the end points \((-1, 1)^T\) and \((1, -1)^T\). Therefore, for any orthogonal matrix \( Z \), there exist at most 5 points in \( X \), say \( x_1, \ldots, x_5 \), such that the equations \( h_2 Z_j(x_i) \lambda_1 + \cdots + h_5 Z_j(x_i) \lambda_5 = 1, i = 1, \ldots, 5 \), define supporting hyperplanes of the polytope \( P_Z \) simultaneously. By Corollary 2.1 the optimal designs can be based on the five points in the set (2.17). For example, it is now easy to show that the \( D \)-optimal design for model (2.16) puts equal weights \( 1/5 \) at the points \((-1, 1)^T, (1, -1)^T, (1, 1)^T, (0, 1)^T \) and \((0, 0)^T\).

The following theorem gives an upper bound on the number of support points of an optimal design.

**Theorem 2.2.** Suppose that there exists a number \( m \in \mathbb{N} \), index sets \( G_1, \ldots, G_m \subseteq \mathbb{R} \) and for \( s = 1, \ldots, m \) a collection of subsets \( \{X_g | g \in G_s\} \) of the design space \( X \), such that

\[ X = \bigcup_{s=1}^m \bigcup_{g \in G_s} X_g, \]

where some sets among \( X_g \) may overlap. Assume further that for each \( N \in \text{NND}(k) \) the following conditions are satisfied:

(a) For each set \( X_g \subseteq X \) the function

\[ d_N|_{X_g} : \{X_g \to \mathbb{R} \}
\]

\[ x \to d_N|_{X_g}(x) := d_N(x) = f^T(x)N \mathbf{f}(x) \]

has exactly one maximizer, say \( x_g^* \).

(b) For each \( s = 1, \ldots, m \), the function

\[ d_{s,N}^* : \{G_s \to \mathbb{R} \}
\]

\[ g \to d_{s,N}^*(g) := d_N|_{X_g}(x_g^*) \]

has at most \( k_s \) local maxima.

Then any \( \phi \)-optimal design has at most \( \sum_{s=1}^m k_s \) support points.

**Example 2.3.** To illustrate Theorem 2.2, we consider two examples.

(a) Motivated by an application of a non-linear model (see Section 4.1 for details) we consider the regression model (2.1) with \( k = 2 \) parameters and a vector of regression functions defined by

\[ f(x, \theta) = \left( \frac{1}{2\theta_1^2} (x_1^2 + \theta_2^2)(1 - x_1^2 - x_2^2), -\frac{\theta_2}{\theta_1} (1 - x_1^2 - x_2^2) \right)^T, \]

where the design space is given by the unit disc \( X = \{(x_1, x_2)^T : x_1^2 + x_2^2 \leq 1\} \) and \( \theta_1 \) and \( \theta_2 \) are constants. For \( g \in [-1, 1] \) define

\[ X_g := \left\{ (g, r\sqrt{1-g^2})^T \mid r \in [-1, 1] \right\}, \]
then $X = \bigcup_{g \in [-1,1]} X_g$ (here $m = 1$) and we obtain for a point $x_g \in X_g$

$$f(x_g, \theta) = f(g, r\sqrt{1 - g^2}, \theta) = (1 - g^2 - r^2(1 - g^2)) \left( \frac{g^2 + \theta_2^2}{2\theta_1} - \theta_2 \right)^T.$$

Consequently, for any nonnegative definite matrix $N \in \mathbb{R}^{2 \times 2}$, the function $d_N|_{X_g}$ attains its maximum on $X_g$ for $r = 0$, which gives

$$(2.19) \quad d_N|_{X_g}(x^*_g) = (1 - g^2)^2 P_2(g^2)$$

for some nonnegative quadratic polynomial $P_2$ in $g^2$ with positive leading coefficient. It is now easy to see that this function has at most 2 local maxima in the interval $[-1,1]$, which means that $k_1 = 2$, and by Theorem 2.2 a $\phi$-optimal design for model 2.1 with regression functions given by (2.18) is supported at two points. Moreover, by these arguments it also follows that the support points of optimal designs must be contained in the set $\{(x_1,0)^T \mid x_1 \in (-1,1)\}$ (which correspond to the case $r = 0$ and $|g| \neq 1$).

(b) In order to demonstrate that Theorem 2.2 can also give useful bounds if the optimal designs are not minimally supported, we consider the model in Example 2.1 again, where $f(x) = (x_1, x_2)^T$ and the design space is $X = [0,1]^2$. Define $G_1 = G_2 = [0,1]$ and for $g_1 \in G_1$, $g_2 \in G_2$ consider the sets

$$X_g^{(1)} = \{(x_1, x_2)^T \mid g_1 x_1 - x_2 = 0, 0 \leq x_1 \leq 1\},$$

$$X_g^{(2)} = \{(x_1, x_2)^T \mid x_1 - g_2 x_2 = 0, 0 \leq x_2 \leq 1\}.$$

Obviously the sets $X_g^{(\ell)}$ ($\ell = 1, 2$) are subsets of $X$ and

$$X = \bigcup_{s=1}^2 \bigcup_{g \in G_s} X_g^{(s)}.$$

If $N \in \text{NND}(2)$ and $x \in X_g^{(1)}$ for some $g_1 \in G_1$, we obtain for the function $d_N|_{X_g^{(1)}}$ the representation

$$d_N|_{X_g^{(1)}}(x) = (x_1, x_2)^T N(x_1, x_2)^T = x_1^2 N(1, g_1)^T,$$

which is increasing in $x_1$. Consequently, it follows that the function $d_N|_{X_g^{(1)}}(x)$ is maximized on $X_g^{(1)}$ for $x_{1,g_1}^* = (1, g_1)^T$ and its maximum is given by $d_N^*(g_1) = (1, g_1)^T N(1, g_1)^T$. This function is a quadratic polynomial in $g_1 \in G_1 = [0,1]$ and has at most 2 local maxima at the points $g_1 = 0$ and $g_1 = 1$ corresponding to the points $(1,0)^T$ and $(1,1)^T$ in $X$. The same argument shows that the function $d_N^*(g_2) = (g_2,1)^T N(g_2,1)^T$ has at most 2 local maxima on the set $G_2$ at the points $g_2 = 0$ and $g_2 = 1$ corresponding to the points $(0,1)^T$ and $(1,1)^T$ in $X$. Therefore, by Theorem 2.2, there exists a $\phi$-optimal design with at most 4 support points, which can be further reduced to 3 support points since there is a common point $(1,1)^T$.

3. Admissibility. In this section, we study the relation between admissibility of a design $\xi$ in the model (2.1) and the admissibility of a corresponding “conditional design” of $\xi$ in a “conditional model” of (2.1), which will be defined below. Note that an application of this concept in the context of models with one qualitative and one quantitative factor has been indicated in Schwabe (1996), see pages 107 -109 in this reference. The following discussion contains this situation as a special case.

Throughout this paper we call a design $\xi_1$ admissible if there does not exist any design $\xi_2$ such that $M(\xi_1) \neq M(\xi_2)$ and $M(\xi_2) \geq M(\xi_1)$, that is the matrix $M(\xi_2) - M(\xi_1)$ 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 \( \mathbf{x} = (x_1, x_2)^\top \) and define the function

\[
(3.1) \quad \mu(\mathbf{x}) = \mu(x_1, x_2) = \sum_{j=1}^{k} f_j(x_1, x_2) \theta_j = \mathbf{f}^\top(\mathbf{x}) \mathbf{\theta},
\]

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

\[
(3.2) \quad \tilde{\mu}_t(\mathbf{x}) = \sum_{j=1}^{k_t} \tilde{f}_{jt}(\mathbf{x}) \tilde{\theta}_t = \tilde{\mathbf{f}}_t^\top (\mathbf{x}) \tilde{\mathbf{\theta}}_t, \quad \mathbf{x} \in \tilde{\mathcal{X}}(t),
\]

where \( \tilde{\mathbf{f}}_t(\mathbf{x}) = (\tilde{f}_{1t}(\mathbf{x}), \tilde{f}_{2t}(\mathbf{x}), \ldots, \tilde{f}_{k_t}(\mathbf{x}))^\top \) is a vector of linearly independent regression functions defining a basis of span \{\( f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x}) \)\} under the condition that \( T(\mathbf{x}) = t \), and \( \tilde{\mathbf{\theta}}_t = (\theta_{1t}, \ldots, \theta_{k_t})^\top \) is a \( k_t \)-dimensional 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 \( T_1(\mathbf{x}) = x_1 \) with range denoted by \( \mathcal{T}_1 \), then for fixed \( x_1 \) the set \( \tilde{\mathcal{X}}(x_1) \) can be identified with the set \( \mathcal{X}_2(x_1) := \{x_2 : (x_1, x_2)^\top \in \mathcal{X}\} \) and we obtain the conditional model for the second factor \( x_2 \) on the design space \( \mathcal{X}_2(x_1) \). Moreover, if the set \( \mathcal{X}_2(x_1) \) is the same for all \( x_1 \), then \( \mathcal{X} \) is a Cartesian product. In this case, if the vector \( \tilde{\mathbf{f}}_{x_1}(\mathbf{x}) \) in the conditional model (3.2) does not contain the first factor \( x_1 \) (where the parameter \( \tilde{\mathbf{\theta}}_{x_1} \), may depend on \( x_1 \)), we use the notation \( \mathcal{X}_2 := \mathcal{X}_2(x_1) \), \( \tilde{\mathbf{f}}_2(x_2) := \tilde{\mathbf{f}}_{x_1}(\mathbf{x}) \), \( \tilde{\mathbf{\theta}}_2 := \tilde{\mathbf{\theta}}_{x_1} \), and the resulting model

\[
(3.3) \quad \tilde{\mu}_2(x_2) = \tilde{\mathbf{f}}_2^\top (x_2) \tilde{\mathbf{\theta}}_2, \quad x_2 \in \mathcal{X}_2,
\]

is called the marginal model for the second factor. In this case we will also use the phrase that the marginal model for the second factor exists. We note again that in model (3.3), the parameter vector \( \tilde{\mathbf{\theta}}_{x_1} \) can depend on \( x_1 \) as reflected by our notation. One can similarly define the conditional model given \( t \) and the existence of a marginal model for the first factor.

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

\[
(3.4) \quad \mu(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2
\]
on the design space \( \mathcal{X} = [0, 1]^2 \). Consider the mapping \( T(\mathbf{x}) = x_1 + x_2 \) 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 \( \{1, x_1, x_2, x_1 x_2\} \) 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

\[
\tilde{\mathbf{f}}_t(\mathbf{x}) = (1, x_1, x_1^2)^\top,
\]

where the conditional design space \( \tilde{\mathcal{X}}(t) \) can be identified with the interval \([0, t]\) if \( t \in [0, 1]\) and with the interval \( \tilde{\mathcal{X}}(t) = [t - 1, 1] \) if \( t \in [1, 2] \).

Moreover, the marginal model for the \( i \)-th factor corresponds to the vector of regression functions \( \tilde{\mathbf{f}}_i(x_i) = (1, x_i)^\top \) and the marginal design space is given by \( \mathcal{X}_i = [0, 1], i = 1, 2. \)
For a design $\xi$ on the design space $X$ and a real-valued function $T$ on $X$ with range $T$, let $\xi_T$ be the induced measure of $\xi$ on $T$ which is defined by $\xi_T(B) := \xi(T^{-1}(B))$ for all measurable sets $B \subset X$. Hence, for each $t \in T$

$$\xi_T(t) := \xi_T(\{t\}) = \int_{\tilde{X}(t)} \xi(dx)$$

is the mass of the set $\tilde{X}(t)$ with respect to the measure $\xi$. If $\xi_T(t) > 0$, the design $\xi$ induces a conditional design $\xi_{x|t}$ on the design region $\tilde{X}(t)$ of the conditional model, which is defined by

$$\xi_{x|t}(x) = \frac{1}{\xi_T(t)} \xi(x), \quad x \in \tilde{X}(t).$$

In addition, we define

$$M_t(\xi_{x|t}) = \int_{\tilde{X}(t)} \tilde{f}_t(x)\tilde{f}_t^T(x)\xi_{x|t}(dx)$$

as the information matrix of the design $\xi_{x|t}$ in the conditional model (3.2) and denote by $\Xi_t$ the set of all approximate designs on the design space $\tilde{X}(t)$. 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 $\Xi$ for the regression model (3.1) is that for every $t \in T$ with $\xi_T(t) > 0$ the conditional design $\xi_{x|t}$ induced by $\xi$ is admissible in the class $\Xi_t$ for the conditional model (3.2).

In particular, Theorem 3.1 holds for $T_1(x) = x_1$ and $T_2(x) = x_2$ corresponding to the conditional models for the two factors, respectively. The final result of this section gives a complete subclass and a bound of the number of support points of an admissible design.

**Corollary 3.1.** Assume that $X = X_1 \times X_2 \subset \mathbb{R}^2$ and that the marginal models exist for both factors. Define $\Xi^A$ as the class of admissible designs for the $i$-th marginal model, $T_i(x) = x_i$ ($i = 1, 2$) and denote by $\Xi^C$ the subclass of designs on $X$, where for every $x_i \in X_i$ with $\xi_T(x_i) > 0$ the conditional designs $\xi_{x_i|x_1}$ and $\xi_{x_i|x_2}$ belong to $\Xi^A_2$ and $\Xi^A_1$, respectively. Then the class of all admissible designs for the model (3.1) is a subset of $\Xi^C$.

Moreover, if the number of support points of every design in $\Xi^A_i$ is at most $k_i$, $i = 1, 2$, and there exists for at least one index $i \in \{1, 2\}$ a set of $k_i$ points in $X_i$ such that all admissible designs in $\Xi^A_i$ are supported among these points, then the designs in $\Xi^C$ are based on at most $k_1 k_2$ points.

**Remark 3.1.** By the assumptions made in Section 2 there exists an optimal design with respect to any optimality criterion. As the $D$-optimality criterion is strictly isotonic, it follows from Section 7.8. in Pukelsheim (2006) that there exists at least one admissible design (namely the $D$-optimal design).

**4. Some applications.** In this section we present 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 model with $q$ factors

$$E[y(x)] = \eta(x, \theta), \quad x \in X \subset \mathbb{R}^q,$$
where $y(x)$ 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 $\eta(x, \theta)$ is continuously differentiable with respect to the parameter $\theta = (\theta_1, \ldots, \theta_k)^\top$ and define

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

as the gradient of $\eta$ with respect to the parameter $\theta$. The information matrix of a design $\xi$ for model (4.1) is given by

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

If $n$ observations are taken according to an approximate design (applying an appropriate rounding procedure) it is well known, that under standard assumptions, the covariance matrix of the maximum likelihood estimator of the parameter $\theta$ is approximately given by the matrix $(\sigma^2/n)M^{-1}(\xi, \theta)$ and a locally optimal design maximizes an information function of the matrix $M(\xi, \theta)$. 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 $f(x, \theta)$ defined in (4.2). We illustrate this in a few examples.

4.1. Electrostatic potential model. Patan (2004) studied optimal designs for the electrostatic potential model

$$
(4.4) \quad \eta(x, \theta) = -\frac{1}{2\theta_1}(x_1^2 + \theta_2^2)(1 - x_1^2 - x_2^2), \quad x \in \mathcal{X} = \{(x_1, x_2)^\top : x_1^2 + x_2^2 \leq 1\}.
$$

The gradient of the function $\eta(x, \theta)$ with respect to the parameter $\theta$ in model (4.4) is given by (2.18). Let $T_2$ denote the projection onto the second coordinate and define for $t \in [0,1)$ the matrix

$$
C(t, \theta) = \begin{pmatrix} \theta_2^2/(2\theta_1^2), 1/(2\theta_1^2) \\ -\theta_2/\theta_1, 0 \end{pmatrix},
$$

then the vector of regression functions corresponding to the conditional model of (2.18) for $T_2(x) = t$ is given by

$$
(4.5) \quad \tilde{f}_t(x) = \begin{pmatrix} 1 - x_1^2 - t^2, x_1^2(1 - x_1^2 - t^2) \end{pmatrix}^\top,
$$

where the design space for the conditional model can be identified with

$$
\tilde{\mathcal{X}}(t) = \{x_1 : -\sqrt{1-t^2} \leq x_1 \leq \sqrt{1-t^2}\}.
$$

By the discussion in Example 2.3(a) the support points of an optimal design must have a vanishing second component and therefore it is sufficient to consider the case $t = 0$ for which

$$
(4.6) \quad \tilde{f}_0(x) = \begin{pmatrix} 1 - x_1^2, x_1^2(1 - x_1^2) \end{pmatrix}^\top
$$

and $\tilde{\mathcal{X}}(0) = [-1 \leq x_1 \leq 1]$. As $\tilde{f}_0(x) = \tilde{f}_0(-x)$, it is sufficient to consider only designs on the interval $[0,1]$ (note that these designs correspond to designs on the set $\{(x_1, x_2)^\top \mid x_1^2 + x_2^2 \leq 1; x_1 \in [0,1]\}$). It follows from Yang (2010) that an admissible design for the conditional model (4.6) on the design region $[0,1]$ is supported at two points $x_1 = 0$ and some point $x_1^* \in (0,1)$. Therefore, by Remark 3.1 and Theorem 3.1 there exist admissible designs for model (4.4) supported at two points $(0,0)^\top$ and $(x_1^*,0)^\top$, where $x_1^* \in (0,1)$. For example, Patan (2004) obtained that the design with equal masses at the points $(0,0)^\top$ and $(1/\sqrt{2},0)^\top$ is $D$-optimal for model (4.4).
4.2. Exponential regression models with two factors. Rodríguez, Ortiz and Martínez (2015) considered the maximin optimal design problem for the two-factor exponential growth model

\begin{equation}
\eta(x, \theta) = \theta_0 + \exp(-\theta_1 x_1) + \exp(-\theta_2 x_2), \tag{4.7}
\end{equation}

(\theta_j \geq 1, j = 1, 2) on the square \( X = [0, 1]^2 \), which has numerous applications in biological and agricultural sciences. In this model the gradient of the function \( \eta(x, \theta) \) in (4.7) is given by

\begin{equation}
f(x, \theta) = (1, -x_1 \exp(-\theta_1 x_1), -x_2 \exp(-\theta_2 x_2))^T \tag{4.8}
\end{equation}

and the two vectors of regression functions corresponding to the marginal models of (4.8) are obtained as

\begin{equation}
\tilde{f}_i(x_i, \theta) = (1, x_i \exp(-\theta_i x_i))^T, \ i = 1, 2. \tag{4.9}
\end{equation}

Note that for every \( x_i \in (0, 1) \) the vector \((1, x_i \exp(-\theta_i x_i))^T\) can be represented as a convex combination of the vectors \((1, 0)^T\) and \((1, \exp(-1)/\theta_i)^T\) corresponding to the points \( x_i = 0 \) and \( x_i = 1/\theta_i \), respectively. Therefore, the admissible designs for the marginal models are supported at the points \( \{0, 1/\theta_i\} \) (i = 1, 2) (see Pukelsheim, 2006, Theorem 8.5), and it now follows from Theorem 3.1 and Corollary 3.1 that the admissible designs for model (4.7) are contained in the class of all designs supported on at most 4 points from the set \( \{(0, 0)^T, (0, 1/\theta_2)^T, (1/\theta_1, 0)^T, (1/\theta_1, 1/\theta_2)^T\} \). 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

\begin{equation}
\eta(x, \theta) = \theta_0 \exp(\theta_1 x_1 + \theta_2 x_2), \tag{4.10}
\end{equation}

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

The gradient of the function \( \eta(x, \theta) \) in model (4.10) is given by

\begin{equation}
f(x, \theta) = \exp(\theta_1 x_1 + \theta_2 x_2)(1, \theta_0 x_1, \theta_0 x_2)^T. \tag{4.11}
\end{equation}

Let \( T(x) = \theta_1 x_1 + \theta_2 x_2 \) and define the matrix

\[ C(t, \theta) = \exp(t) \begin{pmatrix}
1 & 0 & 0 \\
0 & \theta_0 & 0 \\
0 & 0 & \theta_0
\end{pmatrix}, \]

then the vector of regression functions corresponding to the conditional model of (4.11) is given by

\begin{equation}
\tilde{f}_i(x) = (1, x_1, x_2)^T \tag{4.12}
\end{equation}

and the design space for the conditional model is given by \( \tilde{X}(t) = \{x \in X : \theta_1 x_1 + \theta_2 x_2 = t\} \). For every \( t \in T = \{T(x) : x \in X\} \), it is easy to see that admissible designs for the conditional model (4.12) on the design region \( \tilde{X}(t) \) are supported at the two end points of the line segment \( \ell = \{(x_1, x_2)^T \in X : \theta_1 x_1 + \theta_2 x_2 = t\} \). Therefore, by Theorem 3.1, admissible designs for the model (4.10) are supported on the boundary of the design region \( X \).

Moreover, for the boundary corresponding to \( x_1 = 0 \) and \( x_1 = b_1 \) the conditional model for the second factor \( x_2 \) on the design space \( X_2 = [0, b_2] \) is

\begin{equation}
\tilde{f}_{x_1}(x_2, \theta) = \exp(\theta_2 x_2)(1, x_2)^T, \tag{4.13}
\end{equation}

\begin{equation}
\tilde{f}_{x_1}(x_2, \theta) = \exp(\theta_2 x_2)(1, x_2)^T, \tag{4.13}
\end{equation}
and for the boundary corresponding to \( x_2 = 0 \) and \( x_2 = b_2 \) the conditional model for the first factor \( x_1 \) on the design space \( \mathcal{X}_1(x_2) = [0, b_1] \) is

\[
(4.14) \quad f_{\tilde{x}_2}(x_1, \theta) = \exp(\theta_1 x_1) (1, x_1)^\top.
\]

The admissible designs for the \( i \)-th conditional model are supported at one point or two points, one of which is \( b_i \) (\( i = 1, 2 \)) (see Yang and Stufken, 2009, Theorem 4). It now follows from Theorem 3.1 that the admissible designs for model (4.10) have at most 8 support points. Moreover, by the same result, it follows that the admissible designs are contained in the class of all designs supported on at most 5 points, of the form \( (b_1, b_2)^\top, (b_1, v_2)^\top, (u_2, b_2)^\top, (u_1, 0)^\top, (0, v_1)^\top \) where \( u_1, u_2 \in [0, b_1) \) and \( v_1, v_2 \in [0, b_2) \). For example, if there are two support points on each of the two edges \( \{(0, x_2)^\top : x_2 \in [0, b_2]\} \) and \( \{(x_1, 0)^\top : x_1 \in [0, b_1]\} \), then two of these four points are \( (b_2, b_2)^\top \) and \( (b_1, 0)^\top \). As these points lie also on the edges \( \{(b_1, x_2)^\top : x_2 \in [0, b_2]\} \) and \( \{(x_1, b_2)^\top : x_1 \in [0, b_1]\} \), there is at most one additional support point on the edge \( \{(x_1, b_2)^\top : x_1 \in [0, b_1]\} \), which must be the point \( (b_1, b_2)^\top \). However, this point is also a point on the edge \( \{(b_1, x_2)^\top : x_2 \in [0, b_2]\} \). Thus, there are at most 5 support points. The other cases are treated similarly.

4.3. Mixture of exponentials and polynomials. Rodríguez, Ortiz and Martínez (2015) considered the maximin optimal design problem for the model

\[
(4.15) \quad \eta(x, \theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^3 + \exp(-\theta_3 x_2),
\]

which was used in Langseth et al. (2012) for approximating the potentials associated with general hybrid Bayesian networks. The parameter \( \theta_3 \) is assumed to be positive, the design space is given by \( \mathcal{X} = [-1, 1] \times [0, 2] \) and the gradient of the function \( \eta(x, \theta) \) in (4.15) is obtained as

\[
(4.16) \quad f(x, \theta) = (1, x_1, x_1^3, -x_2 \exp(-\theta_3 x_2))^{\top}.
\]

The marginal models of (4.16) exist with corresponding vectors of regression functions given by

\[
(4.17) \quad \tilde{f}_1(x_1, \theta) = (1, x_1, x_1^3)^\top, \quad \mathcal{X}_1 = [-1, 1]
\]

\[
(4.18) \quad \tilde{f}_2(x_2, \theta) = (1, x_2 \exp(-\theta_3 x_2))^\top, \quad \mathcal{X}_2 = [0, 2].
\]

The admissible designs for the marginal model (4.17) have at most 4 support points including the end points \(-1\) and \(1\) (see Yang, 2010, Theorem 8). In addition, it follows from the discussion on invariance in Chapter 13 of Pukelsheim (2006) that for the \( \phi_0 \)-optimality criterion defined in (2.3) the other two (potential) support points, say \( u^* \) and \( v^* \), satisfy \( u^* = -v^* \). For the marginal model (4.18) the admissible designs are supported at the points \( \{0, x_2^b\} \), where \( x_2^b = \min\{1/\theta_3, 2\} \). It now follows from Corollary 3.1 that the admissible designs for model (4.15) are contained in the class of all designs supported at 8 points,

\[
(\pm 1, 0), (\pm 1, x_2^b), (\pm u^*, 0), (\pm v^*, x_2^b),
\]

where \( u^*, v^* \in [0, 1] \). For example, the locally \( D \)-optimal design for model (4.15) is equally supported at the points \((\pm 1, 0)^\top, (\pm 1/\sqrt{3}, 0)^\top, (\pm 1, x_2^b)^\top \) and \((\pm 1/\sqrt{3}, x_2^b)^\top \).

5. Conclusions. Deriving optimal designs for multi-factor experiments is a difficult problem. A common approach for the solution of the optimization problem is the application of equivalence theorems which require the specification of an optimality criterion. Consequently, for different criteria this method can only work in a case-by-case way.
In contrast, the approach presented in this paper is applicable to various criteria simultaneously. It provides sufficient conditions for the occurrence of the de la Garza phenomenon and new criteria for admissibility of designs in multi-factor models. As we have illustrated in Section 4, it can simplify the design problem by identifying a complete subclass, which is composed of relatively simple designs. For any design not belonging to this class, there exists a design in the class with a dominating information matrix with respect to the Loewner ordering. As a result, the determination of optimal designs can be restricted to a much smaller class.

On the other hand, serval numerical algorithms have been developed for searching optimal designs. For example, recently, Masoudi, Holling and Wong (2017) applied the imperialist competitive algorithm to find minimax and standardized maximin optimal designs. Zhang, Wong and Tan (2020) used particle swarm optimization to determine optimal designs for nonlinear regression models with multiple interacting factors. Usually, the number of support points has to be specified in advance when running these algorithms. Therefore, a tight upper bound on the number of support points of optimal designs yields large computational advances. The results of this paper can provide some tools to derive such bounds in multi-factor experiments.

APPENDIX: PROOFS

A.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 \( \phi^\infty : \text{NND}(k) \to [0; \infty) \) of an information function \( \phi : \text{PD}(k) \to (0, \infty) \) is defined by

\[
\phi^\infty(D) = \inf_{C \in \text{PD}(k)} \frac{\text{trace}(CD)}{\phi(C)}.
\]

For every information function function \( \phi \) the corresponding polar function \( \phi^\infty \) is isotonic relative to the Loewner ordering. Define

\[
\mathcal{N} = \{ N \in \text{NND}(k) : f^\top(x)Nf(x) \leq 1 \ \forall x \in \mathcal{X} \},
\]

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

\[
\max_{M \in \mathcal{M}(\Xi)} \phi(M) = \min_{N \in \mathcal{N}} \frac{1}{\phi^\infty(N)}.
\]

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

\[
\phi(M) = \frac{1}{\phi^\infty(N)},
\]

and two matrices \( M \in \mathcal{M}(\Xi) \) and \( N \in \mathcal{N} \) satisfy (A.4) if and only if the conditions

\[
\text{trace}(MN) = 1,
\]

\[
\phi(M)\phi^\infty(N) = \text{trace}(MN)
\]

hold. An application of this result yields the famous general equivalence theorem in optimal design theory.
THEOREM A.1 (Pukelsheim (2006), Theorem 7.17). A positive definite information matrix $M^* \in M(\Xi)$ is $\phi$-optimal for $\theta$ in $M(\Xi)$ if and only if there exists a nonnegative definite $k \times k$ matrix $N \in \mathcal{N}$ that solves the polarity equation

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

and that satisfies the normality inequality

$$(A.7) \quad f^\top(x)NF(x) \leq 1 \forall x \in \mathcal{X}.$$ 

Moreover, if $M^*$ is optimal for $\theta$ in $\Xi$, there is equality for any support point $x_i$ of any $\phi$-optimal design $\xi \in \Xi$, that is any design with $M^* = M(\xi)$.

A.2. Proof of Theorem 2.1. The matrix $N$ in $\text{NND}(k)$ has an eigenvalue decomposition

$$N = Z_N \Lambda_N Z_N^\top,$$

where $\Lambda_N = \text{diag}(\lambda_{N1}, \ldots, \lambda_{Nk})$ is a diagonal matrix containing the ordered eigenvalues $\lambda_{N1} \geq \lambda_{N2} \ldots \geq \lambda_{Nk}$ (counted with their respective multiplicities) of the matrix $N$ and $Z_N = (z_{N1}, \ldots, z_{Nk}) \in \mathbb{R}^{k \times k}$ is an orthogonal matrix with eigenvectors corresponding to the eigenvalues. Denote by $S_Z$ the subset of $\text{NND}(k)$ consisting of matrices which permit an eigenvalue decomposition with the same orthogonal matrix $Z$, i.e.,

$$S_Z = \{N: N = Z\Lambda Z^\top, \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k), \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k\}.$$ 

Then we can express $\text{NND}(k)$ as

$$\text{NND}(k) = \bigcup_{Z \in O(k)} S_Z,$$

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

$$h_Z(x) = Z^\top f(x) = (z_1^\top f(x), \ldots, z_k^\top f(x))^\top = (h_{Z1}(x), \ldots, h_{Zk}(x))^\top,$$

then the set $\mathcal{N}$ in (A.2) can be represented as

$$\mathcal{N} = \bigcup_{Z \in O(k)} \{N \in S_Z: f^\top(x)NF(x) \leq 1 \forall x \in \mathcal{X}\}$$

$$= \bigcup_{Z \in O(k)} \{N \in S_Z: h_Z^\top(x)\Lambda h_Z(x) \leq 1 \forall x \in \mathcal{X}\}$$

$$= \bigcup_{Z \in O(k)} \{N \in S_Z: h_{Z1}^2(x)\lambda_1 + \cdots + h_{Zk}^2(x)\lambda_k \leq 1 \forall x \in \mathcal{X}\}$$

$$\cong \bigcup_{Z \in O(k)} \mathcal{N}_Z,$$

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

$$\mathcal{P}_Z = \{\lambda = (\lambda_1, \ldots, \lambda_k)^\top \in \mathbb{R}_+^k : h_{Z1}^2(x)\lambda_1 + \cdots + h_{Zk}^2(x)\lambda_k \leq 1 \forall x \in \mathcal{X}\}.$$ 

Note that the polar function $\phi^\infty$ is isotonic, hence the optimal solution must be attained at a boundary point of the polytope. Let $\lambda^* = (\lambda_1^*, \ldots, \lambda_k^*)^\top \in \mathbb{R}_+^k$ be the vector corresponding to the extremum. We define $\mathcal{N}^* = Z^*\Lambda^* Z^*^\top$ with $\Lambda^* = \text{diag}(\lambda_1^*, \ldots, \lambda_k^*)$, then $\mathcal{N}^*$ is the optimal matrix of the dual problem. By Theorem A.1, we have

$$1 = f^\top(x_i^*)\mathcal{N}^* f(x_i^*) = h_{Z1}^2(x_i^*)\lambda_1^* + \cdots + h_{Zk}^2(x_i^*)\lambda_k^*$$

for any support point $x_1^*, \ldots, x_m^*$ of the $\phi$-optimal design $\xi^*$. This implies that the rank of the $(k + 1) \times m$ matrix

$$P_m = \{(h_{Z1}^2(x_i^*), \ldots, h_{Zk}^2(x_i^*), 1)^\top, i = 1, \ldots, m\}$$
in (2.9) is at most \( k \) and the equations \( h_Z^2(x_1^*) \lambda_1 + \cdots + h_Z^2(x_k^*) \lambda_k = 1 \) define supporting hyperplanes of the polytope \( \mathcal{P}_Z \) defined by (2.7). Moreover, if there exists two support points, say \( x_i^* \) and \( x_j^* \), corresponding to the same supporting hyperplane, i.e.,
\[
(h_Z^2(x_1^*), \ldots, h_Z^2(x_k^*))^\top = (h_Z^2(x_j^*), \ldots, h_Z^2(x_k^*))^\top,
\]
then we have \( \| f(x_j^*) \|^2 = \| f(x_j^*) \|^2 \) since
\[
\| f(x) \|^2 = \| Z^\top f(x) \|^2 = \| h_Z(x) \|^2,
\]
which completes the proof of Theorem 2.1.

**A.3. Proof of Corollary 2.1.** Part (a) is a direct consequence of Theorem 2.1. For part (b), note that in the case of the existence of a \( \phi \)-optimal design with \( m \geq k + 1 \) support points, Theorem 2.1 implies that the rank of the matrix \( P_m \) in (2.9) is at most \( k \). However, by the assumption in part (b) this rank must be at least \( k + 1 \), which yields a contradiction. Hence there exists no \( \phi \)-optimal design with \( m \geq k + 1 \) support points.

**A.4. Proof of Theorem 2.2.** Let \( x^* \) be a maximizer of \( d_N(x) \) in \( \mathcal{X} \), then, by assumption, \( x^* \) must be the only maximizer of \( d_N|_{\mathcal{X}_g^*}(x) \) in some subset \( \mathcal{X}_g^* \), i.e. \( x^* = x_g^* \). for some \( g^* \in \mathcal{G}_s^* \). Consequently, we have
\[
d_N(x^*) = \max_{x \in \mathcal{X}_g^*} d_N(x) = d_N|_{\mathcal{X}_g^*}(x_g^*) = d_{s^*N}(g^*) .
\]
Therefore \( g^* \) is a maximizer of the function \( d_{s^*N} \) on \( \mathcal{G}_s^* \). However, by condition (b), this function has at most \( k_s \) local maxima on \( \mathcal{G}_s^* \). Consequently, it follows that the number of maxima of the function \( d_N(x) \) on \( \mathcal{X} \) is at most \( \sum_{s=1}^{m} k_s \) and an application of Theorem A.1 yields the assertion of Theorem 2.2.

**A.5. Proof of Theorem 3.1.** For fixed \( t \), there exists a \( k \times k_t \) matrix, say \( C(t) \), such that
\[
f(x) = C(t) \bar{f}_t(x)
\]
on the design space \( \tilde{X}(t) \). Since the elements of the \( k_t \)-dimensional vector \( \bar{f}_t(x) \) are linearly independent and define a basis of \( \text{span}\{f_1(x), \ldots, f_k(x)\} \) on \( \tilde{X}(t) \), the rank of \( C(t) \) is \( k_t \). Suppose there exists some \( t_* \in \mathcal{T} \) with \( \xi_T(t_*) > 0 \) such that the conditional design \( \xi_{x|t_*} \) is inadmissible for the conditional model (3.2). Then there exists a design \( \tilde{\xi}_{x|t_*} \) in the set of all conditional designs \( \Xi_{t_*} \) satisfying
\[
M_{t_*}(\xi_{x|t_*}) \geq M_{t_*}(\tilde{\xi}_{x|t_*}) \quad \text{and} \quad M_{t_*}(\xi_{x|t_*}) \neq M_{t_*}(\tilde{\xi}_{x|t_*}) .
\]
We now define a design \( \tilde{\xi} \) on \( \mathcal{X} \) by
\[
\tilde{\xi}(x) = \begin{cases} \xi(x) & \text{if } x \in \mathcal{X} \setminus \tilde{X}(t_*), \\ \xi_{x|t_*}(x) \xi_T(t_*) & \text{if } x \in \tilde{X}(t_*). \end{cases}
\]
Roughly speaking, \( \tilde{\xi} \) is obtained by replacing the conditional design \( \xi_{x|t_*} \) of \( \xi \) with \( \tilde{\xi}_{x|t_*} \). Then we have \( \tilde{\xi}_T = \xi_T \) and (in the following calculation the symbol \( \leq \) refers to the Loewner ordering and all integrals reduce to finite sums, because all measures have a finite support)
\[
M(\xi) = \int_{\mathcal{X}} f(x) f^\top(x) \xi(dx) = \int_{\mathcal{T}} \int_{\tilde{X}(t)} f(x) f^\top(x) \xi_{x|t}(dx) \xi_T(dt)
\]
Therefore, the design $\xi$ would be inadmissible in the class $\Xi$ for the model (3.1), and this contradiction completes the proof of Theorem 3.1.

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). The authors would like to thank the referees and the associate editor for their constructive comments on three earlier versions of the paper and Martina Stein, who typed parts of the paper with considerable technical expertise. All authors contributed equally to the manuscript.

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.

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 1635–1656.

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

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

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 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 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. G., HAINES, L. M. 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). *Tchebyseff 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.

MASoudi, E., HOLLING, H. and WONG, W. K. (2017). Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs. *Computational Statistics & Data Analysis* **113** 330-345.

PATAN, M. (2004). Optimal observation strategies for parameter estimation of distributed systems, *PhD Thesis.* University of Zielona Góra, Zielona Góra.

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

RANDALL, T., DONEV, A. and ATKINSON, A. C. (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.

SCHWABE, R. (1996). *Optimum Designs for Multi-factor Models, Lecture Notes in Statistics, vol. 113.* Springer, New York.

YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 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.

ZHANG, Z. Z., WONG, W. K. and TAN, K. C. (2020). Competitive swarm optimizer with mutated agents for finding optimal designs for nonlinear regression models with multiple interacting factors. *Memetic Computing* **12** 219-233.