Optimal designs for free knot least squares splines

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

Viatcheslav B. Melas
St. Petersburg State University
Department of Mathematics
St. Petersburg, Russia
email: v.melas@pobox.spbu.ru

Andrey Pepelyshev
St. Petersburg State University
Department of Mathematics
St. Petersburg , Russia
email: andrey@ap7236.spbu.ru

September 25, 2006

Abstract

In this paper D-optimal designs for free knot least squares spline estimation are investigated. In contrast to most of the literature on optimal design for spline regression models it is assumed that the knots of the spline are also estimated from the data, which yields to optimal design problems for nonlinear models. In some cases local D-optimal designs can be found explicitly. Moreover, it is shown that the points of minimally supported D-optimal designs are increasing and real analytic functions of the knots and these results are used for the numerical construction of local D-optimal designs by means of Taylor expansions. In order to obtain optimal designs which are less sensitive with respect to a specification of the unknown knots a maximin approach is proposed and standardized maximin D-optimal designs for least square splines with estimated knots are determined in the class of all minimally supported designs.

Keyword and Phrases: Free knot least squares splines, D-optimal designs, nonlinear models, local optimal designs, robust designs, saturated designs

AMS Subject Classification: Primary 62K05; Secondary: 65D10

1 Introduction

Polynomial regression models have been widely used to analyze functional relations between real valued predictors and response variables. However, in many practical applications a good fit to the data using polynomial models can only be achieved by high degree polynomials coming along with a rather large number of parameters. Because a polynomial function possessing all derivatives at all locations is not flexible for approximating a curve with different degrees of smoothness at different locations, many authors propose to fit piecewise polynomials to the data, which are usually called splines in the literature [see e.g. De Boor (1978), Diercx (1995) or Eubank (1999) among many others]. Smoothing splines owe their origin to Whittaker (1923) and have been further developed by Schoenberg (1964) and Reinsch (1967). As an alternative several authors propose to use least squares splines [see e.g. Gallant and Hudson (1966), Hartley (1961), Fuller (1973) or Eubank (1988) among many authors]. If the knots are assumed to be fixed, this approach is particularly attractive because of its computational simplicity. In this case several authors have studied the problem of constructing optimal designs for the corresponding segmented polynomial regression models [see e.g. Studden and Van Arman (1969), Studden (1971), Murty (1971a,b), Park (1978), Kaishev (1989), Heiligers (1998, 1999), Woods and Lewis (2006) among others]. On the other hand, if the knots are also estimated from the data the estimation problem is a nonlinear least squares problem and the computation of the estimate and appropriate designs is substantially more difficult [see e.g. Jupp (1978), Seber and Wild (1989) or Mao and Zhao (2003)]. In particular - to the knowledge of the authors - optimal designs for least squares splines with estimated knots have not been considered so far in the literature.

The present paper is devoted to the D-optimal design problem for spline regression models with estimated knots, which are introduced in Section 2. In Section 3 we discuss local D-optimal designs, which depend on the unknown knots and have to be found numerically in nearly all applications of practical interest. It is demonstrated that in most cases the support points of minimally supported D-optimal designs are increasing and real analytic functions of the knots. This allows us to represent these designs by means of Taylor expansions and efficient algorithms for their numerical construction are presented and illustrated in several examples. In applications of spline regression models with estimated knots there is usually not much prior information regarding their location and the application of local D-optimal designs could be not robust with respect to a misspecification of the unknown knots. For these reasons a standardized maximin approach is proposed as a robust alternative, which does not require an exact knowledge of the knots before any observations are available. Some theoretical results of minimally supported standardized maximin D-optimal designs are derived, which can be used to construct these designs by means of Taylor expansions. The results are illustrated by several examples, while all more technical details are deferred to an appendix in Section 5.

2 Spline regression models with estimated knots

The general form of a spline regression model is given by

$$E[Y \mid x] = \sum_{i=1}^{k} \theta_i x^{i-1} + \sum_{i=1}^{r} \sum_{j=0}^{k_i - 1} \theta_{ij} (x - \lambda_i)_+^{m-j}, \tag{2.1}$$

where the explanatory variable x varies in a compact interval, say $[a, b]$, $\lambda_1 < \lambda_2 < \ldots < \lambda_r$ denote r knots located in the interval $[a, b]$, $k_i \le m - 1 \quad (i = 1, \ldots, r)$, $k \le m + 1$ and

$$\theta_1, \ldots, \theta_k, \theta_{10}, \ldots, \theta_{1k_1-1}, \ldots, \theta_{r0}, \ldots, \theta_{rk_r-1}, \lambda_1, \ldots, \lambda_r$$

are unknown parameters which have to be estimated from the data. Here and throughout this paper we define $z_+ = \max\{0, z\}$. Note that the model (2.1) is nonlinear in the parameters $\lambda = (\lambda_1, \ldots, \lambda_r)^T$ and linear with respect to the remaining parameters $\theta = (\theta_1, \ldots, \theta_k, \theta_{10}, \ldots, \theta_{rk_r-1})^T$ [see Seber and Wild (1989)]. Following the common convention, we measure the worth of a design by its Fisher information matrix [see Silvey (1980) or Pukelsheim (1993)]. To be precise we define a (approximate) design ξ as a probability measure with finite support on the interval $[a, b]$ [see Kiefer (1974)]. Here the support points $x_1, \ldots, x_n$ represent the locations, where observations are taken and the masses $w_1, \ldots, w_n$ give the proportions of total observations to be taken at the particular points. If N independent observations with constant variance $\sigma^2 > 0$ can be made, an appropriate rounding procedure is applied to determine the number of observations $n_j = N w_j$, taken at each point x_j; $(j = 1, \ldots, n)$ [see e.g. Pukelsheim and Rieder (1992)]. Under the assumption of normality, the covariance matrix of the maximum likelihood estimate of the parameters (θ, λ) is approximately equal to the matrix

$$\frac{\sigma^2}{N} (C_\theta M(\xi, \lambda) C_\theta^T)^{-1} \in \mathbb{R}^{\mu \times \mu} \tag{2.2}$$

where $C_\theta \in \mathbb{R}^{\mu \times \mu}$ denotes a nonsingular matrix, which depends on the parameters $\theta_{10}, \ldots, \theta_{rk_r-1}$ but not on the knots $\lambda_1, \ldots, \lambda_r$ and on the design ξ. Here $\mu = k + \sum_{i=1}^{r} k_i + r$ is the number of parameters,

$$M(\xi, \lambda) = \int_a^b f(x) f^T(x) d\xi(x)$$

is the information matrix of the design ξ and the components of the vector $f(x) = (f_1(x), \ldots, f_\mu(x))^T$ are defined by

$$f_\ell(x) = \begin{cases} x^{\ell-1}; & \ell = 1, \ldots, k \\ (x - \lambda_1)_+^{m+\alpha_0-\ell+1}; & \ell = \alpha_0 + 1, \ldots, \alpha_1 \\ (x - \lambda_2)_+^{m+\alpha_1-\ell+1}; & \ell = \alpha_1 + 1, \ldots, \alpha_2 \\ \vdots & \\ (x - \lambda_r)_+^{m+\alpha_{r-1}-\ell+1}; & \ell = \alpha_{r-1} + 1, \ldots, \alpha_r \end{cases} \tag{2.3}$$

($\ell = 1, \dots, \mu$) and $\alpha_j = k + \sum_{s=1}^{j}(k_s + 1)$; $\quad j = 0, \dots, r$. Usually optimal or efficient designs maximize an appropriate function of the Fisher information matrix. Note that in the particular model under consideration this matrix depends on the nonlinear parameter λ, that is the vector of knots. There are many optimality criteria proposed in the literature [see Silvey (1980) or Pukelsheim (1993)] and in the present paper we concentrate on D-optimal designs which minimize the determinant of the matrix in (2.2). This is equivalent to minimizing the determinant of the matrix $M^{-1}(\xi, \lambda)$, because the matrix C_θ does not depend on the design ξ.

Following Chernoff (1953) we call a design $\xi^*_{D,\lambda}$ local D-optimal if it maximizes

$$\det M(\xi, \lambda). \tag{2.4}$$

For the case of least squares estimation with given knots D-optimal designs have been considered by Park (1978), Kaishev (1989) and Lim (1991), but no results seem to be available for the situation where the knots have also to be estimated from the data. Note that the concept of local D-optimality requires a prior guess for the vector of knots and that local optimal designs are not necessarily robust with respect to a misspecification of the unknown parameters. Therefore this methodology may result in an inefficient design if the (unknown) knots are misspecified. The problem of non-robustness has been mentioned in many publications in the context of nonlinear regression models and several authors propose to use a Bayesian or maximin optimality criterion to obtain robust designs [see e.g. Chaloner and Verdinelli (1995) or Imhof (2001) among many others]. The Bayesian approach requires the specification of a prior for the nonlinear parameters in the models. Because the knots of a spline usually do not have a concrete interpretation it might be difficult to specify such a prior in a concrete situation. As an alternative for the construction of robust designs, we therefore propose in this paper a maximin approach based on the D-optimality criterion, which only requires the specification of a certain range for the unknown knots of the spline regression model. This method determines a design, which maximizes a minimum of D-efficiencies [see Müller (1995), Dette (1997), Imhof (2001)] and is motivated by the fact that in the case of free knot least squares splines it will be difficult to specify an r-dimensional prior for the (unkown) knots $\lambda_1, \dots, \lambda_r$, before any data is available.

A standardized maximin D-optimal design maximizes

$$\min_{\lambda \in \Omega} \frac{\det M(\xi, \lambda)}{\det M(\xi^*_{D,\lambda}, \lambda)} \tag{2.5}$$

where $\Omega \subset \{z = (z_1, \dots, z_r)^T \in \mathbb{R}^r \mid a < z_1 < \dots < z_r < b\}$ is a given compact set for the knots $\lambda_1, \dots, \lambda_r$ and $\xi^*_{D,\lambda}$ is the local D-optimal design for a fixed parameter λ. Throughout this paper we will also consider the corresponding optimization problems in the class of all minimally supported or saturated designs, i.e. the class of all designs with μ support points. In this case the local D-optimal design in the numerator of the expression in (2.5) is also determined in the class of all minimally supported designs.

3 Local D-optimal designs

In most circumstances local D-optimal designs for free knot least squares spline estimation in model (2.1) have to be found numerically. However, in some situations it is possible to derive explicit solutions of the D-optimal design problem. Moreover, it is also possible to derive some analytical properties (as smoothness or monotonicity) of the support points of minimally supported designs.

3.1 Explicit solutions

An explicit solution of the local D-optimal design problem for the least squares spline estimation problem is possible if the regression function in (2.1) has exactly one continuous derivative at the knots $\lambda_1, \dots, \lambda_r$. The following results presents the details.

Theorem 3.1. *Consider the (nonlinear) regression model (2.1) with $m \geq k-1$ and $k_i = m-1; i = 1, \dots, r$. There exists a unique local D-optimal design $\xi^*_{D,\lambda}$ with exactly μ support points, say $x_1 < \dots < x_\mu$ and equal weights $\xi^*_{D,\lambda}(x_j) = 1/\mu \quad j = 1, \dots, \mu$. Moreover, the support points are given by*

$$x_i = a + (\gamma_{i,k}+1)\Big(\frac{\lambda_1 - \lambda_0}{2}\Big); \quad i = 1, \dots, k, \tag{3.1}$$

$$x_{i-1+k+(\ell-1)m} = \lambda_\ell + (\gamma_{i,m+1}+1)\Big(\frac{\lambda_{\ell+1} - \lambda_\ell}{2}\Big); \quad i = 2, \dots, m+1; \ell = 1, \dots, r\,, \tag{3.2}$$

where $\lambda_0 = a, \lambda_{r+1} = b, \gamma_{1,s}, \dots, \gamma_{s,s}$ are the ordered roots of the polynomial $(x^2-1)L'_{s-2}(x)$ and $L_s(x)$ denotes the sth Legendre polynomial orthogonal with respect to the Lebesgue measure.

Proof of Theorem 3.1. Let

$$\xi^*_{D,\lambda} = \begin{pmatrix} x_1^* & \dots & x_n^* \\ w_1 & \dots & w_n^* \end{pmatrix}$$

denote a local D-optimal design for least squares estimation in the nonlinear model (2.1). By the Cauchy Binet formula it is easy to see that there must be at least k support points in the interval $[a, \lambda_1]$ and at least m support points in the intervals $(\lambda_j, \lambda_{j+1}] \quad (j = 1, \dots, r)$. Moreover, the equivalence theorem of Kiefer and Wolfowitz (1960) shows that $\xi^*_{D,\lambda}$ is local D-optimal if and only if the inequality

$$g(x) = f^T(x) M^{-1}(\xi^*_{D,\lambda}, \lambda) f(x) - \mu \leq 0 \tag{3.3}$$

holds for all $x \in [a, b]$, where the vector of regression functions is defined by (2.3). Consequently, it follows that

$$g(x_i^*) = 0 \qquad i = 1, \dots, n \tag{3.4}$$

$$g'(x_i^*) = 0 \qquad i = 2, \dots, n-1. \tag{3.5}$$

Note that g is a polynomial of degree $2k-2$ on the interval $[\lambda_0, \lambda_1] = [a, \lambda_1]$ and a polynomial of degree $2m$ on the interval $[\lambda_1, \lambda_{r+1}] = [\lambda_1, b]$. If $\xi^*_{D,\lambda}$ would have more than $\mu = k + rm$ support points there would exist at least one interval with more than k (for the interval $[\lambda_0, \lambda_1]$) or more than m support points (for the remaining intervals $(\lambda_j, \lambda_{j+1}]; j = 1, \ldots, r$). Both cases would yield a contradiction and as a consequence we have $n = k + mr$. Moreover, the same argument yields

$$\begin{array}{ccccccccc} \lambda_0 & = & x_1^* & < & \ldots & < & x_k^* & = & \lambda_1 \\ \lambda_1 & < & x_{k+1}^* & < & \ldots & < & x_{k+m}^* & = & \lambda_2 \\ \vdots & & & \vdots & & \vdots & \vdots & & \\ \lambda_r & < & x_{k+(r-1)m+1}^* & < & \ldots & < & x_{k+rm}^* & = & \lambda_{r+1}. \end{array} \tag{3.6}$$

A standard argument [see Silvey (1980)] now shows that the weights of the local D-optimal design are all equal, that is $\xi^*_{D,\lambda}(x_j^*) = 1/\mu;\ j = 1, \ldots, \mu$. This implies

$$\det M(\xi^*_{D,\lambda}, \lambda) = \Big(\frac{1}{\mu}\Big)^{\mu} (\det F)^2, \tag{3.7}$$

where

$$F = \begin{bmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_{r+1} \end{bmatrix}$$

denotes a block triangular matrix with blocks in the diagonal given by

$$\begin{aligned} F_1 &= (f_i(x_j^*))_{i,j=1}^{k} \in \mathbb{R}^{k\times k} \\ F_\ell &= (f_i(x_j^*))_{i,j=k+(\ell-1)m+1}^{k+\ell m} \in \mathbb{R}^{m\times m} \quad \ell = 2, \ldots, r+1. \end{aligned}$$

As a consequence we obtain from (3.7) the representation

$$\det M(\xi^*_{D,\lambda}, \lambda) = \Big(\frac{1}{\mu}\Big)^{\mu} \prod_{j=1}^{r+1} (\det F_j)^2,$$

and the blocks can be maximized separately. The first block is a classical Vandermonde determinant with $x_1^* = \lambda_0 = a, x_k^* = \lambda_1 = b$, and consequently maximized for the support points of the local D-optimal design on the interval $[a, \lambda_1]$, which are given by (3.1) [see e.g. Hoel (1958)]. The other determinants are of the form

$$(\det F_\ell)^2 = \begin{vmatrix} (z_1-\lambda_\ell)^m & \ldots & (z_{m-1}-\lambda_\ell)^m & (\lambda_{\ell+1}-\lambda_\ell)^m \\ \vdots & & \vdots & \vdots \\ (z_1-\lambda_\ell)^2 & \ldots & (z_{m-1}-\lambda_\ell)^2 & (\lambda_{\ell+1}-\lambda_\ell)^2 \\ (z_1-\lambda_\ell) & \ldots & (z_{m-1}-\lambda_\ell) & (\lambda_{\ell+1}-\lambda_\ell) \end{vmatrix}^2$$

$$= (\lambda_{\ell+1} - \lambda_\ell)^2 \prod_{j=1}^{m-1} (z_j - \lambda_\ell)^2 (\lambda_{\ell+1} - z_j)^2 \prod_{1 \le i < j \le m-1} (z_i - z_j)^2,$$

where $z_j = x^*_{k+(\ell-1)m+j} \quad (j = 1, \dots, m-1; \ell = 1, \dots, r)$. Now the results of Hoel (1958) show again that this expression is maximized if $z_1, \dots, z_{m-1}$ are the interior support point of the D-optimal design for a polynomial regression of degree m on the interval $[\lambda_\ell, \lambda_{\ell+1}]$, which are given by (3.2). □

Note that Theorem 3.1 generalizes a result of Lim (1991), who considered model (2.1) in the special case $k = m + 1$, where the knots are known and do not have to be estimated from the data.

Example 3.2. Consider the model

$$E[Y \mid x] = \theta_1 + \theta_2 x + \theta_3 x^2 + \sum_{j=1}^{r} \theta_{3+j} (x - \lambda_j)_+^2; \quad x \in [a, b], \tag{3.8}$$

where we have $k = 3; m = 2; k_j = 1 \quad (j = 1, \dots, r)$. According to Theorem 3.1 the local D-optimal design is given by $(\lambda_0 = a, \lambda_{r+1} = b)$

$$\xi^*_{D,\lambda} = \begin{pmatrix} \lambda_0 & \frac{\lambda_0+\lambda_1}{2} & \lambda_1 & \dots & \lambda_r & \frac{\lambda_r+\lambda_{r+1}}{2} & \lambda_{r+1} \\ \frac{1}{2r+3} & \frac{1}{2r+3} & \frac{1}{2r+3} & \dots & \frac{1}{2r+3} & \frac{1}{2r+3} & \frac{1}{2r+3} \end{pmatrix}. \tag{3.9}$$

Table 1: *The non-trivial support points of the local D-optimal designs in the regression model (3.10) The local D-optimal design is given by $\xi^* = \{0, x_2^*(\lambda), \dots, x_5^*(\lambda), 1; 1/6, \dots, 1/6\}$.*

λ	$x_2^*(\lambda)$	$x_3^*(\lambda)$	$x_4^*(\lambda)$	$x_5^*(\lambda)$
0.1	0.033	0.094	0.345	0.750
0.2	0.065	0.180	0.410	0.775
0.3	0.095	0.258	0.473	0.799
0.4	0.124	0.330	0.536	0.824
0.5	0.151	0.398	0.602	0.849
0.6	0.176	0.464	0.670	0.876
0.7	0.201	0.527	0.742	0.904
0.8	0.225	0.590	0.820	0.935

In general optimal designs for least squares slines with estimated knots have to be found numerically. Consider as a typical example a cubic spline regression model (with continuous first and second derivative)

$$E[Y|x] = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 + \theta_5 (x-\lambda)_+^3; \quad x \in [0,1]. \tag{3.10}$$

A straightforward calculation shows that the vector of regression functions in model (2.1) is given by

$$f(x) = (1, x, x^2, x^3, (x-\lambda)_+^3, (x-\lambda)_+^2)^T.$$

Some local D-optimal designs have been calculated numerically for various values of λ. The results are presented in Table 1 and indicate that the support points of the local D-optimal design are strictly increasing functions of the knots. This phenomenon will be further investigated in Section 3.2.

It might be also of interest to study the sensitivity of the local D-optimal design with respect to a misspecification of the initial knots. For this purpose we present in Figure 3.1 the D-efficiencies of the of the local D-optimal design in the spline regression model (3.8) for the values $\lambda = 0.25$ and $\lambda = 0.5$. We observe that the D-efficiencies decrease very rapidly if the knot is misspecified.

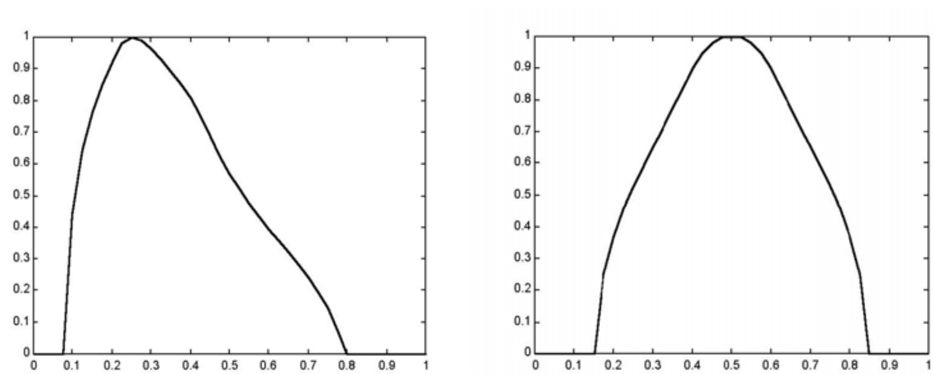

Figure 3.1: *The D-efficiencies of the local D-optimal design in the spline regression model (3.8), where $\lambda = 0.25$ (left panel) and $\lambda = 0.5$ (right panel).*

3.2 Some properties of local D-optimal designs

In this section we discuss two important features of local D-optimal designs for free knot least squares splines. It is indicated in Example 3.2 that the support points of local D-optimal designs are increasing and analytic functions of the knots [see Table 1] and this property will be proved for the case where the local D-optimal design is minimally supported [see Theorem 3.4 below]. Secondly, we prove a symmetry property of local D-optimal designs for least squares splines with estimated knots in the case where there is the same degree of smoothness at each knot. We begin our investigations with the symmetry result.

Theorem 3.3. *Consider the spline regression model (2.1) with knots* $\lambda = (\lambda_1, \dots, \lambda_r)^T$ *and let*

$$\xi^*_{D,\lambda} = \begin{pmatrix} x^*_1, & \dots, & x^*_n \\ w^*_1, & \dots, & w^*_n \end{pmatrix}$$

denote a local D-optimal design. The design

$$\tilde{\xi}_{D,\lambda} = \begin{pmatrix} \tilde{x}_1, & \dots, & \tilde{x}_n \\ w^*_1, & \dots, & w^*_n \end{pmatrix}$$

with $\tilde{x}_i = b + a - x^*_i \quad (i = 1, \dots, n)$ *is local D-optimal for least squares spline estimation in the model (2.1) with knots* $\tilde{\lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_r)^T = (b + a - \lambda_r, \dots, b + a - \lambda_1)^T$.

Proof of Theorem 3.3. The assertion follows from a basic property of the D-optimality criterion observing that the functions

$$1, b + a - x, \dots, (b + a - x)^{k-1}, (x - \tilde{\lambda}_1)_+^{m-k_1}, \dots (x - \tilde{\lambda}_1)_+^m, \dots, (x - \tilde{\lambda}_r)_+^{m-k_r}, \dots, (x - \tilde{\lambda}_r)_+^m$$

and

$$1, x, \dots, x^{k-1}, (x - \lambda_1)_+^{m-k_1}, \dots, (x - \lambda_1)_+^m, \dots, (x - \lambda_r)_+^{m-k_r}, \dots, (x - \lambda_r)_+^m$$

generate the same space. □

Numerical results indicate that local D-optimal designs for free knot least squares splines are minimally supported. In such cases it follows by a standard convexity argument that the local D-optimal design is unique and the following theorems show that in this case the corresponding support points are increasing and analytic functions of the knots, if the condition

$$m - k - 2 \leq k_1 = k_2 = \dots = k_r \leq m - 1 \tag{3.11}$$

is satisfied. The proofs are complicated and therefore deferred to the Appendix.

Theorem 3.4. *Consider the spline regression model (2.1) satisfying (3.11). If any local D-optimal design is minimally supported, then the local D-optimal design* $\xi^*_{D,\lambda}$ *is unique and its support points, which do not coincide with the knots* $a = \lambda_0 < \lambda_1 < \dots < \lambda_r < \lambda_{r+1} = b$, *are strictly increasing functions of any component of the vector* $\lambda = (\lambda_1, \dots, \lambda_r)^T$. *Moreover, the boundary points* a *and* b *of the design space are support points of the local D-optimal design* $\xi^*_{D,\lambda}$.

Theorem 3.5. *Under the assumptions of Theorem 3.4 let*

$$\Omega := \{(\lambda_1, \dots, \lambda_k)^T \mid a < \lambda_1 < \dots < \lambda_k < b\} = \bigcup_{j=1}^{j^*} \Omega_j \tag{3.12}$$

denote a partition of the set of possible knots such that $\Omega_i \cap \Omega_j \neq \emptyset$ *and that for all* $\lambda \in \Omega_j$ *the number of suppport points of the (unique) local* D*-optimal design in each interval* $(\lambda_i, \lambda_{i+1})$ $(i = 0, \ldots, r)$ *is fixed. The support points of the local* D*-optimal design, which do not coincide with the knots* $a = \lambda_0 < \lambda_1 < \ldots < \lambda_r < \lambda_{r+1} = b$ *are real analytic functions on* Ω_j *(for each* $j = 1, \ldots, j^*$*).*

3.3 Taylor expansions for local D-optimal designs

The analytic properties of local D-optimal designs for spline regression models allow an elegant numerical calculation of the support points which will be briefly indicated in this paragraph. The numerical results presented in Example 3.2 were already obtained by this method. To be precise let the assumptions of Theorem 3.5 be satisfied, then the local D-optimal design is unique and of the form

$$\xi_{\tau^*} = \begin{pmatrix} a & \tau_1^* & \ldots & \tau_{\mu-2}^* & b \\ \frac{1}{\mu} & \frac{1}{\mu} & \ldots & \frac{1}{\mu} & \frac{1}{\mu} \end{pmatrix},$$

where the support points $\tau^*(\lambda) = (\tau_1^*, \ldots, \tau_{\mu-2}^*)$ are real analytic functions of the vector of knots $\lambda = (\lambda_1, \ldots, \lambda_r)$ on each set Ω_j defined in (3.12). For the sake of simplicity consider the case $r = 1$, define $\lambda = \lambda_1$ and denote by $\tau_{(0)}^*$ the vector of support points of the local D-optimal design for the knot $\lambda_{(0)} \in \Omega_j$ (for some $j = 1, \ldots, j^*$). From Theorem 3.5 it follows that a Taylor expansion of the form

$$\tau^*(\lambda) = \tau_{(0)}^* + \sum_{i=1}^{\infty} \tau_{(i)}^* (\lambda - \lambda_0)^i \tag{3.13}$$

is valid, where the coefficients are given by

$$\tau_{(s)}^* = \frac{1}{s!} \frac{d^s}{d^s\lambda} \tau^*(\lambda)\Big|_{\lambda=\lambda_0}; \quad s = 0, 1, 2, \ldots$$

Moreover, the coefficients in the expansion can be calculated recursively [see Dette, Melas and Pepelyshev (2004)] using the following recursive relations

$$\tau_{(s+1)}^* = -\frac{1}{(s+1)!} J_{(0)}^{-1} \Big\{ \Big(\frac{d^{s+1}}{d^{s+1}\lambda} \Big) g(\tau_{<s>}^*(\lambda), \lambda) \Big\} \Big|_{\lambda=\lambda_0,} \quad s = 0, 1, \ldots, \tag{3.14}$$

where

$$\begin{aligned} J_{(0)} &= \Big(\frac{\partial^2}{\partial\tau_i \partial\tau_j} \psi(\tau, \lambda_0) \Big|_{\tau=\tau_{(0)}^*} \Big)_{i,j=1}^{\mu-2}, \\ g(\tau, \lambda) &= \Big(\frac{\partial^2}{\partial\tau_i \partial\lambda} \psi(\tau, \lambda) \Big)_{i=1}^{\mu=2}, \\ \psi(\tau, \lambda) &= \Big\{ \det M(\xi_\tau, \lambda) \Big\}^{1/\mu}, \end{aligned}$$

and we define for any (sufficiently differentiable) function h

$$h_{\langle s\rangle}(\lambda) = \sum_{i=0}^{s} \frac{1}{i!}\Big(\frac{d^i}{d\lambda^i}h(\lambda)\Big)\Big|_{\lambda=\lambda_0}(\lambda-\lambda_0)^i \qquad (3.15)$$

We finally note that in the case of several knots (that is $r \geq 2$) an extension of formula (3.13) is given in Melas (2006) and the details are omitted for the sake of brevity.

Example 3.6. Consider the cubic spline regression model (3.10) of Example 3.2. The support points of the local D-optimal designs in Table 1 have been calculated by a Taylor expansion at the point $\lambda = 0.5$. To be precise note that the support points satisfy

$$x_2^*(\lambda) = 1 - x_5^*(1-\lambda) \quad x_3^*(\lambda) = 1 - x_4^*(1-\lambda)$$

[see Theorem 3.3]. In the following we construct Taylor expansions for the support points of the local D-optimal design at the point $\lambda = 0.5$. With the notation $\tau_i^* = x_{i+1}^*$ $(i = 1, \ldots, 4)$, $u = \lambda - 0.5$ we obtain

$$\begin{aligned}
\tau_1^*(u) &= 0.151 + 0.261\,u - 0.0689\,u^2 + 0.0692\,u^3 + 0.0595\,u^4 - 0.0425\,u^5 \\
&\quad +0.0400\,u^6 + 0.0333\,u^7 + 0.0184\,u^8 + 0.0285\,u^9 + 0.0647\,u^{10}, \\
\tau_2^*(u) &= 0.398 + 0.664\,u - 0.153\,u^2 + 0.216\,u^3 + 0.0204\,u^4 + 0.0408\,u^5 \\
&\quad +0.00556\,u^6 + 0.127\,u^7 + 0.0175\,u^8 + 0.146\,u^9 + 0.0569\,u^{10}, \\
\tau_3^*(u) &= 1 - \tau_2^*(-u), \\
\tau_4^*(u) &= 1 - \tau_1^*(-u).
\end{aligned}$$

The support point are depicted in Figure 3.3 as a function of the knot λ. Note that all support points are increasing functions of the nonlinear parameter λ [see Theorem 3.4].

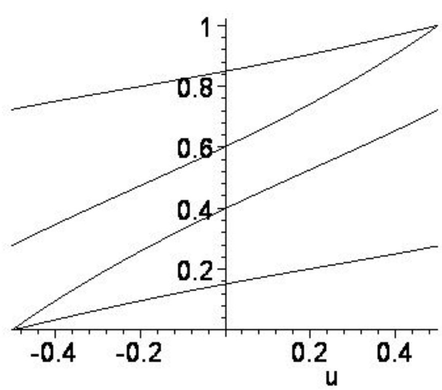

Figure 3.2: *The interior points $\tau_j^* = \tau_j^*(u)$ of the local D-optimal design for the spline regression model (3.10) considered as a function of the parameter u.*

4 Standardized maximin D-optimal designs

If the knots of the spline regression model are estimated from the data there is usually not too much knowledge available with respect to their location. At the same time the numerical results of Section 3 indicate that local D-optimal designs are rather sensitive with respect to the specification of the knots. The standardized maximin optimality criterion (2.5) might be more appropriate for the construction of efficient designs in least squares spline estimation. In the simplest case of model (3.8) with one knot the standardized maximin D-optimal design can be found explicitly in the class of all minimally supported designs.

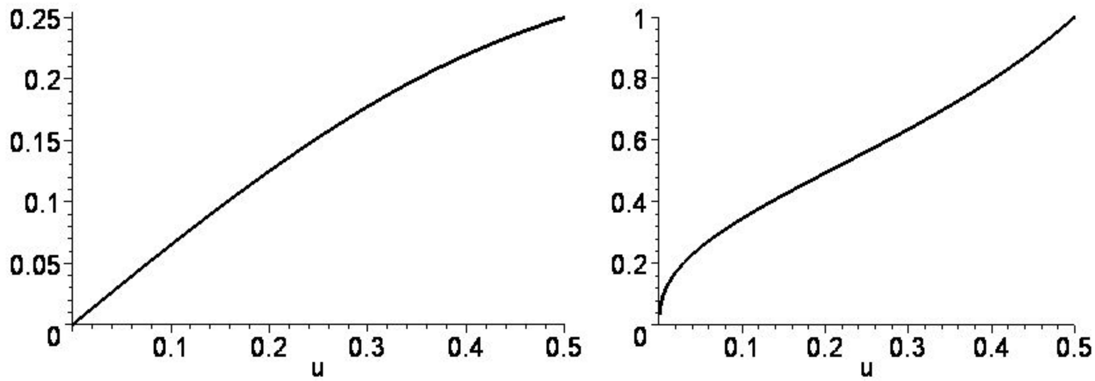

Figure 4.1: *The non-trivial support point $x^* = x(u)$ of the minimally supported standardized maximin D-optimal design (left panel) for the spline regression model (4.2) and its minimal efficiency in the interval $[u, 1-u]$ (right panel).*

Example 4.1. Consider the spline regression model

$$E[Y \mid x] = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 (x-\lambda)_+^2, \tag{4.1}$$

where (without loss of generality) $x \in [0, 1]$. The local D-optimal design is given by (3.9) with $r = 1$ and it is easy to see that the minimally supported standardized maximin D-optimal design must contain the point 0 and 1 in its support [see the proof of Lemma 5.2 in the Appendix]. In the following we consider the set $\Omega = [u, 1-u]$ with $u \in (0, 1/2)$ in the optimality criterion (2.5), then it follows by similar arguments as given in the proof of Theorem 3.3 that the minimally supported standardized maximin D-optimal design is of the form

$$\xi^* = \begin{pmatrix} 0 & x & \frac{1}{2} & 1-x & 1 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix},$$

where $x \in (0, 1/2)$. Consequently, the optimality criterion (2.5) reduces for minimally supported designs to

$$\min_{\lambda \in [u, 1-u]} \frac{\det M(\xi^*, \lambda)}{\det M(\xi^*_\lambda, \lambda)} = 4 \frac{x^2 (x-u)(2x+1)}{(1-u)^3 u^2}$$

Now a straightforward calculation shows that the function on the right hand side is maximal for

$$x^*(u) = \frac{3}{16} + \frac{3}{8}u - \frac{1}{16}\sqrt{(6u-3)^2 + 8u}.$$

The non-trivial support point of the minimally supported design is displayed in Figure 4.1 for various values of $u \in (0, 1/2)$. In the right part of the Figure we display the minimal efficiency of the minimally supported standardized maximin D-optimal design in the interval $[u, 1-u]$, which decreases rapidly with an increasing length of the interval.

In the remaining part of this section we discuss the numerical construction of minimally supported standardized maximin D-optimal designs. In order to derive a Taylor expansion for such designs we consider the following set Ω in the optimality criterion (2.5):

$$\Omega = \Omega_\delta = \{(\lambda_1, \ldots, \lambda_r)^T \mid (1-\delta)c_i \leq \lambda_i \leq (1+\delta)c_i; \; i = 1, \ldots, r\}, \tag{4.2}$$

where $c = (c_1, \ldots, c_r)^T$ with $c_1 < \ldots < c_r$ is a fixed vector (considered as preliminary guess for the unknown vector of knots) and $\delta \in (0, 1)$ is the relative error of this approximation. The following result shows that for sufficiently small δ and minimally supported designs the minimization in the optimality criterion (2.5) can be replaced by a minimization with respect to a two point set. For this purpose let $\bar{\Xi}$ denote the set of all minimally supported designs for the spline regression model (2.1) on the interval $[a, b]$. The proof of the next theorem is complicated and therefore deferred to the Appendix [see Section 5.2].

Theorem 4.2.

(a) *Let Ω_δ be defined by (4.2), then there exists a number $\delta^* > 0$ such that for any $\delta \in [0, \delta^*)$*

$$\max_{\xi \in \bar{\Xi}} \min_{\lambda \in \Omega_\delta} \frac{\det M(\xi, \lambda)}{\sup_{\eta \in \bar{\Xi}} \det M(\eta, \lambda)} = \max_{\xi \in \bar{\Xi}} \min_{\lambda \in \Omega_\delta^*} \frac{\det M(\xi, \lambda)}{\sup_{\eta \in \bar{\Xi}} \det M(\eta, \lambda)},$$

where $\Omega_\delta^* \in \mathbb{R}^r$ is a two point set defined by

$$\Omega_\delta^* = \{(1-\delta)c, (1+\delta)c\}.$$

In other words: If δ is sufficiently small the minimally supported standardized maximin D-optimal design with respect to the set Ω_δ coincides with the minimally supported standardized maximin D-optimal design with respect to the set Ω_δ^*.

(b) *For any $\delta \in [0, \delta^*)$ the support points of the minimally supported standardized maximin D-optimal design are real analytic functions of the parameter $\lambda \in \Omega_\delta$.*

Note that Theorem 4.2 allows us to calculate minimally supported standardized maximin D-optimal designs by means of a Taylor expansion as it was illustrated in Section 3.3 for the

Table 2: *The support points of the minimally supported standardized maximin D-optimal design with respect to the set $\Omega = [u, v]$ in the regression model (4.1). The right column shows the minimal efficiency calculated over the set Ω*

u	v	x_1	x_2	x_3	x_4	x_5	min eff
0.4	0.6	0	0.220	0.5	0.780	1	0.796
0.3	0.7	0	0.178	0.5	0.822	1	0.636
0.2	0.8	0	0.125	0.5	0.875	1	0.494
0.1	0.9	0	0.065	0.5	0.935	1	0.346
0.05	0.95	0	0.033	0.5	0.967	1	0.253
0.5	0.6	0	0.261	0.545	0.789	1	0.890
0.5	0.7	0	0.270	0.581	0.833	1	0.794
0.5	0.8	0	0.274	0.604	0.882	1	0.702
0.5	0.9	0	0.272	0.599	0.937	1	0.594
0.5	0.95	0	0.264	0.564	0.967	1	0.510

case of local D-optimal designs. The corresponding recursive relations are obtained by a slight modification from those presented in Section 3.3 and the details are omitted for the sake of brevity. We conclude this section with a continuation of Example 4.1.

Example 4.3. The concrete values for the support points of the minimally supported standardized maximin D-optimal designs for the spline regression model (4.2) are presented in Table 2, which also shows results for a non-symmetric parameter space $\Omega = [u, v]$. In this case there exists no analytical solution and the designs have been derived by means of a Taylor expansion, which was described in the previous paragraphs. In its last row the table also contains the minimal efficiency of the minimally supported standardized maximin D-optimal design. We observe that these minimal efficiencies decrease substantially, if the range for the free knot λ_1 becomes large. For example, if $\Omega = [u, v] = [0.1, 0.9]$ the minimally supported standardized maximin D-optimal design has only efficiency 34.6 % at some points of the parameter space Ω (note that this is the worst efficiency in the set Ω and that other values $\lambda \in \Omega$ can yield substantially larger efficiencies). On the other hand, if the prior information for the knot λ_1 is rather precise (that is the length $v - u$ of the set Ω is small), the minimally supported designs are rather efficient for all values of the set Ω.

The reason for the loss of efficiency in the situation where the length of the interval $v - u$ approaches 1 is that in this case the standardized maximin D-optimal designs have substantially more support points than the number of parameters in the model. In fact it can be proved using the techniques recently developed by Braess and Dette (2006) that the number of support points of the standardized maximin D-optimal design becomes arbitrary large if $v - u \to 1$. Two illustrative examples are given in Table 3, which shows the standardized maximin D-optimal designs for the parameter spaces $\Omega = [0.45, 0.55]$ and $\Omega = [0.4, 0.6]$, which have already 8 and 10 support points, respectively. If the interval is not symmetric the number of support points grows rapidly with the length of the set Ω. For example, if $\Omega = [0.3, 0.5]$ the standardized maximin D-optimal design has already 14 support points. However designs with with a moderate number of support points yield usually reasonable efficiencies. For example, if $\Omega = [0.3, 0.5]$, the 8-point designs with masses 0.198, 0.170, 0.074, 0.050, 0.045, 0.082, 0.181, 0.199 at the points 0, 0.170, 0.312, 0.372, 0.428, 0.490, 0.725, 1, respectively, is not globally optimal, but its minimal efficiency over the set $\Omega = [0.3, 0.5]$ is given by 0.880.

Table 3: *Globally standardized maximin D-optimal designs with respect to the set $\Omega = [u, v]$ in the regression model (4.1). The right column shows the minimal efficiency of the set Ω.*

u	v		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	min eff
0.45	0.55	x_i	0	0.238	0.452	0.484	0.516	0.548	0.762	1			0.923
		w_i	0.201	0.191	0.073	0.036	0.036	0.073	0.191	0.201			
0.4	0.6	x_i	0	0.225	0.406	0.451	0.484	0.516	0.549	0.594	0.775	1	0.883
		w_i	0.201	0.174	0.069	0.029	0.026	0.026	0.029	0.069	0.174	0.201	

Acknowledgements. The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen") is gratefully acknowledged. The work by V.B.Melas was partly supported by the Russian Foundation of Basic Research (project 07-01-00534). The work of H. Dette was supported in part by a NIH grant award IR01GM072876:01A1. The authors are also grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

5 Appendix: More technical proofs

5.1 Proof of Theorem 3.4 and 3.5.

We start presenting two auxiliary results

Lemma 5.1. *Consider the spline polynomial*

$$\psi(x) = \sum_{i=1}^{\mu} \alpha_i f_i(x), \tag{5.1}$$

where the functions $f_1(x), \dots, f_\mu(x)$ are defined by (2.3) and condition (3.11) is satisfied. If $\sum_{i=1}^{\mu} \alpha_i^2 \neq 0$, the number of isolated roots counted with their multiplicity is at most $\mu - 1$.

Proof. Assume that the spline polynomial in (5.1) has more than $\mu - 1$ isolated roots, then it follows that the function

$$\tilde{\psi}(x) = \left(\frac{d}{dx}\right)^{m-k_1-1} \psi(x)$$

has at least $\mu - m + k_1 + 1$ isolated roots. On the other hand this polynomial is of the form

$$\tilde{\psi}(x) = \sum_{j=0}^{k-m+k_1} \tilde{\alpha}_j x^j + \sum_{i=1}^{r} \sum_{j=1}^{k_1+1} \tilde{\alpha}_{ij} (x - \lambda_i)^j.$$

Therefore $\tilde{\psi}$ is a polynomial of degree $\leq k - m + k_1$ on the interval $[a, \lambda_1]$ and a polynomial of degree $k_1 + 1$ on the remaining r intervals $(\lambda_1, \lambda_2], \ldots, (\lambda_r, \lambda_{r+1}]$. Consequently, $\tilde{\psi}$ has at most

$$\tilde{\mu} := k - m + k_1 + r(k_1 + 1)$$

isolated roots counted with multiplicity, which yields

$$\mu - m + k_1 + 1 \leq \tilde{\mu} = k - m + k_1 + r(k_1 + 1).$$

Observing that $\mu = k + r(k_1 + 1)$ this inequality reduces to $1 \leq 0$, which is a contradiction. □

Lemma 5.2. *Any minimally supported local D-optimal design has the boundary points a and b as its support points.*

Proof. If ξ is a minimally supported local D-optimal design it must have equal weights $1/\mu$ at its support points $x_1 < \ldots < x_\mu$. From the discussion in the proof of Theorem 2.1 it follows that

$$\det M(\xi, \lambda) = \left\{\det(f_i(x_j))_{i,j=1}^{\mu}\right\}^2 \mu^{-\mu}.$$

Now consider the function

$$\psi(x_1) = \det(f_i(x_j))_{i,j=1}^{\mu} = \sum_{i=1}^{\mu} \alpha_i f_i(x_1),$$

where the last identity follows from Laplace's rule and the constants $\alpha_1, \ldots, \alpha_\mu$ depend on the points $x_2, \ldots, x_\mu$ but not on the point x_1. Obviously, $\psi(x_j) = 0$ for $j = 2, \ldots, \mu$ and consequently $\psi'(x)$ vanishes at $\mu - 2$ points $\tilde{x}_j \in (x_j, x_{j+1})$; $(j = 2, \ldots, \mu - 1)$. If $x_1 > a$ we would also have $\psi'(x_1) = 0$. On the other hand it follows from Lemma 5.1 that ψ' has at most $\mu - 2$ roots which is a contradiction. Consequently, $x_1 = a$ and it can be proved by similar arguments that $x_\mu = b$. □

It now follows that a minimally supported local D-optimal design is characterized by its interior support points

$$\tau = (\tau_1, \ldots, \tau_{\mu-2}) = (x_2, \ldots, x_{\mu-1})$$

and consequently we denote candidates for such designs by

$$\xi_\tau = \begin{pmatrix} a & \tau_1 & \dots & \tau_{\mu-2} & b \\ \frac{1}{\mu} & \frac{1}{\mu} & \dots & \frac{1}{\mu} & \frac{1}{\mu} \end{pmatrix}.$$

Therefore the problem of determining minimally supported local D-optimal designs reduces to the maximization of the function

$$\psi(\tau,\lambda) = [\det M(\xi_{\tau,\lambda})]^{1/\mu} \tag{5.2}$$

over the set

$$T = \{\tau = (\tau_1,\dots,\tau_{\mu-2})^T \mid a \le \tau_1 \le \dots \le \tau_{\mu-2} \le b\}, \tag{5.3}$$

where

$$\lambda \in \Omega := \{(\lambda_1,\dots,\lambda_k)^T \mid a < \lambda_1 < \dots < \lambda_k < b\} \tag{5.4}$$

is a fixed parameter. Note that under the assumptions of Theorem 3.4 this optimization problem has a unique solution, say $\tau^* = \tau^*(\lambda)$, which satisfies the necessary conditions for an extremum, i.e.

$$\frac{\partial}{\partial \tau_i}\psi(\tau,\lambda)\Big|_{\tau=\tau^*} = 0; \quad i = 1,\dots,\mu-2. \tag{5.5}$$

Using the same arguments as in Melas (2006), p. 65-66, it now follows from Lemma 5.1 that the Jacobi matrix of equation (5.5),

$$J(\lambda) := \Big(\frac{\partial^2}{\partial\tau_i\partial\tau_j}\psi(\tau,\lambda)\Big|_{\tau=\tau^*(\lambda)}\Big)_{i,j=1}^{\mu-2},$$

is non-singular and

$$(J^{-1}(\lambda))_{ij} < 0; \quad i,j = 1,\dots,\mu-2 \tag{5.6}$$

$$\frac{\partial^2}{\partial\tau_i\partial\lambda_j}\psi(\tau,\lambda)(-1)^{s(i)}\mid_{\tau=\tau^*} < 0; \quad i=1,\dots,\mu-2; \quad j=1,\dots,r \tag{5.7}$$

where $s(i) \in \{1,2\}$. Note that there could exist several solutions of (5.5) corresponding to local extrema of the function ψ. However, from the assumptions of the theorem it follows that for a fixed parameter $\lambda_0 \in \Omega$ there exists a global maximum of the function ψ and we denote by $\overline{\tau} = \tau^*(\lambda_0)$ a solution of (5.5) corresponding to this global maximum. From the implicit function theorem [see Gunning and Rossi (1965)] it therefore follows that the function $\tau^*(\lambda)$ is a unique continuous solution of (5.5) such that $\bar\tau = \tau^*(\lambda_0)$. By the same theorem we obtain for $j = 1,\dots,r; i = 1,\dots,\mu-2$

$$\frac{\partial}{\partial\lambda_j}\tau_i^*(\lambda) = \Big(J^{-1}(\lambda)G_j(-1)^{s(i)}\Big)_i > 0,$$

where the vector G_j is defined by

$$G_j = \Big(\frac{\partial^2}{\partial\tau_\ell\partial\tau_j}\psi(\tau,\lambda)\Big|_{\tau=\tau^*(\lambda)}\Big)_{\ell=1}^{\mu-2}.$$

As a consequence the support points of the local D-optimal design for the spline regression model are increasing functions of the knots. Finally, if λ is an interior point of one of the sets Ω_j in the partition (3.12), the function $\psi(\tau,\lambda)$ is real analytic and by the implicit function theorrem the solution $\tau(\lambda)$ of (5.5) is also real analytic.

5.2 Proof of Theorem 4.2.

Note that a minimally supported standardized maximin D-optimal design (with respect to any set Ω) must have equal weights. Recall the definition of the function ψ in (5.2), define

$$\varphi(\tau,\lambda) = \frac{\psi(\tau,\lambda)}{\psi(\tau^*(\lambda),\lambda)}, \tag{5.8}$$

where $\tau^* = \tau^*(\lambda)$ is the vector of support points of the minimally supported local D-optimal design. Obviously, we have

$$\min_{\lambda\in\Omega_\delta^*}\varphi(\tau,\lambda) = \min_{\alpha\in[0,1]}\varphi(\tau,\alpha,\delta) \tag{5.9}$$

with

$$\varphi(\tau,\alpha,\delta) = (1-\alpha)\varphi(\tau,(1-\delta)c) + \alpha\varphi(\tau,(1+\delta)c). \tag{5.10}$$

Consequently, the problem of finding the minimally supported standardized maximin D-optimal design with respect to the set Ω_δ^* can be reduced to finding a solution $(\hat\tau,\hat\alpha)$ of

$$\max_{\tau\in T}\min_{\alpha\in[0,1]}\varphi(\tau,\alpha,\delta), \tag{5.11}$$

where the set T is defined by

$$T = \{\tau = (\tau_1,\dots,\tau_{\mu-2}) \mid a < \tau_1 < \ldots < \tau_{\mu-2} < b\}$$

(if two components of the vector τ would be equal the determinant would vanish). The necessary conditions for an extremum yield

$$\begin{aligned} &\frac{\partial}{\partial\tau_i}\varphi(\tau,\alpha,\delta)\Big|_{\tau=\hat\tau} = 0; \quad i=1,\dots,\mu-2, \\ &\frac{\partial}{\partial\alpha}\varphi(\tau,\alpha,\delta)\Big|_{\alpha=\hat\alpha} = 0, \end{aligned} \tag{5.12}$$

which will be further investigated using the following parameterization

$$\Phi(u,\delta) = \varphi(\tau^* + \rho\delta^2, \frac{1}{2} + \beta\delta, \delta) \cdot \frac{\psi(\tau^*, c)}{\delta^2}. \tag{5.13}$$

Here $u = (\rho, \beta) = (\rho_1, \ldots, \rho_{\mu-2}, \beta)$ and τ^* denotes the vector of interior support points of the minimally supported local D-optimal design for the vector $c = (c_1, \ldots, c_r)$; i.e. $\tau^* = \tau^*(c)$. Obviously, the equations (5.12) are equivalent to

$$\frac{\partial}{\partial u_i}\Phi(u,\delta)\Big|_{u=\hat{u}} = 0, \quad i = 1, \ldots, \mu - 1, \tag{5.14}$$

and the solutions $\hat{u} = (\hat{\rho}, \hat{\beta})$ and $(\hat{\tau}, \hat{\alpha})$ are related by

$$\hat{\tau} = \tau^* + \hat{\rho}\delta^2; \hat{\alpha} = \frac{1}{2} + \hat{\beta}\delta. \tag{5.15}$$

Assume that δ^* is sufficiently small and define the set

$$\mathcal{U}_\rho := \Big\{u = (\rho, \beta) \Big| \frac{a - \tau^*}{\delta^2} < \rho_1 < \ldots < \rho_{\mu-2} < \frac{b - \tau^*}{\delta^2}; -\frac{1}{2\delta} \le \beta \le \frac{1}{2\delta}\Big\},$$

then we prove the following assertions.

(I) There exists a unique continuous function

$$\hat{u} : \begin{cases} (-\delta^*, \delta^*) & \rightarrow \ \mathcal{U} \\ \qquad \delta & \rightarrow \ \hat{u}(\delta) \end{cases} \tag{5.16}$$

such that for each $\delta \in (-\delta^*, \delta^*)$ the value $\hat{u}(\delta)$ is a solution of the system (5.14).

(II) The function defined in (I) is real analytic and the coefficients in the corresponding Taylor expansion

$$\hat{u}(\delta) = \sum_{j=0}^{\infty} u_{(j)}\delta^j$$

can be calculated recursively as

$$\begin{aligned} u_{(0)} &= -\hat{J}^{-1}[h(0,\delta)]_{(2)}, \\ u_{(s+1)} &= -\hat{J}^{-1}[h(u_{\langle s\rangle}(\delta), \delta)]_{(s+3)}, \quad s = 0, 1, 2, \ldots, \end{aligned} \tag{5.17}$$

where $u_{\langle s\rangle}$ is defined in (3.15),

$$h(u,\delta) = \Big(\frac{\partial}{\partial u_1}\Phi(u,\delta), \ldots, \frac{\partial}{\partial u_{\mu-1}}\Phi(u,\delta)\Big)^T \tag{5.18}$$

$$A = \Big(\frac{\partial^2}{\partial\tau_i\partial\tau_j}\psi(\tau,c)\Big|_{\tau=\tau^*}\Big)_{i,j=1}^{\mu-2}$$

$$b = \Big(\sum_{j=1}^{r} c_j \frac{\partial^2}{\partial\tau_i\partial c_j}\psi(\tau,c)\Big|_{\tau=\tau^*}\Big)_{i=1}^{\mu-2}$$

$$\hat{J} = \begin{pmatrix} A & b \\ b^T & 0 \end{pmatrix} \in \mathbb{R}^{\mu-1\times\mu-1}. \tag{5.19}$$

(III) The design

$$\xi_{\hat\tau} = \begin{pmatrix} a & \hat\tau_1 & \dots & \hat\tau_{u-2} & b \\ \frac{1}{\mu} & \frac{1}{\mu} & \dots & \frac{1}{\mu} & \frac{1}{\mu} \end{pmatrix}$$

is the unique minimally supported standardized maximin D-optimal design with respect to the set Ω_δ^*.

(IV) The design $\xi_{\hat\tau}$ is the unique minimally supported standardized maximin D-optimal design with respect to the set Ω_δ.

For a proof of (I) and (II) we note that $h(u,\delta)$ is a real analytic vector valued function in a neighbourhood of the point $(u^*,\delta^*) = (0,0)$, with components satisfying

$$h_i(0,0) = \frac{\partial}{\partial u_i} h(u,\delta)\Big|_{(u,\delta)=(0,0)} = 0; \quad i = 1,\dots,\mu-1,$$

and

$$\Big(\frac{\partial}{\partial u_j} h_i(u,\delta)\Big)_{i,j=1}^{\mu-1} = \delta^2\hat{J} + O(\delta^3),$$

where the matrix $\hat{J}$ is defined in (5.19). Obviously,

$$\det\hat{J} = -(\det A) b^T A^{-1} b,$$

where $\det A \neq 0$ as demonstrated in the proof of Theorem 3.4 and 3.5. A similar argument shows that $b \neq 0$ and therefore the matrix $\hat{J}$ is non singular. The implicit function theorem [see Gunning and Rossi (1965)] now shows the existence of a unique real analytic solution $\hat{u}$ of (5.14) in a sufficiently small interval $(-\delta^*,\delta^*)$. The recursive relation (5.17) for the coefficients in the corresponding Taylor expansion is now a consequence of from Theorem 5.3 in Melas (2005).

In order to prove (III) we note that it follows from the uniqueness of the minimally supported local D-optimal design for any $\delta \in (0,1)$

$$\min_{0\le\alpha\le1} (1-\alpha)\frac{\psi(\tau,(1-\delta)c)}{\psi(\tau^*((1-\delta)c),(1-\delta)c)} + \alpha\frac{\psi(\tau,(1+\delta)c)}{\psi(\tau^*((1+\delta)c),(1+\delta)c)} < 1. \tag{5.20}$$

For $\delta \in [0,1]$ define as $(\tilde{\tau}, \tilde{\alpha})$ a point where the optimum in (5.11) is attained, that is

$$\varphi(\tilde{\tau}, \tilde{\alpha}, \delta) = \max_{\tau \in T} \min_{\alpha \in [0,1]} \varphi(\tau, \alpha, \delta).$$

If $\tilde{\alpha} = 0$ we would obtain

$$\varphi(\tilde{\tau}, \tilde{\alpha}, \delta) = \varphi(\tilde{\tau}, 0, \delta) = \max_{\tau \in T} \frac{\psi(\tau, (1-\delta)c)}{\psi(\tau^*((1-\delta)c), (1-\delta)c)} = 1,$$

which contradicts (5.20). Similary, we can exclude the case $\tilde{\alpha} = 1$. The matrix A in (5.18) is nonsingular and the Hesse matrix of the function $\psi(\tau, c)$ evaluated at the extreme point τ^* must be negative definite. Consequently, it follows that for sufficiently small δ the function $\varphi(\tau, \alpha, \delta)$ defined in (5.10) is a concave function of τ in a neighbourhood of the point τ^*. This means that $(\hat{\tau}, \hat{\alpha}) = (\tilde{\tau}, \tilde{\alpha})$ and consequently the design $\xi_{\hat{\tau}}$ is the unique minimally supported standardized maximin D-optimal design with respect to the set Ω_δ^*.

Finally, we prove assertion (IV), which follows from the equation

$$\min_{\lambda \in \Omega_\delta} \varphi(\hat{\tau}, \lambda) = \min_{\lambda \in \Omega_\delta^*} \varphi(\hat{\tau}, \lambda) \tag{5.21}$$

To prove (5.21) we define the rescaled quantities $\gamma_i = (\lambda_i - c_i)/(\delta c_i)$ $(i = 1, \ldots, r)$ and note that $|\gamma_i| \leq 1$ if $\lambda \in \Omega_\delta$. A straightforward but tedious calculation yields

$$\varphi(\hat{\tau}, \lambda) = 1 + \delta^2 \gamma^T B^T A B \gamma + O(\delta^3), \tag{5.22}$$

where $\gamma = (\gamma_1, \ldots, \gamma_r)^T$, $B = A^{-1} D$, the matrix D is defined by

$$D = \Big(\frac{\partial^2 h(\tau, c)}{\partial \tau_i \partial c_j} \Big|_{\tau = \tau^*} \Big)_{i=1,\ldots,\mu-2}^{j=1,\ldots,r},$$

and the elements of the matrix A^{-1} and D are negative and positive, respectively (this follows by similar arguments as given in Melas (2006), p. 56-57). Consequently, the elements of the matrix $D^T A^{-1} D$, say z_{ij} $(i, j = 1, \ldots, r)$, are negative and (5.22) yields

$$\varphi(\hat{\tau}, \lambda) = 1 + \delta^2 \sum_{i,j=1}^{r} z_{ij} \gamma_i \gamma_j + O(\delta^3).$$

Therefore, if δ is sufficiently small, the minimum of $\varphi(\hat{\tau}, \lambda)$ is attained if all components of $\gamma = (\gamma_1, \ldots, \gamma_r)$ have the same sign and are equal to $+1$ or -1. Consequently, the minimum is attained either at $\lambda = (1-\delta)c$ or $\lambda = (1+\delta)c$.

□

References

C. De Boor (1978). A practical guide to splines. Applied Mathematical Sciences. Vol. 27. New York-Heidelberg-Berlin: Springer-Verlag. XXIV.

K. Chaloner, I. Verdinelli (1995). Bayesian experimental design: a review. Statistical Science 10, 273-304.

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

H. Dette , V.B. Melas, A. Pepelysheff (2004). Optimal designs for estimating individual coefficients in polynomial regression - a functional approach. Journal of Statistical Planning and Inference 118, 201 - 219.

P. Dierckx (1995). Curve and surface fitting with splines. Monographs on Numerical Analysis. Oxford: Clarendon Press.

R. L. Eubank (1999). Nonparametric regression and spline smoothing. 2nd ed. Statistics: Textbooks and Monographs. 157. New York, NY: Marcel Dekker.

A. R. Gallant, W. A. Fuller (1973). Fitting segmented polynomial regression models whose join points have to be estimated. J. American Statist. Soc. 68, 144-147.

H. O. Hartley (1961). The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. Technometrics 3, 269-280.

B. Heiligers (1998). E-optimal designs for polynomial spline regression. J. Stat. Plann. Inference 75, 159-172.

B. Heiligers (1999). Experimental design for polynomial spline regression. Tatra Mt. Math. Publ. 17, 157-165.

P. G. Hoel (1958) Efficiency problems in polynomial estimation. Ann. Math. Stat. 29, 1134-1145.

D. J. Hudson (1966). Fitting segmented curves whose join points have to be estimated. J. American. Statist. Assoc. 61, 1097-1125.

L. A. Imhof (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. Ann. Statist. 29, 561-576.

D. L. B. Jupp (1978). Approximation to data by splines with free knots. SIAM J. Numer. Anal. 15, 328-343.

V. K. Kaishev (1989). Optimal experimental designs for the B-spline regression. Comput. Stat. Data Anal. 8, 39-47.

S. Karlin, W. J. Studden (1966). Tchebycheff Systems: with Applications in Analysis and Statistics. Interscience, New York.

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

J. Kiefer, J. Wolfowitz (1960). The equivalence of two extremum problems. Can. J. Math. 12, 363-366.

Y. B. Lim (1991). D-optimal designs in polynomial spline regression. The Korean Journal of applied Statistics 4, 171-178.

W. Mao, L. H. Zhao (2003). Free-knot polynomial splines with confidence intervals. J. Roy. Statist. Soc. 65, 901-919.

V.B.Melas (2005). On the functional approach to optimal designs for nonlinear models. Journal of Statistical Planning and Inference, 132, 93–116.

V.B.Melas (2006). Functional approach to optimal experimental design. Springer, Heidelberg.

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

V. N. Murty (1971a). Optimal designs with a Tchebycheffian spline regression function. Ann. Math. Stat. 42, 643-649.

V. N. Murty (1971b). Optimal designs with a polynomial spline regression with a single multiple knot at the center. Ann. Math. Stat. 42, 952-960.

S. H. Park (1978). Experimental designs for fitting segmented polynomial regression models. Technometrics 20, 151-154.

C. Reinsch (1967). Smoothing by spline functions. Numer. Math. 10, 177-183.

I. J. Schoenberg (1964). Spline functions and the problem of graduation. Proc. Nat. Acad. Sci. 52, 947-950.

G. A. F. Seber, C. J. Wild (1989). Nonlinear Regression. Wiley, N.Y.

W. J. Studden (1971). Optimal designs and spline regression. Optimizing Meth. Statist., Proc. Sympos. Ohio State Univ.

W. J. Studden, D. J. VanArman (1969). Admissible designs for polynomial spline regression. Ann. Math. Stat. 40, 1557-1569.

E. T. Wittaker (1923). On a new method of graduation. Proc. Edinburgh. Math. Soc. 41, 63-75.

D. Woods, S. Lewis (2006). All-bias designs for polynomial spline regression models. Aust. N.Z.J. Stat. 48, 49 -58.