Complete classes of designs for nonlinear regression models and principal representations of moment spaces

Holger Dette, Kirsten Schorning
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

November 16, 2012

Abstract

In a recent paper Yang and Stufken (2012a) gave sufficient conditions for complete classes of designs for nonlinear regression models. In this note we demonstrate that their result is a simple consequence of the fact that boundary points of moment spaces generated by Chebyshev systems possess unique representations.

Keywords and Phrases: locally optimal design; admissible design; Chebyshev system; principle representations; moment spaces; complete classes of designs
AMS Subject Classification: 62K05

1 Introduction

The construction of locally optimal designs for nonlinear regression models has found considerable interest in recent years [see for example He et al. (1996), Dette et al. (2006), Khuri et al. (2006), Fang and Hedayat (2008), Yang and Stufken (2012b) among others]. While most of the literature focuses on specific models or specific optimality criteria, general results characterizing the structure of locally optimal designs are extremely difficult to obtain due to the complicated structure of the corresponding nonlinear optimization problems. In a series of remarkable papers Yang and Stufken (2009), Yang (2010), Dette and Melas (2011) and Yang and Stufken (2012a) derived several complete classes of designs with respect to the Loewner Ordering of the information matrices. The first paper in this direction of Yang and Stufken (2009) investigates nonlinear regression models with two parameters. These results were generalized by Yang (2010) and Dette and Melas (2011) to identify small complete classes for nonlinear regression models with more than two parameters. The most general contribution is the recent paper of Yang and Stufken (2012a), which provides a sufficient condition for a complete class of designs and is applicable to most of the commonly used regression models. The proof of this statement is complicated and requires several auxiliary results.

The purpose of the present paper is to demonstrate that conditions of this type are intimately related to the characterization of boundary points of moment spaces associated with a nonlinear regression model. Our main tool is a Chebyshev system [Karlin and Studden (1966)] appearing in (a transformation of) the Fisher Information matrix of a given design. The complete class of designs can essentially be characterized as the set of measures corresponding to the lower and upper principal representation of the boundary points of the corresponding moment spaces. With this insight the main result in the paper of Yang and Stufken (2012a) is a simple consequence of the fact that a representation of a boundary point of a $k+1$-dimensional moment space associated with a Chebyshev system depends only on the first k functions which are used to generate the moment space.

In Section 2 we present some facts on moment spaces associated with Chebyshev systems which are of general interest for constructing admissible designs. The design problem and Theorem 1 of Yang and Stufken (2012a) are stated in Section 3, where we also present our alternative proof. We finally note that the paper of Yang and Stufken (2012a) contains numerous interesting examples and provided a further result which are not discussed in this note for the sake of brevity.

2 Chebyshev systems and associated moment spaces

A set of k real valued functions $\Psi_0, \ldots, \Psi_{k-1} : [A, B] \to \mathbb{R}$ is called Chebychev system on the interval $[A, B]$ if and only if it fulfills the inequality

$$\det \begin{pmatrix} \Psi_0(x_0) & \ldots & \Psi_0(x_{k-1}) \\ \vdots & \ddots & \vdots \\ \Psi_{k-1}(x_0) & \ldots & \Psi_{k-1}(x_{k-1}) \end{pmatrix} > 0$$

for any points $x_0, \ldots, x_{k-1}$ with $A \leq x_0 < x_1 \ldots < x_{k-1} \leq B$. The moment space associated with a Chebychev system is defined by

$$\mathcal{M}_{k-1} = \Big\{ c = (c_0, \ldots, c_{k-1})^T \;\Big|\; c_i = \int_A^B \Psi_i(x) d\sigma(x)\,, i = 0, \ldots, k-1\,, \sigma \in \mathbb{P}([A, B]) \Big\}, \quad (2.1)$$

where $\mathbb{P}([A, B])$ denotes the set of all finite measures on the interval $[A, B]$. It can be characterized as the smallest convex cone containing the curve

$$\mathcal{C}_{k-1} = \Big\{ (\Psi_0(t), \ldots, \Psi_{k-1}(t))^T \;\Big|\; t \in [A, B] \Big\}$$

[see Karlin and Studden (1966)]. By Caratheodory's theorem any point of $\mathcal{M}_{k-1}$ can be described as a linear combination of at most $k+1$ points in $\mathcal{C}_{k-1}$, where the coefficients are positive. Moment spaces can be defined for any set of linearly independent functions, but if the functions $\{\Psi_0, \dots, \Psi_{k-1}\}$ generate a Chebychev system, the moment space has several additional interesting properties. In particular, less points of $\mathcal{C}_{k-1}$ are required for the representation of points in $\mathcal{M}_{k-1}$. To be precise, we define for a point $c^0 \in \mathcal{M}_{k-1}$ its index $I(c^0)$ as the minimal number of points in $\mathcal{C}_{k-1}$ which are required to represent c^0, where the points $(\Psi_0(A), \dots, \Psi_{k-1}(A))^T$ and $(\Psi_0(B), \dots, \Psi_{k-1}(B))^T$ corresponding to the boundary of the interval $[A, B]$ are counted by $1/2$. The index of a finite measure σ on $[A, B]$ is defined as the index of the point $c = \int_A^B (\Psi_0(x), \dots, \Psi_{k-1}(x))^T d\sigma(x)$. The measure σ is also called representation of the point $c \in \mathcal{M}_{k-1}$.
With this convention it follows that the point $c^0 \in \mathcal{M}_{k-1}$ is a boundary point of $\mathcal{M}_{k-1}$ if and only if its index satisfies $I(c^0) < \frac{k}{2}$. Similarly, c^0 is in the interior of $\mathcal{M}_{k-1}$ if its index is $\frac{k}{2}$ or $\frac{k+1}{2}$. Following Karlin and Studden (1966) we denote a representation of an interior point c^0 as *principal*, if $I(c^0) = \frac{k}{2}$. These authors proved that for each interior point $c^0 \in \mathcal{M}_{k-1}$ there exist exactly two principal representations. The first is called *upper* principal representation and contains the point $(\Psi_0(B), \dots, \Psi_{k-1}(B))^T$ corresponding to the right boundary of the interval $[A, B]$, whereas the second is called *lower* principal representation and does not use this point. The corresponding measures are denoted by σ^+ and σ^-. If k is odd the lower and upper principal representation have $\frac{k+1}{2}$ support points. On the other hand, if k is even the lower and upper principal representation have $\frac{k}{2}$ and $\frac{k+2}{2}$ support points respectively. The following Lemma is crucial in the following investigations and a direct consequence of the discussion on page 55-56 in Karlin and Studden (1966).

Lemma 2.1 *Let $\Psi_j : [A, B] \to \mathbb{R}$ $(j = 0, \dots, k-1)$; $\Omega : [A, B] \to \mathbb{R}$ denote real valued functions and assume that the systems $\{\Psi_0, \dots, \Psi_{k-1}\}$ and $\{\Psi_0, \dots, \Psi_{k-1}, \Omega\}$ are Chebychev systems on the interval $[A, B]$. If $c^0 = (c_1^0, \dots, c_{k-1}^0)^T \in \mathcal{M}_{k-1}$, then the upper and lower principal representation σ^+ and σ^- of c^0 are uniquely determined and satisfy*

$$\max \left\{ \int_A^B \Omega(t) d\sigma(t) \Big| \; \sigma \in \mathbb{P}([A, B]),\, c_i^0 = \int_A^B \Psi_i(t) d\sigma(t), i = 0, \dots, k-1 \right\} = \int_A^B \Omega(t) d\sigma^+(t),$$

$$\min \left\{ \int_A^B \Omega(t) d\sigma(t) \Big| \; \sigma \in \mathbb{P}([A, B]),\, c_i^0 = \int_A^B \Psi_i(t) d\sigma(t), i = 0, \dots, k-1 \right\} = \int_A^B \Omega(t) d\sigma^-(t).$$

In particular both representation do not depend on the function $\Omega : [A, B] \to \mathbb{R}$.

3 A complete class of designs for regression models

Consider the common non linear regression model

$$E[Y|x] = \eta(x, \theta) \tag{3.1}$$

where $\theta \in \mathbb{R}^p$ is the vector of unknown parameters, x denotes a real valued covariate from the design space $[A, B] \subset \mathbb{R}$ and different observations are assumed to be independent with variance σ^2. The function η is called regression function [see Seber and Wild (1989) or Ratkowsky (1990)] and assumed to be continuous and differentiable with respect to the variable θ. A design is defined as a probability measure ξ on the interval $[A, B]$ with finite support [see Kiefer (1974)]. If the design ξ has masses w_i at the points $x_i\,(i = 1, \ldots, l)$ and n observations can be made by the experimenter, this means that the quantities $w_i n$ are rounded to integers, say n_i, satisfying $\sum_{i=1}^{l} n_i = n$, and the experimenter takes n_i observations at each location x_i $(i = 1, \ldots, l)$. If the design ξ contains l support points $x_1, \ldots, x_l$ such that the vectors $\frac{\partial}{\partial\theta}\eta(x_1, \theta), \ldots, \frac{\partial}{\partial\theta}\eta(x_l, \theta)$ are linearly independent and observations are taken according to this procedure it follows from Jennrich (1969) that the covariance matrix of the non-linear least squares estimator is approximately (if $n \to \infty$) given by

$$\frac{\sigma^2}{n} M^{-1}(\xi, \theta) = \frac{\sigma^2}{n} \left(\int_A^B \left(\frac{\partial}{\partial\theta}\eta(x, \theta) \right) \left(\frac{\partial}{\partial\theta}\eta(x, \theta) \right)^T d\xi(x) \right)^{-1}, \tag{3.2}$$

An optimal design maximizes an appropriate functional of the matrix $\frac{n}{\sigma^2} M(\xi, \theta)$ and numerous criteria have been proposed in the literature to discriminate between competing designs [see Pukelsheim (2006)]. Note that the matrix (3.2) depends on the unknown parameter θ and following Chernoff (1953) we call the maximizing designs locally optimal designs. These designs require an initial guess of the unknown parameters in the model and are used as benchmarks for many commonly used designs or for the construction of more sophisticated optimality criteria which require less information regarding the parameters of the model [Chaloner and Verdinelli (1995) and Dette (1997)].

Most of the available optimality criteria are positively homogeneous, that is $\Phi\big(\frac{n}{\sigma^2} M(\xi, \theta)\big) = \frac{n}{\sigma^2}\Phi(M(\xi, \theta))$ [Pukelsheim (2006)]. Therefore it is sufficient to consider maximization of functions of the matrix $M(\xi, \theta)$, which is called information matrix in the literature. Moreover, the commonly used optimality criteria also satisfy a monotonicity property with respect to the Loewner ordering, that is $\Phi(M(\xi_1, \theta)) \geq \Phi(M(\xi_2, \theta))$, whenever $M(\xi_1, \theta) \geq M(\xi_2, \theta)$, where the parameter θ is fixed, ξ_1, ξ_2 are two competing designs on the interval $[A, B]$ and Φ denotes an information function in the sense of Pukelsheim (2006). Throughout this paper we call a design ξ admissible if there does not exist any design ξ_1, such that $M(\xi_1, \theta) \neq M(\xi, \theta)$ and

$$M(\xi_1, \theta) \geq M(\xi, \theta). \tag{3.3}$$

Yang and Stufken (2012a) derive a complete class theorem in this general context which characterizes the class of designs, which cannot be improved with respect to the Loewner ordering of their information matrices. For the sake of completeness and because of its importance we will state this result here again. In particular, we demonstrate that the complete class specified by these authors corresponds to upper and principal representations of a moment space generated by the regression functions. For this purpose we denote by $P(\theta)$

a regular $p \times p$ matrix, which does not depend on the design ξ, such that the representation

$$M(\xi, \theta) = P(\theta) C(\xi, \theta) P^T(\theta) \tag{3.4}$$

holds, where the $p \times p$ matrix $C(\xi, \theta)$ is defined by

$$C(\xi, \theta) \quad = \quad \int_A^B \begin{pmatrix} \Psi_{11}(x) & \dots & \Psi_{1p}(x) \\ \vdots & \ddots & \vdots \\ \Psi_{p1}(x) & \dots & \Psi_{pp}(x) \end{pmatrix} d\xi(x) \;=\; \int_A^B \begin{pmatrix} C_{11}(x) & C_{21}^T(x) \\ C_{21}(x) & C_{22}(x) \end{pmatrix} d\xi(x)$$

and $C_{11}(x) \in \mathbb{R}^{p-p_1 \times p-p_1}$, $C_{21}(x) \in \mathbb{R}^{p_1 \times p-p_1}$, $C_{22}(x) \in \mathbb{R}^{p_1 \times p_1}$ are appropriate block matrices ($1 \leq p_1 \leq p$). Obviously, $P(\theta)$ could be chosen as identity matrix, but in concrete applications other choices might be advantageous [see Yang and Stufken (2012b), Section 4, for numerous interesting examples]. A similar comment applies to the choice of p_1 which is used to represent the matrix C in a 2×2 block matrix. Note that the inequality (3.3) is satisfied if and only if the inequality

$$C(\xi_1, \theta) \geq C(\xi, \theta) \tag{3.5}$$

holds. Following Yang and Stufken (2012a) we define $\Psi_0(x) = 1$, denote the different elements among $\{\Psi_{ij} | 1 \leq i \leq p,\, j \leq p - p_1\}$ in the matrices $C_{11}(x)$ and $C_{21}(x)$ which are not constant by $\Psi_1, \dots, \Psi_{k-1}$ and define for any vector $Q \in \mathbb{R}^{p_1} \setminus \{0\}$ the function

$$\Psi_k^Q(x) = Q^T C_{22}(x) Q. \tag{3.6}$$

We are now in apposition to state and prove the main result of this paper.

Theorem 3.1 [Yang and Stufken (2012a)]

1. *If $\{\Psi_0, \dots, \Psi_{k-1}\}$ and $\{\Psi_0, \dots, \Psi_{k-1}, \Psi_k^Q\}$ are Chebychev systems for every non-zero vector Q, then for any design ξ there exists a design ξ^+ with at most $\frac{k+2}{2}$ support points, such that $M(\xi^+, \theta) \geq M(\xi, \theta)$.*
   If the index of ξ satisfies $I(\xi) < \frac{k}{2}$, then the design ξ^+ is uniquely determined in the set
   $$\Big\{\eta \Big| \int_A^B \Psi_i(x) d\eta(x) = \int_A^B \Psi_i(x) d\xi(x),\, i = 1, \dots, k-1\Big\} \tag{3.7}$$
   and coincides with the design ξ.

   If the index of ξ satisfies $I(\xi) \geq \frac{k}{2}$, then the following cases are discriminated:

   (a) If k is odd, then the design ξ^+ has at most $\frac{k+1}{2}$ support points and it can be chosen such that B is a support point of the design ξ^+.

   (b) If k is even, then the design ξ^+ has at most $\frac{k+2}{2}$ support points and it can be chosen such that A and B are support points of the design ξ^+.

2. *If* $\{\Psi_0, \ldots, \Psi_{k-1}\}$ *and* $\{\Psi_0, \ldots, \Psi_{k-1}, -\Psi_k^Q\}$ *are Chebychev systems for every non-zero vector* Q*, then for any design* ξ *there exists a design* ξ^- *with at most* $\frac{k+1}{2}$ *support points, such that* $M(\xi^-, \theta) \geq M(\xi, \theta)$.
   If the index of ξ *satisfies* $I(\xi) < \frac{k}{2}$*, then the design* ξ^- *is uniquely determined in the set of measures satisfying* (3.7) *and coincides with the design* ξ.
   If the index of ξ *satisfies* $I(\xi) \geq \frac{k}{2}$*, then the following cases are discriminated:*
   (a) *If* k *is odd, then the design* ξ^- *has at most* $\frac{k+1}{2}$ *support points and it can be chosen such that* A *is a support point of the design* ξ^-.
   (b) *If* k *is even, then the design* ξ^- *has at most* $\frac{k}{2}$ *support points.*

Proof. We only present the proof of the first part of the theorem, the second part follows by similar arguments. Yang and Stufken (2012a) showed that a design ξ_1 satisfies (3.3) if the conditions

$$\begin{aligned} \int_A^B \Psi_i(x) d\xi_1(x) &= \int_A^B \Psi_i(x) d\xi(x) \quad i = 1, \ldots, k-1 \\ \int_A^B \Psi_k^Q(x) d\xi_1(x) &\geq \int_A^B \Psi_k^Q(x) d\xi(x) \end{aligned} \tag{3.8}$$

are satisfied for all vectors $Q \neq 0$. Consequently an improvement of the design ξ is obtained by maximizing the "k-th moment" $\int_A^B \Psi_k^Q(x) d\xi_1(x)$ in the set of all designs satisfying (3.8). If $I(\xi) < \frac{k}{2}$, then this set is a singleton and the maximizing design ξ_Q^+ coincides with ξ. Otherwise, by Lemma 2.1 the maximizing measure ξ_Q^+ corresponds to the upper principal presentation of the moment point $(\int_A^B \Psi_0(x) d\xi(x), \ldots, \int_A^B \Psi_{k-1}(x) d\xi(x))^T$, which does not depend on the vector Q. Finally, assertion 1(a) or 1(b) of Theorem 3.1 follow from the discussion regarding the number of support points of principal representations given at the end of Section 2.

□

Acknowledgements. The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG).

References

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

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

Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

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

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

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

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

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40:633–643.

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

Khuri, A., Mukherjee, B., Sinha, B., and Ghosh, M. (2006). Design issues for generalized linear models. *Statistical Science*, 21(3):376–399.

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

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

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*. Dekker, New York.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley and Sons Inc., New York.

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

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

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

Yang, M. and Stufken, J. (2012b). On locally optimal designs for generalized linear models with group effects. *Statistica Sinica*, 22:1765–1786.