Statistica Sinica **18**(2008), 1047-1062

OPTIMAL DESIGNS FOR FREE KNOT LEAST SQUARES SPLINES

Holger Dette[1], Viatcheslav B. Melas[2] and Andrey Pepelyshev[2]

[1]*Ruhr-Universität Bochum and* [2]*St. Petersburg State University*

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; 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 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.

Key words and phrases: Free knot least squares splines, D-optimal designs, nonlinear models, local optimal designs, robust designs, saturated designs.

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. Because a polynomial function possessing all derivatives at all locations is not flexible enough for approximating a curve with different degrees of smoothness at different locations, many authors propose fitting piecewise polynomials, or splines, to the data (see e.g., De Boor (1978), Dierckx (1995) or Eubank (1999), among many others). Smoothing splines owe their origin to Wittaker (1923) and have been further developed by Schoenberg (1964) and Reinsch (1967). As an alternative several authors propose least squares splines (see e.g., Hartley (1961), Gallant and Fuller (1973) or Eubank (1999), among many authors). If the knots are assumed 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 VanArman (1969), Studden (1971) Murty (1971), Park (1978), Kaishev (1989), Heiligers (1998), among others). While in numerical analysis splines are used to approximate functions, experimental design optimality considerations require that the assumed model be true.

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; several authors have worked on the nonlinear estimation problem (see e.g., Jupp (1978), Seber and Wild (1989) or Mao and Zhao (2003), among others). In particular - to the knowledge of the authors − optimal designs for least squares splines with estimated knots have not been investigated extensively in the literature. Pazman (2002) discussed the problem of optimal design for nonlinear models with constraints on the parameters, and put the free knot spline in this context. However, he did not determine optimal designs for free knot splines. Woods (2005) and Woods and Lewis (2006) considered a mean squared error criterion which addresses model misspecification when both the number and location of the knots are unknown, but did not consider the problem of estimating the knots.

The present paper is devoted to the D-optimal design problem for spline regression models with estimated knots; this is introduced in Section 2. In Section 3 we discuss local D-optimal designs; these 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; this does not require exact knowledge of the knots before any observations are available. Some theoretical results on minimally supported standardized maximin D-optimal designs are derived and can be used to construct these designs by means of Taylor expansions. The results are illustrated by several examples, while the more technical details are presented in an on-line supplement.

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 $[a, b]$, $\lambda_1 < \lambda_2 < \cdots < \lambda_r$ denote r knots located in the interval $[a, b]$, $k_i \leq m - 1 \quad (i = 1, \ldots, r)$, $k \leq m+1$, 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 $z_+ = \max\{0, z\}$. Note that the model (2.1) is nonlinear in the parameters $\lambda = (\lambda_1, \ldots, \lambda_r)^T$, and linear in the remaining parameters $\theta = (\theta_1, \ldots, \theta_k, \theta_{10}, \ldots, \theta_{rk_r-1})^T$ (see Seber and Wild (1989)). Note also that we use the truncated power basis, but there are other bases for which similar results as ours could be derived. We prefer to work with the truncated power basis, because of its similarity to that of ordinary polynomials.

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 $n_j = Nw_j$, taken at each point x_j; $(j = 1, \ldots, n)$ (see e.g., Pukelsheim (1992)). Under the assumption of normality, the covariance matrix of the maximum likelihood estimate of the parameters (θ, λ) is approximately $(\sigma^2/N)(C_\theta M(\xi, \lambda) C_\theta^T)^{-1} \in \mathbb{R}^{\mu \times \mu}$, where $C_\theta \in \mathbb{R}^{\mu \times \mu}$ denotes a nonsingular matrix that depends on the parameters $\theta_{10}, \ldots, \theta_{rk_r-1}$, but not on the knots $\lambda_1, \ldots, \lambda_r$ nor the design ξ. Here $\mu = k + \sum_{i=1}^r k_i + r$ is the number of parameters, $M(\xi, \lambda) = \int_a^b f(x, \lambda) f^T(x, \lambda) d\xi(x)$ is the information matrix of the design ξ, and the components of $f(x, \lambda) = (f_1(x, \lambda), \ldots, f_\mu(x, \lambda))^T$ are defined by

$$f_\ell(x, \lambda) = \begin{cases} x^{\ell-1}; & \ell = 1, \ldots, k \\ (x - \lambda_1)_+^{m+\alpha_0-\ell+1}; & \ell = \alpha_0 + 1, \ldots, \alpha_1 \\ (x - \lambda_2)_+^{m+\alpha_1-\ell+1}; & \ell = \alpha_1 + 1, \ldots, \alpha_2 \\ \vdots & \\ (x - \lambda_r)_+^{m+\alpha_{r-1}-\ell+1}; & \ell = \alpha_{r-1} + 1, \ldots, \alpha_r \end{cases} \tag{2.2}$$

$(\ell = 1, \ldots, \mu)$, and $\alpha_j = k + \sum_{s=1}^j (k_s + 1)$ for $j = 0, \ldots, r$. Usually optimal or efficient designs maximize an appropriate function of the Fisher information matrix. In our particular model 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)); in the present paper we concentrate on D-optimal designs that minimize the determinant of the matrix in $(C_\theta M(\xi, \lambda) C_\theta^T)^{-1}$. This is equivalent to minimizing the determinant of the matrix $M^{-1}(\xi, \lambda)$, because the matrix C_θ does not depend on ξ.

Following Chernoff (1953), we call a design $\xi^*_{D,\lambda}$ local D-optimal if it maximizes $\det M(\xi, \lambda)$. For the case of least squares estimation with given knots, D-optimal designs have been considered by Park (1978), Kaishev (1989) and Lim (1991). Pazman (2002) discussed the problem of optimal design for nonlinear models with constraints on the parameters, and put the free knot spline in this context. However, no explicit 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 the use of 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 is difficult to specify such a prior in a concrete situation. As an alternative, we propose a maximin approach based on the D-optimality criterion; this only requires the specification of a certain range for the unknown knots of the spline regression model. The method determines a design that maximizes a minimum of D-efficiencies (see Müller (1995), Dette (1997) and Imhof (2001)), and is motivated by the fact that, in the case of free knot least squares splines, it is difficult to specify an r-dimensional prior for the (unknown) knots before any data is available.

A standardized maximin D-optimal design maximizes

$$\min_{\lambda \in \Omega} \frac{\det M(\xi, \lambda)}{\det M(\xi^*_{D,\lambda}, \lambda)}, \tag{2.3}$$

where $\Omega \subset \{z = (z_1, \ldots, z_r)^T \in \mathbb{R}^r \mid a < z_1 < \cdots < z_r < b\}$ is a given compact set for the knots $\lambda_1, \ldots, \lambda_r$, and $\xi^*_{D,\lambda}$ is the local D-optimal design for a fixed parameter λ. We 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 (2.3) 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 (such 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 $k_i = m - 1; i = 1, \dots, r$. There exists a unique local D-optimal design $\xi^*_{D,\lambda}$ with exactly μ support points, say $x_1 < \cdots < x_\mu$ and equal weights $\xi^*_{D,\lambda}(x_j) = 1/\mu,\ j = 1, \dots, \mu$. The support points are given by*

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

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

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 to all polynomial of degree less or equal than $s - 1$ with respect to Lebesgue measure.

Proof of Theorem 3.1. Let $x^*_1, \dots, x^*_n$ and $w^*_1, \dots, w^*_n$ denote the support points and corresponding weights of a local D-optimal design $\xi^*_{D,\lambda}$ for least squares estimation in the nonlinear model (2.1). 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 $(\lambda_j, \lambda_{j+1}]$ $\quad (j = 1, \dots, r)$, because otherwise the determinant of the corresponding information matrix would vanish. Moreover, the equivalence theorem of Kiefer and Wolfowitz (1960) shows that $\xi^*_{D,\lambda}$ is local D-optimal if and only if the inequality

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

holds for all $x \in [a, b]$, where the vector of regression functions is defined by (2.2). 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

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

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 $F = \mathrm{diag}(F_1, F_2, \ldots, F_{r+1})$ denotes a block triangular matrix with blocks in the diagonal given by $F_1 = (f_i(x_j^*, \lambda))_{i,j=1}^k \in \mathbb{R}^{k\times k}$, $F_\ell = (f_i(x_j^*, \lambda))_{i,j=k+(\ell-1)m+1}^{k+\ell m}$ $\in \mathbb{R}^{m\times m}$, $\ell = 2, \ldots, r+1$. As a consequence, we obtain from (3.7) that $\det M(\xi_{D,\lambda}^*, \lambda) = (1/\mu)^\mu \prod_{j=1}^{r+1} (\det F_j)^2$, 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 is 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 $z_j = x_{k+(\ell-1)m+j}^*$ $(j = 1, \ldots, m-1; \ell = 1, \ldots, r)$. Now the results of Hoel (1958) show again that this expression is maximized if $z_1, \ldots, z_{m-1}$ 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 $k = 3; m = 2; k_j = 1 \quad (j = 1, \ldots, r)$. According to Theorem 3.1 the local D-optimal design is given by $(\lambda_0 = a, \lambda_{r+1} = b)$

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

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 $\xi^* = \{0, x_2^*(\lambda), \ldots, x_5^*(\lambda), 1; 1/6, \ldots, 1/6\}$.

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 splines with estimated knots have to be found numerically. Consider, as a typical example, a cubic spline regression model (with $r = 1$ and continuous first and second derivative)

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

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.

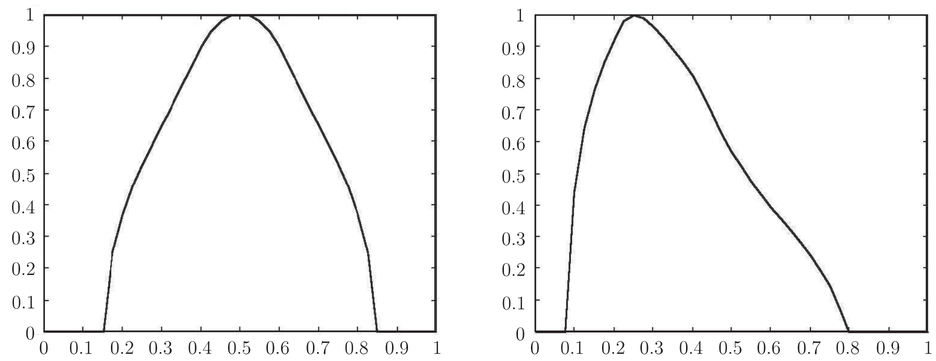

Figure 3.1. The D-efficiencies of the local D-optimal design in the spline regression model (3.8), where $\lambda = 0.5$ (left panel) and $\lambda = 0.25$ (right panel).

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 a symmetry result.

Theorem 3.3. *Consider the spline regression model* (2.1) *with knots* $\lambda = (\lambda_1, \ldots, \lambda_r)^T$ *and let* $\xi^*_{D,\lambda}$ *denote a local* D*-optimal design with masses* $w^*_1, \ldots, w^*_n$ *at the points* $x^*_1, \ldots, x^*_n$*, respectively. The design* $\tilde{\xi}_{D,\lambda}$ *with masses* $w^*_1, \ldots, w^*_n$ *at the points* $\tilde{x}_1, \ldots, \tilde{x}_n$ *with* $\tilde{x}_i = b - a - x^*_i \quad (i = 1, \ldots, n)$ *is local* D*-optimal for least squares spline estimation in the model* (2.1), *with knots* $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_r)^T = (b - a - \lambda_r, \ldots, b - a - \lambda_1)^T$.

Proof of Theorem 3.3. The assertion follows from a basic property of the D-optimality criterion observing that the functions $1, b-a-x, \ldots, (b-a-x)^{k-1}, (x-\tilde{\lambda}_1)_+^{m-k_1}, \ldots (x - \tilde{\lambda}_1)_+^m, \ldots, (x - \tilde{\lambda}_r)_+^{m-k_r}, \ldots, (x - \tilde{\lambda}_r)_+^m$ and $1, x, \ldots, x^{k-1}, (x - \lambda_1)_+^{m-k_1}, \ldots, (x-\lambda_1)_+^m, \ldots, (x-\lambda_r)_+^{m-k_r}, \ldots, (x-\lambda_r)_+^m$ 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, provided

$$m - k - 2 \le k_1 = k_2 = \ldots = k_r \le m - 1\,. \tag{3.11}$$

The proofs are complicated and are therefore presented in an on-line supplement.

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 $a = \lambda_0 < \lambda_1 < \cdots < \lambda_r < \lambda_{r+1} = b$, are strictly increasing functions of any component of the vector $\lambda = (\lambda_1, \ldots, \lambda_r)^T$. 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*

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

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 $a = \lambda_0 < \lambda_1 < \cdots < \lambda_r < \lambda_{r+1} = b$, 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 is 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 ξ_{τ^*} is unique and has equal masses $1/\mu$ at the points $a, \tau^*_1, \ldots, \tau^*_{\mu-2}, b$, 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 $\tau^*_{(s)} = (1/s!)(d^s/d\lambda^s)\tau^*(\lambda)\Big|_{\lambda=\lambda_0}$; $s = 0, 1, 2, \ldots$ The region of convergence of this series depends on the specific model under consideration (see Example 3.6). Moreover, the coefficients in the expansion can be calculated recursively (see Melas (2006)) using the recursive relations

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

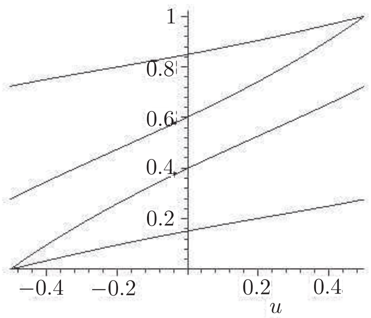

Figure 3.3. The interior points $\tau_j^* = \tau_j^*(u)$ of the local D-optimal design for the spline regression model (3.10), considered as a function of the parameter u.

where $J_{(0)} = \big(\frac{\partial^2}{\partial\tau_i\partial\tau_j}\psi(\tau,\lambda_0)\big|_{\tau=\tau^*_{(0)}}\big)_{i,j=1}^{\mu-2}, g(\tau,\lambda) = \big(\frac{\partial^2}{\partial\tau_i\partial\lambda}\psi(\tau,\lambda)\big)_{i=1}^{\mu-2}, \psi(\tau,\lambda) = \big\{\det M(\xi_\tau,\lambda)\big\}^{1/\mu}$, and we define, for any (sufficiently differentiable) function h,

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

We finally note that in the case of at least two knots, an extension of formula (3.13) is given in Melas (2006). 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 $x_2^*(\lambda) = 1 - x_5^*(1-\lambda), x_3^*(\lambda) = 1 - x_4^*(1-\lambda)$ (see Theorem 3.3). In the following we construct Taylor expansions for the support points of the local D-optimal design at the point $\lambda = 0.5$. The radius of convergence for this series is $\rho = 0.5$, which was obtained by numerical calculations. In principle one can construct coefficients corresponding to an arbitrary large order in the expansion, using the recursive relations. For the determination of the support points with a precision 10^{-2} an expansion of order 10 is sufficient in the present example. With the notation $\tau_i^* = x_{i+1}^*$ $(i = 1,\dots,4)$, $u = \lambda - 0.5$ we obtain

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

$\tau_3^*(u) = 1 - \tau_2^*(-u), \tau_4^*(u) = 1 - \tau_1^*(-u)$. 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). We finally note that if $\lambda \to 0$, then $u \to -0.5$ and $\lim_{u \to -0.5}(\tau_1(u), \tau_2(u), \tau_3(u), \tau_4(u)) = (0, 0, x^*, y^*)$, where x^*, y^* maximize the function $xy(y-x)(1-x)(1-y)$ in the set $\{(x, y)|0 < x < y < 1\}$. Similarly, if $\lambda \to 1$, it follows by symmetry that $\lim_{u \to 0.5}(\tau_1(u), \tau_2(u), \tau_3(u), \tau_4(u)) = (y^*, x^*, 1, 1)$.

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 to the specification of the knots. The standardized maximin optimality criterion (2.3) 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.

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 points 0 and 1 in its support (see the proof of Lemma 5.2 in the on-line supplement). In the following we consider the set $\Omega = [u, 1-u]$ with $u \in (0, 1/2)$ in the optimality criterion (2.3), then it follows by similar arguments as given in the proof of Theorem 3.3 that the minimally supported standardized maximin D-optimal design ξ^* has masses 1/5 at the points $0, x, 1/2, 1-x, 1$, where $x \in (0, 1/2)$. Consequently, the optimality criterion (2.3) 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 $x^*(u) = 3/16 + 3/8u - \sqrt{(6u-3)^2 + 8u}/16$. 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 as the length of the interval increases.

In the remaining part of this section we discuss the numerical construction of minimally supported standardized maximin D-optimal designs. In order to

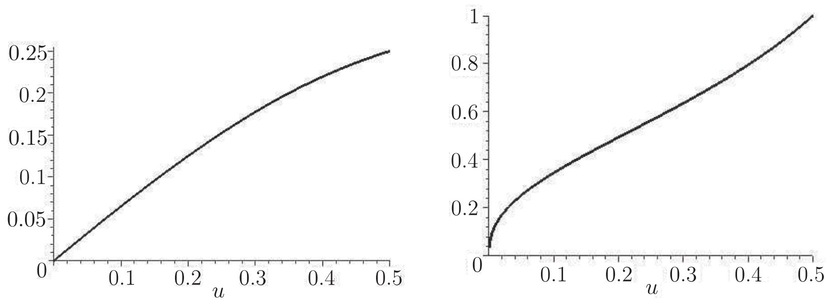

Figure 4.1. The non-trivial support point $x^* = x(u)$ of the minimally supported standardized maximin D-optimal design (left panel) for the spline regression model (4.1) and its minimal efficiency in the interval $[u, 1-u]$ (right panel).

derive a Taylor expansion for such designs we consider the following set Ω in the optimality criterion (2.3):

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

where $c = (c_1, \ldots, c_r)^T$, with $c_1 < \cdots < c_r$, 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 (2.3) 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 presented in an on-line supplement.

Theorem 4.2.

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

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

where $\bar{\Omega}_\delta \in \mathbb{R}^r$ is a two point set defined by $\bar{\Omega}_\delta = \{(1-\delta)c, (1+\delta)c\}$.

(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 was illustrated in Section 3.3 for the case of local D-optimal designs. The corresponding recursive relations are obtained, by a slight modification, from those presented in Section 3.3. The details are omitted for the sake of brevity. It is difficult to determine δ^* theoretically, but in all our numerical examples we observed that $\delta^* = 1$. We conclude this section with a continuation of Example 4.1.

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

Example 4.3. The concrete values for the support points of the minimally supported standardized maximin D-optimal designs for the spline regression model (4.1) 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 the Taylor expansion, as described before. 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 efficiency only 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. It is also worthwhile to mention that for fixed u the efficiencies, as function of v, are nearly linear, and a similar statement can be made if the efficiency is considered as a function of v, where $u = 1 - v$.

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 (2007) that the number of support points of the standardized maximin D-optimal design becomes arbitrary large if $v - u \to 1$. These designs are calculated numerically, increasing successively the number of support points of the designs under consideration until there is no further improvement in the value of the optimality criterion. 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 14 support points. However designs 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 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. The authors would also like to thank the reviewers for their constructive comments on an earlier version of this manuscript.

References

Braess, D. and Dette, H. (2007). On the number of support points of maximin and Bayesian optimal designs. *Ann. Statist.* **35**, 772-792.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statist. Sci.* **10**, 273-304.

De Boor, C. (1978). A Practical Guide to Splines. Applied Mathematical Sciences, Vol. 27, Springer-Verlag, New York-Heidelberg-Berlin.

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

Dierckx, P. (1995). Curve and Surface Fitting with Splines. Monographs on Numerical Analysis. Clarendon Press, Oxford.

Eubank, R. L. (1999). Nonparametric Regression and Spline Smoothing. 2nd ed., Marcel Dekker, New York.

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

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

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

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

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

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

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

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

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

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canad. J. Statist.* **12**, 363-366.

Lim, Y. B. (1991). *D*-optimal designs in polynomial spline regression. *Korean J. Appl. Statist.* **4**, 171-178.

Mao, W. and Zhao, L. H. (2003). Free-knot polynomial splines with confidence intervals. *J. Roy. Statist. Soc.* **65**, 901-919. **132**, 93-116.

Melas, V. B. (2006). Functional Approach to Optimal Experimental Design. Springer, Heidelberg.

Müller, Ch. H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *J. Statist. Plann. Inference* **44**, 117-132.

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

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

Pazman, A. (2002). Optimal design of nonlinear experiments with parameter constraints. *Metrika* **56**, 113-130.

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

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763-770.

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

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

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

Silvey, S. D. (1980). *Optimum Design*, Chapman & Hall, London.

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

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

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

Woods, D. (2005). Designing experiments under random contamination with applications to polynomial spline regression. *Statist. Sinica* **15**, 619-633

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

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany.

E-mail: holger.dette@rub.de

Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia.

E-mail: v.melas@pobox.spbu.ru

Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia.

E-mail: andrey@ap7236.spb.edu

(Received September 2006; accepted January 2007)