# Optimal designs for smoothing 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


August 10, 2007


**Abstract**

In the common nonparametric regression model we consider the problem of constructing optimal designs, if the unknown curve is estimated by a smoothing spline. A new basis for the space of natural splines is derived, and the local minimax property for these splines is used to derive two optimality criteria for the construction of optimal designs. The first criterion determines the design for a most precise estimation of the coefficients in the spline representation and corresponds to $D$-optimality, while the second criterion is the $G$-criterion and corresponds to an accurate prediction of the curve. Several properties of the optimal designs are derived. In general $D$- and $G$-optimal designs are not equivalent. Optimal designs are determined numerically and compared with the uniform design.



Keyword and Phrases: Smoothing spline, nonparametric regression, $D$- and $G$-optimal designs, saturated designs

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

# 1 Introduction

Consider the common nonparametric regression model on the interval $[a, b]$

$$Y_i = g(t_i) + \varepsilon_i \ , \ i = 1, \ldots, n; \tag{1.1}$$

where $a \leq t_1 < \ldots < t_n \leq b$ are the design points, the errors are independent identically distributed with mean 0 and variance $\sigma^2 > 0$. There are several procedures to estimate the unknown regression function $g$ nonparametrically, including kernel type, series estimators and polynomial splines [see e.g. the monographs of Fan and Gijbels (1996) or Efromovich (1999)]. Because of its similarity to polynomials and its conceptual simplicity many authors propose to fit polynomial splines to the data [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); see also the monographs of Eubank (1999) and Wahba (1990). The basic idea of this estimate is rather simple. Because the minimization of the residual sum of squares

$$\sum_{i=1}^{n} (Y_i - g(t_i))^2 \tag{1.2}$$

with respect to the function $g$ would yield an interpolating curve with too many rapid fluctuations, a roughness penalty is introduced, which restricts the class of plausible curves with respect to their smoothness properties. More precisely, if $W^{(m)}([a, b])$ denotes the Sobolev space of all $m$ times continuously differentiable functions defined on the interval $[a, b]$ with $\int_a^b |g^{(m)}(t)|^2 dt < \infty$, then the sum of squares in (1.2) is minimized in the class

$$\mathcal{F}_\rho = \Big\{ g \in W^{(m)}([a, b]) \mid \int_a^b |g^{(m)}(t)|^2 dt \leq \rho^2 \Big\}, \tag{1.3}$$

where $\rho > 0$ is a given roughness penalty or smoothing parameter, that is

$$\min_{g \in \mathcal{F}_\rho} \sum_{i=1}^{n} (Y_i - g(t_i))^2. \tag{1.4}$$

Introducing a Lagrange multiplier it can be shown that this constrained optimization problem is equivalent to minimizing

$$\sum_{i=1}^{n} (Y_i - g(t_i))^2 + \lambda \int_a^b |g^{(m)}(t)|^2 dt \tag{1.5}$$

for some well defined constant $\lambda > 0$. The case $m = 2$ corresponds to smoothing cubic splines, which have been extensively studied and widely used because of the availability of fast and efficient algorithms for its calculation [see Reinsch (1967), Silverman (1985), Eubank (1999) or Green and Silverman (1994) among many others]. While many statistical properties of these estimates have been considered in the literature the problem of designing experiments for non-parametric estimation with smoothing splines has - to the knowledge of the authors - not been investigated so far.

Design problems for spline estimation with a well defined truncated power basis have a long history. If the knots are assumed to be fixed and the basis is given, the estimation problem reduces to a linear regression problem and optimal designs have been investigated by Studden and Van Arman (1969), Studden (1971), Murty (1971a,b), Park (1978), Kaishev (1989), Heiligers (1998, 1999). If the knots are also estimated from the data, then the resulting splines are called free knot splines and the estimation problem is a nonlinear least squares problem [see Jupp (1978) of Mao and Zhao (2003)]. The construction of $D$-optimal designs for splines with estimated knots was considered recently by Dette, Melas and Pepelyshev (2007). Other types of optimal designs for the estimation with splines taking the bias into account have been discussed by Woods (2005) and Woods and Lewis (2006).

On the other hand no optimal designs are available for estimation with the smoothing spline. This gap can be partially explained by the particular difficulties, which emerge from the implicit definition of the basis and the knots in the solution of the optimization problem (1.4). The goal of the present paper is to fill this gap and present some useful tools for constructing optimal designs for the estimation of the regression with smoothing splines. In Section 2 we review some basic terminology and recall a local minimax property of the smoothing spline. The value of the corresponding minimax criterion will be the basis for the definition of optimal design problems. In Section 3 we introduce a new basis for the space of natural splines, which is of own interest and fundamental for the solution of the optimal design problems discussed in Section 4. Two optimality criteria for the determination of optimal designs are introduced corresponding a most precise estimation of the coefficients in the spline representation ($D$-optimality) and an accurate prediction of the curve ($G$-optimality). Several properties of the optimal designs are derived. In particular – in contrast to approximate design theory [see Kiefer (1974)] – $D$- and $G$-optimal designs are not equivalent. Finally, some numerical results and a comparison of the optimal design with the commonly used uniform design in this context are presented in Section 5. The numerical study also includes an $I$-optimality criterion which is more appropriate for mean squared error considerations.

# 2 The local minimax property

It is well known that in the case $n \geq m$ there exists a unique solution of the constrained optimization problem (1.4) or (1.5), which is a natural spline [see e.g. Eubank (1999), Theorem 5.3]. The set of natural splines, say $N^{2m}(t_1, \dots, t_n)$, is defined as the set of all functions $g$ defined on the interval $[a, b]$ with the following properties

(i) $g$ is a piecewise polynomial of order $2m - 1$ on any subinterval $(t_i, t_{i+1}); i = 1, \dots, n - 1$.

(ii) $g$ has $2m-2$ continuous derivatives.

(iii) $g$ has $2m-1$ derivatives.

(iv) $g$ is a polynomial of degree $m-1$ on the intervals $(-\infty, t_1)$ and $(t_n, \infty)$.

The dimension of the vector space $N^{2m}(t_1, \dots, t_n)$ is precisely $n$ [see Eubank (1999)], and if $\varphi_1, \dots \varphi_n$ denotes a basis of $N^{2m}(t_1, \dots, t_n)$, then the unique smoothing spline has a local minimax property, which will be recalled here, because it is essential for the construction of optimal designs. For this purpose we consider the $n \times n$ matrix

$$X = X(T) = (\varphi_i(t_j))_{i,j=1,\dots,n} \tag{2.1}$$

(here the notation $X(T)$ reflects the fact that the basis functions depend on the knots $T = \{t_1, \dots, t_n\}$ and is used whenever it is necessary to emphazise this dependence) and define the vector $\varphi(t) = (\varphi_1(t), \dots, \varphi_n(t))^T$. If

$$g(t) = \theta^T \varphi(t) = \sum_{j=1}^{n} \theta_j \varphi_j(t) \tag{2.2}$$

denotes a natural spline, then it is easy to see that the defining constraint (1.3) for the set $\mathcal{F}_\rho$ is $\theta^T B \theta \le \rho^2$, where the matrix $B$ is given by

$$B = \Big( \int_a^b \varphi_i^{(m)}(t) \varphi_j^{(m)}(t) dt \Big)_{i,i=1,\dots,n}. \tag{2.3}$$

In what follows we consider the set

$$\Omega = \{\theta \in \mathbb{R}^n \mid \theta^T B \theta \le \rho^2\} \tag{2.4}$$

of all vectors $\theta$ corresponding to a function in $N^{2m}(t_1, \dots, t_n) \cap \mathcal{F}_\rho$ and discuss for $k = 1, \dots, n$ the following minimax problem

$$\inf_{a \in \mathbb{R}^n} \sup_{g \in \mathcal{F}_\rho} E[(g(t_k) - a^T Y)^2], \tag{2.5}$$

where $Y = (Y_1, \dots, Y_n)^T$ denotes the vector of all observations. The following Lemma can be found in Eubank (1999). We will present a proof here, because we need the optimal value of the minimax problem (2.5) for the definition of the optimality criteria, and could not find this in the literature.

**Lemma 2.1.** *Assume that $n > m$ and $k \in \{1, \dots, n\}$. The solution of the minimax problem (2.5) is given by*

$$a^* = X(X^T X + \lambda B)^{-1} \varphi(t_k), \tag{2.6}$$

*and the minimum value is given by*

$$\inf_{a\in\mathbb{R}^n} \sup_{g\in\mathcal{F}_\rho} E[(g(t_k)-a^TY)^2] = \sigma^2\varphi^T(t_k)(X^TX+\lambda B)^{-1}\varphi(t_k), \tag{2.7}$$

*where* $\lambda = \sigma^2/\rho^2$.

**Proof.** For fixed $k \in \{1,\dots,n\}$ we define for any vector $a \in \mathbb{R}^n$ the function

$$H(a) = \sup_{g\in\mathcal{F}_\rho} E[(g(t_k)-a^TY)^2] \tag{2.8}$$

and note that this function depends on the class $\mathcal{F}_\rho$ only through the points $t_1,\dots,t_n$. Consequently, it follows from well known properties of natural splines [see e.g. Karlin and Studden (1966), Section 11.9 ] that the supremum in (2.8) is attained at a function $g^* \in N^{2m}(t_1,\dots,t_n)\cap\mathcal{F}_\rho$, that is

$$H(a) = \sup_{g\in N^{2m}(t_1,\dots,t_n)\cap\mathcal{F}_\rho} E[(g(t_k)-a^TY)^2]\ . \tag{2.9}$$

Note that all functions $g \in N^{2m}(t_1,\dots,t_n)\cap\mathcal{F}_\rho$ are of the form (2.2) for some $\theta\in\Omega$, and define

$$\mathcal{L} = \Big\{ \frac{\varphi(t_k)a^T}{\varphi^T(t_k)\varphi(t_k)} \Big| \, a \in \mathbb{R}^n \Big\} \subset \mathbb{R}^{n\times n}.$$

Obviously each vector $a\in\mathbb{R}^n$ can be represented as $a = L^T\varphi(t_k)$ with $L\in\mathcal{L}$, and we obtain

$$\begin{aligned}
\inf_{a\in\mathbb{R}^n} H(a) &= \inf_{L\in\mathcal{L}}\sup_{\theta\in\Omega} E[\{\varphi^T(t_k)(\theta-LY)\}^2] \geq \inf_{S\in\mathbb{R}^{n\times n}}\sup_{\theta\in\Omega} E[\{\varphi^T(t_k)(\theta-SY)\}^2] \\
&= \sigma^2\varphi^T(t_k)(X^TX+\lambda B)^{-1}\varphi(t_k),
\end{aligned}$$

where $\lambda$ is defined by $\lambda = \sigma^2/\rho^2$ and the last equality follows from Kuks and Olman (1971) with $S^* = (X^TX+\lambda B)^{-1}X^T$ [see also Toutenburg (1982), Chap. 4]. Note that

$$\varphi^T(t_k)S^* = \varphi^T(t_k)\frac{\varphi(t_k)\varphi^T(t_k)S^*}{\varphi^T(t_k)\varphi(t_k)},$$

and that $\varphi(t_k)\varphi^T(t_k)S^*/\varphi^T(t_k)\varphi(t_k) \in \mathcal{L}$. Consequently we have

$$\inf_{a\in\mathbb{R}^n} H(a) = \sigma^2\varphi^T(t_k)(X^TX+\lambda B)^{-1}\varphi(t_k),$$

and the infimum is attained for the vector $a^*$ defined in (2.6). □

The value of the minimax criterion (2.6) will be the basic criterion for constructing optimal designs for estimation with smoothing splines. More precisely, note that the matrix

$$\sigma^2(X^TX+\lambda B)^{-1} \tag{2.10}$$

is the analogue of the Fisher information matrix and depends on the design points $T = \{t_1, \dots, t_n\}$. Therefore a good design, specified by an appropriate choice of $T$, should maximize a real valued function of the matrix

$$M(T) = X^T X + \lambda B \tag{2.11}$$

[see Silvey (1980) or Pukelsheim (1993)]. However, in contrast to classical design theory for regression, the functions $\varphi_i$ defining the basis of the set of natural splines $N^{2m}(t_1, \dots, t_n)$ also depend on the design points. Therefore it is not clear that the "minimization" of a real valued function of the matrix (2.10) will yield a "small" value for the optimum in (2.8), and alternative optimality criteria could be used to reflect this dependence more appropriately. For the sake of brevity we restrict ourselves in this paper to one optimality criterion which depends only on the matrix (2.11) and one criterion, which takes also into account that the functions $\varphi_i$ depend on the design points.

To be precise recall that a design $T = \{t_1, \dots, t_n\}$ is called $D$-optimal if it maximizes the determinant $\det M(T)$. Similarly a design is called $G$-optimal if it minimizes the expression

$$\max_{t \in [a,b]} \varphi^T(t, T) M^{-1}(T) \varphi(t, T), \tag{2.12}$$

where the notation $\varphi(t, T) = \varphi(t)$ reflects the fact that the vector basis functions depend also on the design points. Note that in classical approximate design theory $D$- and $G$-designs are identical [see Kiefer and Wolfowitz (1960)], but this is not necessarily true in the present context, because on the one hand we do not consider approximate designs here, and on the other hand the functions $\varphi_i$ in (2.12) also depend on the design points. In the following section we construct a special basis for $N^{2m}(t_1, \dots, t_n)$, which will be useful for deriving some properties of the optimal designs with respect to the two criteria.

# 3 A new basis for natural splines

There are several bases, which could be used in this context [see for example Eubank (1999), Section 5.3.3], and it is worthwhile to mention again that basis functions depend on the knots $t_1, \dots, t_n$. As pointed out in the previous section the optimal design problem consists in the determination of the points $t_1, \dots, t_n$, such that an efficient estimate is obtained by the smoothing spline. In this section we will present a new basis of $N^{2m}(t_1, \dots, t_n)$, which is particularly useful for this purpose. We could not find this basis in the literature and therefore it might be also of its own interest. To be precise consider the nonparametric regression model (1.1) and define

$$F = F(T) = \begin{pmatrix} 1 & t_1 & \dots & t_1^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & \dots & t_n^{n-1} \end{pmatrix} \in \mathbb{R}^{n \times n}, \tag{3.1}$$

$$d_i = d_i(T) = (\det F)(F^T)^{-1} e_{m+i}\ ,\ i = 1, \dots, n-m; \tag{3.2}$$

where $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$ denotes the $j$th unit vector in $\mathbb{R}^n$ $(j = 1, \dots, n)$. Finally, we introduce the vector

$$\psi(t) = \psi(t, T) = ((t - t_1)_+^{2m-1}, \dots, (t - t_n)_+^{2m-1})^T. \tag{3.3}$$

**Theorem 3.1.** *The functions*

$$\varphi_1(t) = 1 \ , \ \varphi_2(t) = t \ , \ \dots \ , \ \varphi_m(t) = t^{m-1} \tag{3.4}$$

$$\varphi_{m+1}(t) = d_1^T \psi(t) \ , \ \dots \ , \ \varphi_n(t) = d_{n-m}^T \psi(t) \tag{3.5}$$

*define a basis of the set of natural splines $N^{2m}(t_1, \dots, t_n)$.*

**Proof.** Recalling the definition of the natural splines [see for example Eubank (1999), Chapter 5] the assertion of Theorem 3.1 follows if the following two assertions can be established:

(a) The functions $\varphi_{m+1}, \dots \varphi_n$ are polynomials of degree $m - 1$ on the interval $[t_n, \infty)$

(b) The functions $\varphi_1, \dots \varphi_n$ are linearly independent on the interval $[t_1, t_n]$.

We begin with a proof of assertion (a). Note that for $t > t_n$ the functions $\varphi_{m+i}$ are of the form

$$\begin{aligned} \varphi_{m+i}(t) &= \sum_{j=1}^n d_{ij}(t - t_j)^{2m-1} = \sum_{j=1}^n \sum_{\ell=0}^{2m-1} d_{ij} \binom{2m-1}{\ell} t^\ell t_j^{2m-1-\ell} \\ &= \sum_{\ell=0}^{2m-1} \binom{2m-1}{\ell} \Big\{ \sum_{j=1}^n d_{ij} t_j^{2m-1-\ell} \Big\} t^\ell \ , \end{aligned}$$

where $d_{ij}$ denotes the $j$th component of the vector $d_i = (d_{i1}, \dots, d_{in})^T$ defined in (3.2). Consequently the functions $\varphi_{m+i}$ are polynomials of degree $2m - 1$. Moreover, if $\ell \geq m$ we have for the coefficients of the monomials $t^\ell$

$$\sum_{j=1}^n d_{ij} t_j^{2m-1-\ell} = d_i^T F e_{2m-1-\ell} = (\det F) e_{m+i}^T e_{2m-1-\ell} = 0,$$

which proves assertion (a). For a proof of (b) we show that the determinant of the matrix $X = (\varphi_i(t_j))_{i,j=1,\dots,n}$ with the functions $\varphi_i$ defined in (3.4) and (3.5) is not vanishing. Obviously for this basis the matrix $B$ in (2.3) is of the form

$$B = \begin{pmatrix} 0 & 0 \\ 0 & V \end{pmatrix}$$

(here 0 denotes a matrix of appropriate dimensions with all entries equal to 0), where the matrix $V$ is given by

$$V = \Big( \int_a^b \varphi_{m+i}^{(m)}(t) \varphi_{m+j}^{(m)}(t) dt \Big)_{i,i=1,\dots,n-m} \in \mathbb{R}^{n-m \times n-m}.$$

Define for $a \le s < t \le b$

$$Q(s,t) = \int_a^b (s-u)_+^{m-1}(t-u)_+^{m-1}du, \tag{3.6}$$

then it is known that the matrix

$$Q_n = (Q(t_i,t_j))_{i,j=1,\ldots,n} \tag{3.7}$$

is positive definite [see Eubank (1999), p. 208]. With the notation $D = (d_{ij})_{i=1,\ldots,n-m}^{j=1,\ldots,n} \in \mathbb{R}^{n-m\times n}$ we have

$$V = DQ_nD^T.$$

Note that by Cramer's rule the vectors $d_i$ are the rows of the matrix $(\det F)F^{-1}$ and as consequence linearly independent. Therefore we have $\det V > 0$, which shows that the matrix $B$ has rank $n-m$. Moreover, Lemma 5.1 in Eubank (1999) yields,

$$\int_a^b \varphi_{m+i}^{(m)}(t)\varphi_{m+j}^{(m)}(t)dt = (-1)^m(2m-1)!\sum_{\ell=1}^n d_{j\ell}\varphi_{m+i}(t_\ell),$$

which gives $V = \kappa DRD^T$, where $\kappa = (-1)^m(2m-1)!$ and

$$R = \Big(\frac{|t_i-t_j|^{2m-1}}{2}\Big)_{i,j=1}^n .$$

Recalling the definition of

$$X = \begin{pmatrix} 1 & t_1 & \ldots & t_1^{m-1} & \varphi_{m+1}(t_1) & \ldots & \varphi_n(t_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \ldots & t_n^{m-1} & \varphi_{m+1}(t_n) & \ldots & \varphi_n(t_n) \end{pmatrix} \in \mathbb{R}^{n\times n},$$

we obtain for its determinant by Laplace's identity

$$\det X = \sum_{1\le\alpha_1<\ldots<\alpha_m\le n} (-1)^{\sum_{i=1}^m \alpha_i + m(m+1)/2} \det(t_{\alpha_i}^{j-1})_{i,j=1}^m \det(\varphi_{m+i}(t_{\beta_j})_{i,j=1})^{n-m},$$

where the indices $\beta_1,\ldots,\beta_{n-m} \in \{1,\ldots,n\}\setminus\{\alpha_1,\ldots,\alpha_m\}$ are arranged in natural order. On the other hand due to the well known formula for minors of the inverse matrix (see e.g. Gantmacher (1998), Section 1.3) we obtain

$$\det(t_{\alpha_i}^{j-1})_{i,j=1}^m = (-1)^{\sum_{i=1}^m \alpha_i + m(m+1)/2} \det(d_{i\beta_j})_{i,j=1}^{n-m}(\det F)^{n-m-1}$$

and we have

$$\begin{aligned} \det X &= (\det F)^{n-m-1} \sum_{1\le\beta_1<\ldots<\beta_{n-m}\le n} \det(d_{i\beta_j})_{i,j=1}^{n-m} \det(\varphi_{m+i}(t_{\beta_j})_{i,j=1}^{n-m} \\ &= (\det F)^{n-m-1}\det(DRD^T) = \kappa^{n-m}(\det V)(\det F)^{n-m-1} \ne 0, \end{aligned} \tag{3.8}$$

which completes the proof of Theorem 3.1. □

# 4 $D$- and $G$-optimal designs for smoothing splines

In this section we present several properties of the $D$- and $G$-optimal designs for estimation with smoothing splines (see the definitions at the end of Section 2). We begin with an invariance property of the optimal designs, which reduces the design problem to the interval $[-1, 1]$.

**Theorem 4.1.** *Let $T = \{t_1, \ldots t_n\}$ denote the D- (G-) optimal design for estimation with the smoothing spline on the interval $[-1, 1]$ with penalty $\rho^2$, then the design $\bar{T} = \{\bar{t}_1, \ldots \bar{t}_n\}$ with,*

$$\bar{t}_i = \frac{b-a}{2} t_i + \frac{b+a}{2}\ , \quad i = 1, \ldots, n$$

*is D- (G-) optimal for estimation with the smoothing spline on the interval $[a, b]$ with penalty $(b-a)\rho^2/2$.*

**Proof.** Let $T = \{t_1, \ldots, t_n\}$ denote an arbitrary design on the interval $[-1, 1]$, define $c = \frac{b-a}{2}$ and $d = \frac{b+a}{2}$. We introduce the notation $\varphi_i(t, T)$ for the basis functions $\varphi_i$ reflecting the dependence on the design $T = \{t_1, \ldots, t_n\}$ and we will show at the end of this proof that the vector $\varphi(t, T) = (\varphi_1(t, T), \ldots, \varphi_n(t, T))^T$ defined by (3.3) and (3.4) satisfies

$$\varphi(ct + d, \bar{T}) = S\varphi(t, T), \tag{4.1}$$

where $\bar{T} = \{ct_1 + d, \ldots, ct_n + d\}$ denotes the transformed design, the $n \times n$ matrix $S$ is given by

$$S = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}, \tag{4.2}$$

$J_1$ is a lower $m \times m$ triangular matrix with diagonal elements $1, c \ldots, c^{m-1}$,

$$J_2 \;=\; c^s \mathrm{diag}\left(c^{m-2}, c^{m-3}, \ldots, c^{2m-n-1}\right) \in \mathbb{R}^{n-m \times n-m} \tag{4.3}$$

and $s = n(n-1)/2$ . With this representation we obtain ($\bar{t}_i = ct_i + d$)

$$\begin{aligned} M(\bar{T}) &= X^T(\bar{T})X(\bar{T}) + \lambda B \\ &= \sum_{i=1}^n \varphi(\bar{t}_i, \bar{T})\varphi^T(\bar{t}_i, \bar{T}) + \lambda \int_a^b \varphi^{(m)}(t, \bar{T})(\varphi^{(m)}(t, \bar{T}))^T dt \\ &= \sum_{i=1}^n S\varphi(t_i, T)\varphi^T(t_i, T)S^T + \frac{\lambda(b-a)}{2} \int_{-1}^1 S\varphi^{(m)}(t, T)(\varphi^{(m)}(t, T))^T S^T dt \\ &= S\underline{M}(T)S, \end{aligned}$$

where

$$\underline{M}(T) = \sum_{i=1}^n \varphi(t_i, T)\varphi^T(t_i, T) + \frac{\lambda(b-a)}{2} \int_{-1}^1 \varphi^{(m)}(t, T)(\varphi^{(m)}(t, T))^T dt.$$

Because the matrix $S$ does not depend on the design $T$, the $D$-optimal design on the interval $[a,b]$ for a given $\lambda$ (maximizing $\det M(\bar{T})$ with respect to $\bar{T}$) can be obtained from the $D$-optimal design on the interval $[-1,1]$ for $\lambda(b-a)/2$ (maximizing $\det M(T)$ with respect to $T$) by the linear transformation $\bar{t} = ct + d$. The assertion for $G$-optimality follows by the same argument observing the identity

$$\max_{t\in[-1,1]} \varphi^T(t,T)\underline{M}^{-1}(T)\varphi(t,T) = \max_{t\in[a,b]} \varphi^T(t,\bar{T})M^{-1}(\bar{T})\varphi(t,\bar{T}),$$

and it remains to prove the identity (4.1). For this purpose recall the definition of the matrix $F(T)$ and the vector $d_i(T)$ in (3.1) and (3.2), then it is easy to see that

$$\begin{aligned} \det F(\bar{T}) &= c^s \det F(T), \\ d_i(\bar{T}) &= c^{s-(m+i)} d_i(T)\ ,\ i = 1, \dots n-m,\ s = n(n-1)/2. \end{aligned}$$

A simple calculation shows that

$$(ct + d - \bar{t}_j)_+^{2m-1} = (c(t-t_j))_+^{2m-1} = c^{2m-1}(t-t_j)_+^{2m-1}$$

(note that $c > 0$), and a multiplication of this equation by the $jth$ component $d_{ij}(\bar{T})$ of the vector $d_i(\bar{T})$ yields

$$\varphi_{m+i}(ct+d,\bar{T}) = \sum_{j=1}^n d_{ij}(\bar{T})c^{2m-1}(t-t_j)_+^{2m-1} = c^{s+m-1-i}\varphi_{m+i}(t,T).$$

This proves the representation

$$(\varphi_{m+1}(ct+d,\bar{T}),\dots,\varphi_n(ct+d,\bar{T}))^T = J_2(\varphi_{m+1}(t,T),\dots,\varphi_n(t,T))^T,$$

where the matrix $J_2$ is defined in (4.3). Finally, the representation

$$(\varphi_1(ct+d,\bar{T}),\dots,\varphi_m(ct+d,\bar{T}))^T = J_1(\varphi_1(t,T),\dots,\varphi_m(t,T))^T$$

with a lower triangular matrix $J_1$ is obvious and the assertion of Theorem 4.1 follows. □

Theorem 4.1 shows that the determination of $D$- and $G$-optimal designs for the estimation with smoothing splines can be restricted to the design space $[-1,1]$. The next theorem gives some more information about the structure of the optimal designs.

**Theorem 4.2.** *Consider the estimation problem with smoothing splines on the interval* $[-1,1]$ *with* $n > m$ *and* $\rho \geq 0$.

*(i) Any $D$-optimal design for estimation with the smoothing spline contains the boundary points of the design space , i.e. $t_1 = -1$, $t_n = 1$.*

*(ii) There exists a G-optimal design for estimation with the smoothing spline which contains the boundary points* $t_1 = -1$, $t_n = 1$.

*(iii) If the D- (G-) optimal design for estimation with the smoothing spline is unique, then it is necessarily symmetric, i.e.* $t_{n-i+1} = -t_i$, $i = 1, \ldots, n$.

**Proof.**

(i) Assume that $T = \{t_1, \ldots, t_n\}$ denotes a $D$-optimal design for estimation with the smoothing spline on the interval $[-1, 1]$ with $t_1 \neq -1$ or $t_n \neq 1$ . We define $c = 2/(t_n - t_1) > 1$ and $d = -(t_n + t_1)/(t_n - t_1)$, then the transformation

$$\bar{t} = ct + d$$

maps the interval $[t_1, t_n]$ onto $[-1, 1]$. From the proof of Theorem 4.1 we have for the design $\bar{T} = \{\bar{t}_1, \ldots, \bar{t}_n\}$ on the interval $[-1, 1]$

$$M(\bar{T}) = S\underline{M}(T)S^T \geq SM(T)S^T,$$

where the matrix $\underline{M}(T)$ is defined in the proof of Theorem 4.1 and $(b-a)/2$ is replaced by $c > 1$. This implies

$$\det(M(\bar{T})) \geq (\det S)^2 \det(M(T))$$

where $(\det S)^2 > 1$, which contradicts to the $D$-optimality of the design $T$.

(ii) This follows similarly observing the identity

$$\varphi^T(t, T)\underline{M}^{-1}(T)\varphi(t, T) = \varphi^T(t, \bar{T})M^{-1}(\bar{T})\varphi(t, \bar{T}).$$

(iii) Assume that the $D$-optimal design for the smoothing spline on the interval $[-1, 1]$ is unique, consider the transformation $\bar{t} = -t$ and define the design $-T = \{-t_n, \ldots, -t_1\}$. A similar calculation as given in the proof of Theorem 4.1 shows

$$\det M(-T) = \det M(T),$$

and therefore we obtain from the uniqueness of the $D$-optimal design that $-T = T$, that is $t_{n-i+1} = -t_i$, $i = 1, \ldots, n$. The corresponding statement for $G$-optimality follows by similar arguments. □

**Lemma 4.3.** *If there exists a design* $\bar{T}$ *on the interval* $[-1, 1]$*, which satisfies*

$$\max_{t \in [-1,1]} \varphi^T(t, \bar{T})M^{-1}(\bar{T})\varphi(t, \bar{T}) = tr\left(X^T(\bar{T})X(\bar{T})M^{-1}(\bar{T})\right), \tag{4.4}$$

$$tr\left(X^T(\bar{T})X(\bar{T})M^{-1}(\bar{T})\right) = \min_T tr\left(X^T(T)X(T)M^{-1}(T)\right), \tag{4.5}$$

*then $\bar{T}$ is a G-optimal design for estimation with the smoothing spline on the interval* $[-1,1]$.

**Proof.** We have for any design $T=\{t_1,\dots,t_n\}$ on the interval $[-1,1]$

$$\begin{aligned}\psi(T) &= \max_{t\in[-1,1]} \varphi^T(t,T)M^{-1}(T)\varphi(t,T) \geq \sum_{i=1}^n \varphi^T(t_i,T)M^{-1}(T)\varphi(t_i,T) \\ &= \operatorname{tr}\left(X^T(T)X(T)M^{-1}(T)\right) \geq \min_{\tilde{T}} \operatorname{tr}\left(X^T(\tilde{T})X(\tilde{T})M^{-1}(\tilde{T})\right).\end{aligned}$$

Consequently we obtain from (4.4) and (4.5)

$$\psi(\bar{T}) = \min_T \max_{t\in[-1,1]} \varphi^T(t,T)M^{-1}(T)\varphi(t,T),$$

which establishes the $G$-optimality of the design $\bar{T}$. □

Note that in the case $\rho=\infty$ or equivalently $\lambda=0$ we have $\operatorname{tr}(X^T(\bar{T})X(\bar{T})M^{-1}(\bar{T}))=n$ and condition (4.5) of Lemma 4.3 is trivially satisfied. Note also that the case $\lambda=0$ corresponds to an interpolating spline and is for this reason not of practical interest for estimation. However, our numerical studies presented in the following section show that $D$-optimal designs for estimation with smoothing splines are rather robust with respect to the choice of the smoothing parameter. Therefore these optimal designs are also of some interest in the case $\lambda=0$, which will now be considered for the remaining part of this section. In particular we will show that in this case $D$- and $G$-optimal designs for the smoothing splines are identical, if $n=m+1$. In the case $n>m+1$ this statement is not true anymore. Note that the famous equivalence theorem of Kiefer and Wolfowitz (1960) cannot be directly applied in the present context, because we consider exact designs in this paper and the regression functions in the optimality criteria depend on the design points. In order to apply approximate design theory in the present context we consider a linear regression model with a basis of $N^{2m}(u_1,\dots,u_n)$ as regression functions, where the knots $U=\{u_1,\dots,u_n\}$ do not necessarily coincide with the points $T=\{t_1,\dots,t_n\}$, where observations are taken, i.e.

$$Y = X(T,U)\theta+\varepsilon \tag{4.6}$$

Here $Y=(Y_1,\dots,Y_n)^T$ denotes the vector of observations in the nonparametric regression model (1.1), $\theta=(\theta_1,\dots,\theta_n)^T$, $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)^T$ denotes a vector of centered and uncorrelated random variables with equal variances,

$$X(T,U) = (\varphi_i(t_j,U))_{i,j=1}^n \tag{4.7}$$

and $\varphi_1(t,U)),\dots,\varphi_n(t,U))$ are the basis functions of $N^{2m}(u_1,\dots,u_n)$ defined by (3.4) and (3.5), where the points $t_j$ in (3.3) have been replaced by $u_j$ $(j=1,\dots,n)$. In what follows define

$$\begin{aligned}
&(4.8) & R(T,U) &= ((t_i-u_j)_+^{2m-1})_{i,j=1}^n, \\
&(4.9) & F(T) &= (t_i^{j-1})_{i,j=1}^m\ , \quad F(U)=(u_i^{j-1})_{i,j=1}^m, \\
&(4.10) & d_i(T) &= (\det F(T))(F^T(T))^{-1}e_{m+i}\ , \quad d_i(U)=(\det F(U))(F^T(U))^{-1}e_{m+i}, \\
&(4.11) & D(T) &= (d_{ij}(T))_{i=1,\dots,n-m}^{j=1,\dots,n}\ , \quad D(U)=(d_{ij}(U))_{i=1,\dots,n-m}^{j=1,\dots,n},
\end{aligned}$$

[here $d_{ij}(T)$ denotes the $j$th component of the vector $d_i(T)$], then the following result holds.

**Lemma 4.4.** *We have*

$$[\det(D(T)R(T,U)D^T(U))]^2 \leq \det(D(T)R(T,T)D^T(T))\det(D(U)R(U,U)D^T(U))$$

*with equality if and only if $T = U$.*

**Proof.** With the same arguments as presented in the proof of Theorem 3.1 it follows

$$\kappa^{n-m}\det(D(T)R(T,U)D^T(U)) = \det V(T,U), \tag{4.12}$$
$$\kappa^{m-n}\det(D(T)R(T,T)D^T(T)) = \det V(T,T), \tag{4.13}$$

where the $(n-m)\times(n-m)$ matrix $V(T,U)$ is defined by

$$V(T,U) = \Big(\int_a^b \varphi_{m+i}^{(m)}(t,T)\varphi_{m+j}^{(m)}(t,U)dt\Big)_{i,i=1,\dots,n-m}$$

On the other hand we have from the Cauchy Binet formula [see e.g. Karlin and Studden (1966), p. 14] and the Cauchy Schwarz inequality

$$\begin{aligned}
\det V(T,U) &= \det\Big(\int_a^b \varphi_{m+i}^{(m)}(t,T)\varphi_{m+j}^{(m)}(t,U)dt\Big)_{i,i=1,\dots,n-m} \\
&= \int_a^b \cdots \int_a^b \det\Big(\varphi_{m+i}^{(m)}(z_j,T)\Big)_{i,j=1}^{n-m} \det\Big(\varphi_{m+j}^{(m)}(z_i,U)\Big)_{i,j=1}^{n-m} dz_1,\dots dz_{n-m} \\
&\leq \Big\{\int_a^b \cdots \int_a^b \Big[\det\Big(\varphi_{m+i}^{(m)}(z_j,T)\Big)_{i,j=1}^{n-m}\Big]^2 dz_1\dots dz_{n-m} \\
&\qquad \times \int_a^b \cdots \int_a^b \Big[\det\Big(\varphi_{m+i}^{(m)}(z_i,U)\Big)_{i,j=1}^{n-m}\Big]^2 dz_1\dots dz_{n-m}\Big\}^{1/2} \\
&= \{\det V(T,T)\det V(U,U)\}^{1/2}.
\end{aligned}$$

Note that the equality takes place if and only if for some constant $c \in \mathbb{R}$

$$\det\Big(\varphi_{m+i}^{(m)}(z_j,T)\Big)_{i,j=1}^{n-m} = c\det\Big(\varphi_{m+j}^{(m)}(z_i,U)\Big)_{i,j=1}^{n-m}$$

almost everywhere with respect to measure $dz_1\dots dz_{n-m}$ on $[a,b]^{n-m}$. But since both sides are piecewise polynomials, this takes place if and only if $U = T$ and the assertion of Lemma 4.4 follows. For example, consider the case $n-m = 1$. In this case the condition is of the form

$$\sum_{j=1}^{m+1} d_{1j}(T)(z-t_j)_+^{m-1} - c\sum_{j=1}^{m+1} d_{1j}(U)(z-u_j)_+^{m-1} = 0$$

almost everywhere on the interval 8 $[a,b]$. This equality is impossible if $U \neq T$ since the left side will not be $m-1$ times continuously differentiable in point $z = t_j$ for some $j \in \{1,\ldots,n\}$. The general case $n > m+1$ can be considered in a similar (but slightly more complicated) way. $\square$

For the formulation of our next result we consider an approximate design $\xi$ with weights $\xi_1,\ldots,\xi_N$ at the points $t_1,\ldots,t_N$ [see e.g. Kiefer (1974)] and its information matrix

$$\tilde{M}(\xi,U) = \sum_{i=1}^{N} \varphi(t_i,U)\varphi^T(t_i,U)\xi_i \,. \tag{4.14}$$

**Proposition 4.5.** *Consider the smoothing spline defined by (1.4) with* $n = m+1$ *and* $\rho = \infty$*. Let* $T^* = \{t_1^*,\ldots,t_n^*\}$ *denote a* $D$*-optimal design for estimation with the smoothing spline. If there exists an approximate* $D$*-optimal design with* $N = n$ *support points maximizing* $\det \tilde{M}(\xi,T^*)$ *in the class of all approximate design, then the design* $T^* = \{t_1^*,\ldots,t_n^*\}$ *is* $G$*-optimal for estimation with the smoothing spline.*

**Proof.** Let $T^* = \{t_1^*,\ldots,t_n^*\}$ denote a $D$-optimal design for estimation with the smoothing spline and $\tilde{\xi}$ be a $D$-optimal design maximizing $\det \tilde{M}(\xi,T^*)$ with $N = n$ support points, say $\tilde{T} = \{\tilde{t}_1,\ldots,\tilde{t}_n\}$. A standard result from approximate design theory shows that $\tilde{\xi}_i = \frac{1}{n}$ $i = 1,\ldots,n$. Obviously we have

$$\det \tilde{M}(\tilde{\xi},T^*) = (\frac{1}{n})^n \det M(\tilde{T},T^*) = (\frac{1}{n})^n \max_T \det M(T,T^*)$$

(otherwise the design $\tilde{\xi}$ would not be the approximate $D$-optimal design). We obtain from Lemma 4.4 and formula (3.8)

$$(\det M(\tilde{T},T^*))^2 = (\det(D(\tilde{T})R(\tilde{T},T^*)D^T(T^*)^2 \leq \det M(\tilde{T},\tilde{T}) \cdot \det M(T^*,T^*)$$

with equality if and only if $T^* = \tilde{T}$. Consequently it follows that $\tilde{T} = T^*$. If $\xi^* = \tilde{\xi}$ is the approximate $D$-optimal design with equal masses at the points $t_1^*,\ldots,t_n^*$, it follows by the equivalence theorem of Kiefer and Wolfowitz (1960) that $\xi^*$ is also an approximate $G$-optimal design, which means that it minimizes

$$\Phi(\xi) = \max_{t\epsilon[a,b]} \varphi^T(t,T^*)\tilde{M}^{-1}(\xi,T^*)\varphi(t,T^*)$$

among all approximate designs. Moreover, $\Phi(\xi^*) = n$, which implies (observing the identity $n\tilde{M}(\xi^*,T^*) = M(T^*,T^*)$)

$$\max_{t\epsilon[a,b]} \varphi^T(t,T^*)M^{-1}(T^*,T^*)\varphi(t,T^*) = 1 \,.$$

On the other hand we have for any design $T = \{t_1, \dots, t_n\}$ for estimation with the smoothing spline

$$\begin{aligned}\max_{t\epsilon[a,b]} \varphi^T(t,T)M^{-1}(T,T)\varphi(t,T) &\geq \frac{1}{n}\sum_{i=1}^{n} \varphi^T(t_i,T)M^{-1}(T,T)\varphi(t_i,T) \\ &= \frac{1}{n} tr(M^{-1}(T,T)M(T,T)) = 1\,,\end{aligned}$$

which proves that the design $T^*$ is also $G$-optimal. $\Box$.

We will conclude this section with a more detailed discussion of the case $m = 1$ for which the resulting optimization problem for the determination of the $D$-optimal design can be formulated more explicitly.

**Theorem 4.6.** *Consider the quadratic smoothing spline ($m = 1$) with $\rho = \infty$, then the $D$-optimal design $T^* = \{t_1^*, \dots, t_n^*\}$ on the interval $[-1, 1]$ is unique, $t_1^* = -1$, $t_n^* = 1$ and $t_2^*, \dots, t_{n-1}^*$ maximize the function*

$$\det X(T) = \prod_{i=2}^{n}(t_i - t_{i-1})\Big(\prod_{1\leq i<j\leq n}(t_j - t_i)\Big)^{n-2}$$

*Moreover, the $D$-optimal designs $T^*$, is symmetric, i.e. $t_{n-i+1}^* = -t_i^*$, $i = 1, \dots, n$.*

**Proof.** Note that we have by assumption $m = 1$, which implies for any design $T = \{t_1, \dots, t_n\}$

$$X(T) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & & Q & \\ 1 & & & \end{pmatrix} \in \mathbb{R}^{n\times n},$$

where the $(n-1)\times(n-1)$ matrix $Q = Q_1 Q_2$ is defined by

$$\begin{aligned} Q_1 &= \begin{pmatrix} t_2 - t_1 & 0 & 0 & \dots & 0 \\ t_3 - t_1 & t_3 - t_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_n - t_1 & t_n - t_2 & t_n - t_3 & \dots & t_n - t_{n-1} \end{pmatrix} \in \mathbb{R}^{n-1\times n-1}, \\ Q_2 &= \begin{pmatrix} d_{1,1} & \dots & d_{1,n-1} \\ \vdots & \vdots & \vdots \\ d_{n-1,1} & \dots & d_{n-1,n-1} \end{pmatrix} \in \mathbb{R}^{n-1\times n-1}. \end{aligned}$$

Note that the matrix $Q_2$ is a minor of the matrix $(F^T)^{-1}$ and we obtain using the known formula for minors of the inverse matrix [see Gantmacher (1998), Section 1.3] that

$$\det Q_2 = (\det F)^{n-2},$$

which yields

$$\det X(T) = \det Q_1 \det Q_2 = \prod_{i=2}^{n}(t_i - t_{i-1})\Big(\prod_{1\le i<j\le n}(t_j - t_i)\Big)^{n-2}. \tag{4.15}$$

It was shown in Theorem 4.2 that the the $D$-optimal design satisfies $t_1^* = -1$ and $t_n^* = 1$. Moreover, for $t_1 = -1$ and $t_n = 1$ the function on the right hand side of (4.15) is strictly concave as a function of $(t_2, \ldots, t_{n-1})$, which implies the uniqueness of the $D$-optimal design. The assertion regarding the symmetry follows form part (iii) of Theorem 4.3. $\Box$.

# 5 Some examples

In this section we present several numerical results illustrating the theory. We begin with a discussion of the case $\rho = \infty$, which allows an explicit calculation of the optimal designs for small sample sizes. The second example refers to some numerical calculation and comparison of $D$- and $G$-optimal designs for estimation with linear and quadratic smoothing splines.

**Example 5.1.** Consider the case $m = 1$ and $\rho = \infty$. For a small sample sizes the optimization problem in Theorem 4.6 can be solved explicitly and the $D$-optimal designs can be found explicitly. In the case $n = 2$ and $n = 3$ it follows from Theorem 4.6 that the $D$-optimal design is given by $\{-1, 1\}$ and $\{-1, 0, 1\}$, respectively. In the case $n = 4$ the symmetry implies that the $D$-optimal design is of the form $\{-1, -x, x, 1\}$ where $x \in [0, 1]$ maximizes the function $x^3(x-1)^6(x+1)^4$, that is $x = (-1 + 2\sqrt{10})/13$. Similarly, in the case $n = 5$ the $D$-optimal design is given by $\{-1, -x, 0, x, 1\}$, where $x = (-1 + 2\sqrt{69})/25$. Note that by Proposition 4.5 the $D$-optimal design for $n = 2$ is also $G$-optimal.

**Example 5.2.** Note that in all cases of practical interest the optimal designs for estimation with the smoothing spline have to be calculated numerically. We have performed such calculations for the case $m = 1$ considered in Theorem 4.6 and the smoothing cubic spline (i.e. $m = 2$).

In Table 1 we present the $D$- and $G$-optimal designs for the estimation with the quadratic ($m = 1$) and cubic ($m = 2$) spline in the case of $n = 6$ and $n = 8$ knots, while Table 2 contains the corresponding results if $n = 10$ knots are used. It is interesting to note that the $D$-optimal designs for estimation with the quadratic and cubic spline do not change substantially for different values of the parameter $\lambda$. In the case of $G$-optimal designs the situation is completely different. Here a larger value of the smoothing parameter $\lambda$ yields to a $G$-optimal design which is more concentrated at the boundary. This corresponds to intuition, because a larger value of $\lambda$ yields to a more smooth function (in the extreme case a line), for which it is better to take observations at the boundary of the design space. A comparison of $D$- and $G$-optimal designs for fixed $n$ shows that in the case of no smoothing ($\lambda = 0$) the $G$-optimal designs are slightly more concentrated at the interior of the design space. On the other hand, if $\lambda > 0$, the $D$-optimal designs do not change, but the knots corresponding to the $G$-optimal designs are more concentrated at the boundary.

A comparison of the designs for the quadratic and cubic spline model shows nearly no differences for the $D$-optimality criterion. Similarly, in the case $\lambda = 0$ the $G$-optimal designs show the same pattern. On the other hand if $\lambda > 0$ the differences between the $G$-optimal designs for the quadratic and cubic spline model are clearly visible. In particular in the case $m = 1$ the $G$-optimal designs are more concentrated at the boundary of the design space.

It might also be of interest to investigate the function

$$d(t,T) = \varphi^T(t)(X^TX + \lambda B)^{-1}\varphi(t) \tag{5.1}$$

for the different designs. In the upper panel of Figure 1 we present this function for the $G$-optimal design and the uniform design. The results show that for $\lambda > 0$ the $G$-optimal design has a better performance at the boundary of the design interval, while in the interior of the interval $[-1, 1]$ the uniform design has advantages. For $\lambda = 0$ both designs yield similar curves $t \to d(t,T)$. Note that a $G$-optimal design minimizes the worst case, which appears at the boundary of the design interval. Thus the price which has to be paid for this better performance at the boundary is a worse behaviour in the interior of the interval $[-1, 1]$. If a minimax approach might not be appropriate for the construction of optimal designs one could alternatively determine optimal designs which minimize the integrated "variance"

$$\int_{-1}^{1} \varphi^T(t)(X^TX + \lambda B)^{-1}\varphi(t)dt$$

and are called $I$-optimal designs for estimation with smoothing splines. Some $I$-optimal designs are shown for the cases $m = 1$ and $m = 2$ in Table 3 and 4. It is interesting to note that there appear not too substantial differences between the $I$-optimal designs for estimation with the quadratic and cubic splines. A comparison of $G$- and $I$-optimal designs shows that the $I$-optimal designs are more concentrated in the interior of the design interval $[-1, 1]$. As a consequence we observe in the lower panel of Figure 1 that in the case $\lambda > 0$ the uniform design yields smaller values for the function $d(t,T)$ defined in (5.1) and is smaller at the boundary of the interval $[-1, 1]$. The opposite behaviour is observed in the interior of the interval $[-1, 1]$. Finally, if $\lambda = 0$, there are no substantial differences between $I$-, $G$-optimal designs and the uniform design.

**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 04-01-00519). The work of H. Dette was supported in part by a NIH grant award IR01GM072876:01A1.

# References


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

H. Dette, V.B. Melas and A.N. Pepelyshev (2007).Optimal designs for free knot least squares splines.Preprint Ruhr-Universität Bochum. http://www.ruhr-uni- bochum.de/mathematik3/preprint.htm Statistica Sinica, to appear.

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

S. Efromovich (1999). Nonparametric Curve Estimation : Methods, Theory and Applications. Springer, New York.

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

J. Fan, I. Gijbels (1996). Local polynomial modelling and its applications. Chapman and Hall, London.

P. Green and B. Silverman (1994). Nonparametric Regression and Generalized Linear Models, Chapman and Hall.

Gantmacher, F.R. (1998). *The Theory of Matrices*. Chelsea, Providence. RI.

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.

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 and W.J. Studden (1966). Tchebycheff systems: with applications in anlysis and statistics. Wiley, Interscience.

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.

J. Kuks W. Olman (1971). Minimaxnaja linejnaja ozenka coefficientov regressii II,Izv. Akad. Nauk. Eston. SSR 21, 6672.

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

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.

B. W. Silverman (1985). Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting. Journal of the Royal Statistical Society. Series B, 47, 1-52.

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

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

H. Toutenburg (1982). Prior Information in Linear Models, Wiley, New York.

G. Wahba (1990). Spline Models for Observational Data, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59, SIAM, Philadelphia.

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

D.C. Woods (2005). Designing experiments under random contamination with application to polynomial spline regression. Statistica Sinica, 15, 619-635.

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

**Table 1:** *$D$- and $G$-optimal designs for estimation with the quadratic and cubic smoothing spline*

| | | quadratic spline ($m=1$) | | | | | | cubic spline ($m=2$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $D$-optimal | | | $G$-optimal | | | $D$-optimal | | | $G$-optimal | | |
| $n$ | $\lambda$ | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 |
| | $t_1$ | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| | $t_2$ | -0.74 | -0.75 | -0.75 | -0.68 | -1 | -1 | -0.73 | -0.73 | -0.74 | -0.65 | -0.78 | -0.85 |
| 6 | $t_3$ | -0.27 | -0.28 | -0.28 | -0.33 | -1 | -1 | -0.26 | -0.26 | -0.27 | -0.21 | -0.30 | -0.24 |
| | $t_4$ | 0.27 | 0.28 | 0.28 | 0.33 | 1 | 1 | 0.26 | 0.26 | 0.27 | 0.21 | 0.30 | 0.24 |
| | $t_5$ | 0.74 | 0.75 | 0.75 | 0.68 | 1 | 1 | 0.73 | 0.73 | 0.74 | 0.65 | 0.78 | 0.85 |
| | $t_6$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | $\lambda$ | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 |
| | $t_1$ | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| | $t_2$ | -0.86 | -0.86 | -0.87 | -0.79 | -1 | -0.95 | -0.85 | -0.86 | -0.86 | -0.77 | -0.89 | -1 |
| | $t_3$ | -0.58 | -0.58 | -0.59 | -0.51 | -0.32 | -0.95 | -0.57 | -0.57 | -0.58 | -0.47 | -0.54 | -0.60 |
| 8 | $t_4$ | -0.21 | -0.21 | -0.21 | -0.20 | -0.30 | -0.87 | -0.20 | -0.20 | -0.20 | -0.15 | -0.20 | -0.17 |
| | $t_5$ | 0.21 | 0.21 | 0.21 | 0.20 | 0.30 | 0.87 | 0.20 | 0.20 | 0.20 | 0.15 | 0.20 | 0.17 |
| | $t_6$ | 0.58 | 0.58 | 0.59 | 0.51 | 0.32 | 0.95 | 0.57 | 0.57 | 0.58 | 0.47 | 0.54 | 0.6 |
| | $t_7$ | 0.86 | 0.87 | 0.87 | 0.78 | 1 | 0.95 | 0.85 | 0.86 | 0.86 | 0.77 | 0.89 | 1 |
| | $t_8$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table 2:** *$D$- and $G$-optimal designs for estimation with the quadratic and cubic smoothing spline*

| | | quadratic spline ($m=1$) | | | | | | cubic spline ($m=2$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $D$-optimal | | | $G$-optimal | | | $D$-optimal | | | $G$-optimal | | |
| $n$ | $\lambda$ | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 | 0 | 0.01 | 0.1 |
| | $t_1$ | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| | $t_2$ | -0.91 | -0.92 | -0.92 | -0.82 | -1 | -1 | -0.91 | -0.91 | -0.91 | -0.83 | -0.95 | -1 |
| | $t_3$ | -0.73 | -0.73 | -0.74 | -0.58 | -0.32 | -1 | -0.72 | -0.73 | -0.73 | -0.60 | -0.66 | -0.89 |
| | $t_4$ | -0.47 | -0.47 | -0.48 | -0.41 | -0.31 | -0.23 | -0.46 | -0.47 | -0.47 | -0.35 | -0.43 | -0.33 |
| 10 | $t_5$ | -0.16 | -0.16 | -0.16 | -0.20 | 0 | 0 | -0.16 | -0.16 | -0.16 | -0.12 | -0.12 | -0.31 |
| | $t_6$ | 0.16 | 0.16 | 0.16 | 0.20 | 0 | 0 | 0.16 | 0.16 | 0.16 | 0.12 | 0.12 | 0.31 |
| | $t_7$ | 0.47 | 0.47 | 0.48 | 0.41 | 0.32 | 0.23 | 0.46 | 0.47 | 0.47 | 0.35 | 0.43 | 0.33 |
| | $t_8$ | 0.73 | 0.73 | 0.74 | 0.58 | 0.32 | 1 | 0.72 | 0.73 | 0.73 | 0.60 | 0.66 | 0.89 |
| | $t_9$ | 0.91 | 0.92 | 0.92 | 0.82 | 1 | 1 | 0.92 | 0.91 | 0.91 | 0.83 | 0.95 | 1 |
| | $t_{10}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table 3:** *$I$- optimal designs for estimation with the quadratic and cubic smoothing spline*

| | | quadratic spline ($m = 1$) | | | cubic spline ($m = 2$) | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\lambda$ | 0 | 0.001 | 0.01 | 0 | 0.01 | 0.1 |
| 6 | $t_1$ | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 |
| | $t_2$ | -0.60 | -0.60 | -0.63 | -0.63 | -0.58 | -0.49 |
| | $t_3$ | -0.22 | -0.20 | -0.21 | -0.21 | -0.18 | -0.22 |
| | $t_4$ | 0.22 | 0.20 | 0.21 | 0.21 | 0.18 | 0.22 |
| | $t_5$ | 0.60 | 0.60 | 0.63 | 0.63 | 0.58 | 0.49 |
| | $t_6$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 8 | $t_1$ | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 |
| | $t_2$ | -0.75 | -0.72 | -0.75 | -0.73 | -0.66 | -0.75 |
| | $t_3$ | -0.48 | -0.44 | -0.43 | -0.43 | -0.40 | -0.30 |
| | $t_4$ | -0.18 | -0.15 | -0.14 | -0.14 | -0.13 | -0.30 |
| | $t_5$ | 0.18 | 0.15 | 0.14 | 0.14 | 0.13 | 0.30 |
| | $t_6$ | 0.48 | 0.44 | 0.43 | 0.43 | 0.40 | 0.30 |
| | $t_7$ | 0.75 | 0.72 | 0.75 | 0.73 | 0.66 | 0.75 |
| | $t_8$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

**Table 4:** *$I$- optimal designs for estimation with the quadratic and cubic smoothing spline*

| | | quadratic spline ($m = 1$) | | | cubic spline ($m = 2$) | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\lambda$ | 0 | 0.001 | 0.01 | 0 | 0.01 | 0.1 |
| 10 | $t_1$ | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 | -1.00 |
| | $t_2$ | -0.80 | -0.79 | -0.83 | -0.79 | -0.72 | -0.94 |
| | $t_3$ | -0.59 | -0.57 | -0.69 | -0.56 | -0.55 | -0.46 |
| | $t_4$ | -0.41 | -0.35 | -0.29 | -0.33 | -0.32 | -0.46 |
| | $t_5$ | -0.19 | -0.12 | -0.13 | -0.11 | -0.11 | 0 |
| | $t_6$ | 0.19 | 0.12 | 0.13 | 0.11 | 0.11 | 0 |
| | $t_7$ | 0.41 | 0.35 | 0.29 | 0.33 | 0.32 | 0.46 |
| | $t_8$ | 0.59 | 0.57 | 0.69 | 0.56 | 0.55 | 0.46 |
| | $t_9$ | 0.80 | 0.79 | 0.83 | 0.79 | 0.72 | 0.94 |
| | $t_{10}$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

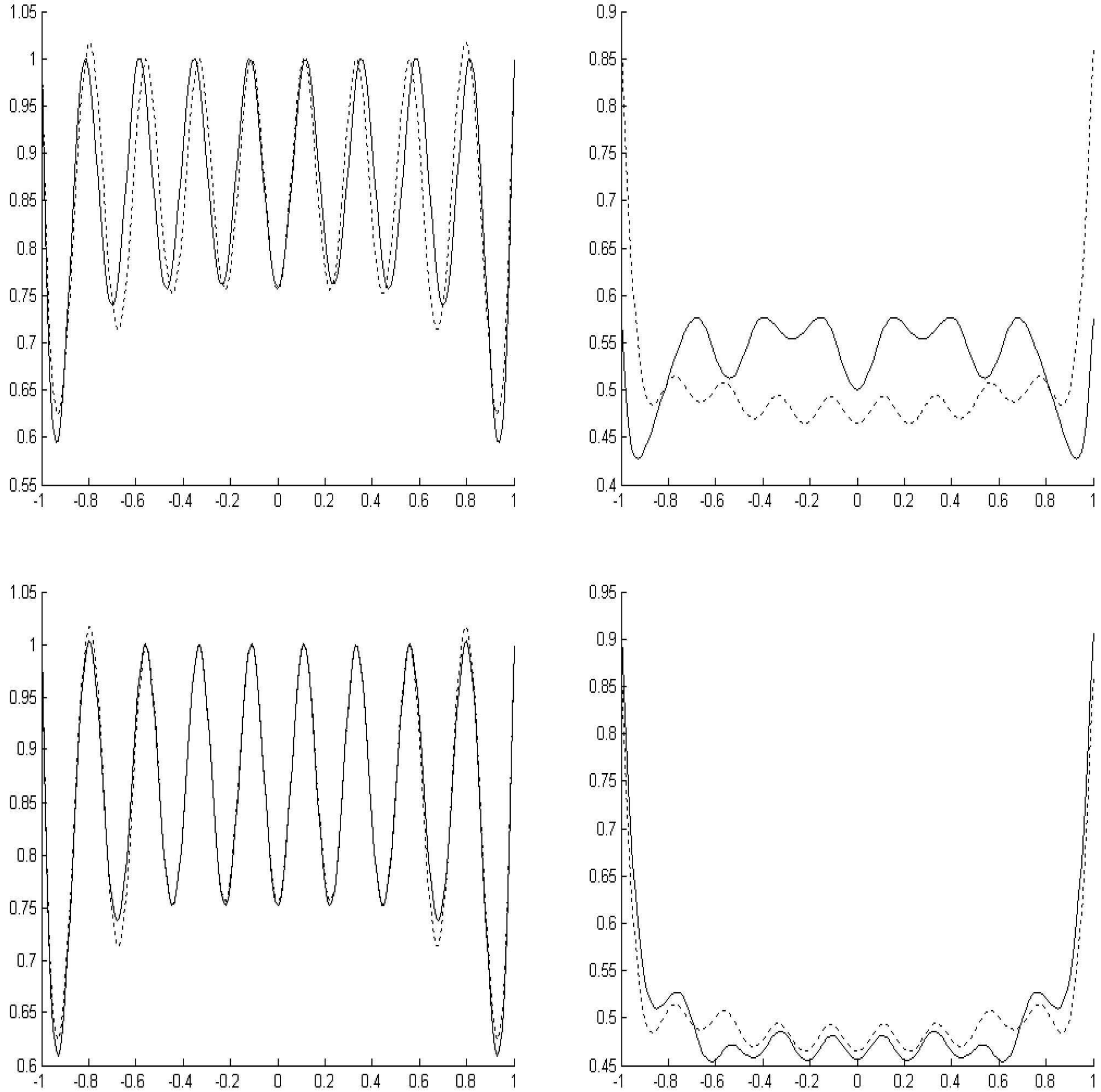


**Figure 1** *The function $d(t, \xi)$ defined in (5.1) for various optimal designs and different values of the parameter $\lambda$ (left part: $\lambda = 0$, right part: $\lambda = 0.01$. Upper panel: comparison of $G$-optimal (solid line) and uniform design (dashed line). Lower panel: comparison of $I$-optimal (solid line) and uniform design (dashed line). The number of observations is $n = 10$.*