Ann Inst Stat Math (2011) 63:981–1003
DOI 10.1007/s10463-009-0265-x

Optimal design for smoothing splines

Holger Dette · Viatcheslav B. Melas · Andrey Pepelyshev

Received: 27 August 2008 / Revised: 13 July 2009 / Published online: 17 November 2009
© The Institute of Statistical Mathematics, Tokyo 2009

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 special basis for the space of natural splines is introduced 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-optimality 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.

Keywords Smoothing spline · D- and G-optimal designs · Saturated designs

1 Introduction

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

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

where $a \le t_1 < \cdots < t_n \le b$ are the design points, the errors are independent identically distributed with mean 0 and variance $\sigma^2 > 0$. There are several procedures to

H. Dette (✉)
Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
e-mail: holger.dette@rub.de

V. B. Melas · A. Pepelyshev
Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia

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, Dierckx 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

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

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 $\int_a^b |g^{(m)}(t)|^2 \,\mathrm{d}t < \infty$, then the sum of squares in (2) is minimized in the class

$$\mathcal{F}_\rho = \left\{ g \in W^{(m)}([a, b]) \;\middle|\; \int_a^b |g^{(m)}(t)|^2 \,\mathrm{d}t \le \rho^2 \right\}, \tag{3}$$

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

$$\min_{g \in \mathcal{F}_\rho} \sum_{i=1}^{n}(Y_i - g(t_i))^2. \tag{4}$$

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

$$\sum_{i=1}^{n}(Y_i - g(t_i))^2 + \lambda \int_a^b |g^{(m)}(t)|^2 \,\mathrm{d}t \tag{5}$$

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; 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 nonparametric estimation with smoothing splines has—to the knowledge of the authors—not been investigated so far. The only paper which discussed design aspects for smoothing spline models is Butler (2001), where properties of G-optimal designs for polynomial regression are studied if these designs are used for smoothing spline models. In the same paper, sufficient conditions are given under which optimal designs for polynomials are also optimal for smoothing splines.

On the other hand, 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 VanArman (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; Mao and Zhao 2003). The construction of D-optimal designs for splines with estimated knots was considered recently by Dette et al. (2008). 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).

The fact that no optimal designs are available for estimation with the smoothing spline 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 (4). The goal of the present paper is to partially fill this gap and present some useful tools for constructing optimal designs for the estimation of the regression with smoothing splines. In Sect. 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 Sect. 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 Sect. 4. Two optimality criteria for the determination of optimal designs are introduced corresponding to 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 (in general) not equivalent. Some numerical results and a comparison of the optimal design with the commonly used uniform design in this context are presented in Sect. 5. The numerical study also includes optimal designs with respect to an I-optimality criterion which is in some cases more appropriate for mean squared error considerations. Finally, most technical details and proofs are deferred to an Appendix.

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 (4) or (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 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, ∞).

The dimension of the vector space $N^{2m}(t_1, \dots, t_n)$ is precisely n (see Eubank 1999), and if $\varphi_1(t), \dots, \varphi_n(t)$ 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

$$X = X(T) = (\varphi_i(t_j))_{i,j=1,\dots,n} \tag{6}$$

(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 emphasize this dependence) and define the vector $\varphi(t) = (\varphi_1(t), \dots, \varphi_n(t))^T$. If

$$g(t) = \theta^T \varphi(t) = \sum_{j=1}^{n} \theta_j \varphi_j(t) \tag{7}$$

denotes a natural spline, then it is easy to see that the defining constraint (3) for the set $\mathcal{F}_\rho$ is $\theta^T B\theta \le \rho^2$, where the matrix B is given by

$$B = \left(\int_a^b \varphi_i^{(m)}(t) \varphi_j^{(m)}(t)\, \mathrm{d}t \right)_{i,j=1,\dots,n}. \tag{8}$$

In what follows we consider the set

$$\Omega = \{\theta \in \mathbb{R}^n \mid \theta^T B\theta \le \rho^2\} \tag{9}$$

of all vectors θ 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:

$$\inf_{a \in \mathbb{R}^n} \sup_{g \in \mathcal{F}_\rho} E[(g(t_k) - a^T Y)^2], \tag{10}$$

where $Y = (Y_1, \dots, Y_n)^T$ denotes the vector of all observations. The following Lemma can be found in Eubank (1999). In the Appendix we will present a proof, because we need the optimal value of the minimax problem (10) for the definition of the optimality criteria, and could not find this in the literature.

Lemma 1 *Assume that* $n > m$ *and* $k \in \{1, \dots, n\}$. *The solution of the minimax problem* (10) *is given by*

$$a^* = X(X^T X + \lambda B)^{-1} \varphi(t_k), \tag{11}$$

and the minimum value is given by

$$\inf_{a \in \mathbb{R}^n} \sup_{g \in \mathcal{F}_\rho} E[(g(t_k) - a^T Y)^2] = \sigma^2 \varphi^T(t_k)(X^T X + \lambda B)^{-1} \varphi(t_k), \tag{12}$$

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

The value of the minimax criterion (12) will be the basic criterion for constructing optimal designs for estimation with smoothing splines. More precisely, note that the matrix

$$\sigma^2(X^T X + \lambda B)^{-1} \tag{13}$$

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

$$M(T) = X^T X + \lambda B \tag{14}$$

(see Silvey 1980; Pukelsheim 1993). However, in contrast to classical design theory for regression, the functions φ_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 (13) will yield a "small" value for the optimum in (12) 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 (14) and one criterion, which takes also into account that the functions φ_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

$$\max_{t \in [a,b]} \varphi^T(t, T) M^{-1}(T) \varphi(t, T), \tag{15}$$

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 φ_i in (15) also depend on the design points. Note also that here the design problem is defined on a noncompact set. While this fact does not cause any difficulties for the D-criterion, the existence of G-optimal designs is not guaranteed. To overcome this difficulty we introduce in the next section a modified definition of G-optimality. Moreover, the D-optimal design may depend on the choice of the basis and 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 Optimal designs and a special basis for natural splines

In the following any ordered sequence of points $T = (t_1, \dots, t_n)$ such that

$$a \le t_1 \le t_2 \le \dots \le t_n \le b$$

will be called experimental design. If $t_1 < t_2 < \cdots < t_n$ the design will be called nondegenerated; otherwise, if $t_i = t_{i+1}$ for at least some i, $1 \le i \le n-1$ the design will be called degenerated. The set of all nondegenerated designs is not compact. Therefore, we will consider a compact subset

$$A_\delta = \{T = (t_1, \ldots, t_n) \mid a \le t_1 < \cdots < t_n \le b;\ t_{i+1} - t_i \ge \delta,\\ i = 1, 2, \ldots, n-1\},$$

where δ is a given positive number. Let $\{\varphi_1, \ldots, \varphi_n\}$ be an arbitrary basis in the space of natural spline polynomials $N^{2m}(t_1, \ldots, t_n)$, where $\varphi_i(t) = \varphi_i(t, T)$, $i = 1, \ldots, n$, $T = (t_1, \ldots, t_n)$ is a nondegenerated design on the interval $[a, b]$. A design $T^* \in A_\delta$ will be called G_δ-optimal if

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

A design T^* will be called nongenerated G-optimal design if it is G_δ-optimal for all sufficiently small positive δ. Due to compactness of the set A_δ a G_δ-optimal design always exists. The proof can be found in the Appendix.

Lemma 2 *A G_δ-optimal design does not depend on the basis of the set $N^{2m}(t_1, \ldots, t_n)$*

Proof of Lemma 2 Let $T = (t_1, \ldots, t_n)$ be a nondegenerated design, $\varphi(t) = (\varphi_1(t), \ldots, \varphi_n(t))^T$, $\varphi(t) = \varphi(t, T)$ be an arbitrary basis in $N^{2m}(t_1, \ldots, t_n)$. If $\tilde{\varphi}(t)$ is another basis, then

$$\tilde{\varphi}(t) = K\varphi(t),$$

where K is a non-singular $n \times n$ matrix, which can depend on the design points $t_1, \ldots, t_n$ but does not depend on variable t. Let

$$M(T) = \sum_{i=1}^{n} \varphi(t_i)\varphi^T(t_i) + \lambda \int_a^b \varphi^{(m)}(t) \left(\varphi^{(m)}(t)\right)^T \mathrm{d}t$$

and

$$\bar{M}(T) = K M(T) K^T$$

be the information matrix for basis φ and $\tilde{\varphi}$, respectively. Then we obtain

$$\tilde{\varphi}^T(t) M^{-1}(T) \tilde{\varphi}(t) = \varphi^T(t) K^T \left[K M(T) K^T\right]^{-1} K\varphi(t) = \varphi^T(t) M^{-1}(T) \varphi(t),$$

which implies the independence of the G_δ-optimal design problem with respect to the basis. □

Conjecture 1 *For arbitrary $n > m$ and $\lambda = 0$ there exists a nondegenerated G-optimal design.*

A proof of Conjecture 1 for the cases $m = 1$ and 2 is contained in the proofs of Theorems 4 and 5, respectively. Throughout the remaining part of this paper we denote a nondegenerated G-optimal design as G-optimal design. From Lemma 2 it follows that a G_δ-optimal design does not depend on the basis. However, it should be noted that the D-optimal design problem depends, generally speaking, on the given basis of $N^{2m}(t_1, \ldots, t_n)$ since

$$\det \overline{M}(T) = (\det K)^2 \det M(T)$$

and the matrix K can depend on the design T. In the following we will introduce and study a special basis which turns out to be very useful for the solution of the D-optimal design problem.

It is known (see e.g. Eubank 1999, Chapter 5) that a basis in $N^{2m}(t_1, \ldots, t_n)$ can be formed from the monomial and truncated power functions

$$f_1(t) = 1, \;\; f_2(t) = t, \ldots, t_m(t) = t^{m-1},$$
$$\psi_1(t) = (t - t_1)_+^{2m-1}, \ldots, \psi_n(t) = (t - t_n)_+^{2m-1}.$$

Let us introduce the following notations:

$$f(t) = (f_1(t), \ldots, f_m(t))^T = \left(1, t, \ldots, t^{m-1}\right)^T, \quad \psi(t) = (\psi_1(t), \ldots, \psi_n(t))^T,$$
$$F = F(T) = \left(t_i^{j-1}\right)_{i,j=1}^{n}, \quad Q = Q(T) = \left(t_i^{j-1}\right)_{i,j=1}^{nm},$$
$$d_{ij}(T) = h(T) e_{i+m}^T F^{-1} e_j, \qquad i = 1, \ldots, n-m, \quad j = 1, \ldots, n,$$

$e_1 = (1, 0, \ldots, 0)^T, \ldots, e_n = (0, \ldots, 0, 1)^T \in \mathbb{R}^n$, $h(T) = (\det F)^{\frac{1}{n-m}}$

$$D = D(T) = \left(d_{ij}(T)\right)_{i,j=1}^{n-m\, n}, \quad R = R(T) = \left((t_i - t_j)_+^{2m-1}\right)_{i,j=1}^{n}. \tag{16}$$

The following two results are proved in the Appendix.

Lemma 3 *The following relations hold:*

(a)

$$\det \begin{pmatrix} d_{1\,j_1} & \ldots & d_{1\,j_{n-m}} \\ \vdots & \ddots & \vdots \\ d_{n-m\,j_1} & \ldots & d_{n-m\,j_{n-m}} \end{pmatrix} = \det \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ t_{l_1}^{m-1} & \ldots & t_{l_m}^{m-1} \end{pmatrix} \tag{17}$$

for arbitrary $1 \le j_1 < \cdots < j_{n-m} \le n$, $1 \le l_1 < \cdots < l_m \le n$, $l_i \in \{1, \ldots, n\} \backslash \{j_1, \ldots, j_{n-m}\}$, $i = 1, \ldots, m$.

(b)

$$DQ = \mathbf{O}, \tag{18}$$

where $\mathbf{O}$ *is the matrix of size* $(n-m) \times m$ *with all entries equal to 0.*

Lemma 4 *Let $a \le t_1 < t_2 < \cdots < t_n \le b$. Then the functions*

$$1, t, \ldots, t^{m-1}, \sum_{j=1}^{n} d_{1j}\psi_j(t), \ldots, \sum_{j=1}^{n} d_{n-m,j}\psi_j(t) \tag{19}$$

generate a basis in $N^{2m}(t_1, \ldots, t_n)$.

The basis defined by (19) possesses a number of remarkable properties which will be presented in the following Lemma. For its formulation we introduce the notation

$$\begin{aligned} \bar{\varphi}_i(t) &= t^{i-1}, \quad i = 1, 2, \ldots, m, \\ \bar{\varphi}_{i+m}(t) &= \sum_{i=1}^{n} d_{ij}(t - t_j)_+^{2m-1}, \quad i = 1, \ldots, n-m, \\ \bar{X}(T) &= (\bar{\varphi}_i(t_i))_{i,j=1}^{n}, \\ \bar{V}(T) &= \left(\int_a^b \bar{\varphi}_{i+m}^{(m)}(t)\bar{\varphi}_{j+m}^{(m)}(t)\,\mathrm{d}t\right)_{i,j=1}^{n-m}. \end{aligned} \tag{20}$$

Lemma 5 *For arbitrary m and $n > m$ the following identities are valid:*

$$\det \bar{X}(T) = \det(DRD^T) = (-1)^{(n+1)} \det \begin{pmatrix} 0 & Q^T \\ Q & R \end{pmatrix},$$
$$\det \bar{V}(T) = \kappa^{(n-m)} \det \bar{X}(T).$$

where $\kappa = (-1)^m(2m-1)!$

It follows from Lemma 5, which will be proved in the Appendix, that a design maximizing determinant $\bar{X}(T)\bar{X}(T)^T$ will also maximize $\det \bar{V}(T)$. Note that for arbitrary λ the optimality criterion is of the form $\det[\bar{X}(T)\bar{X}^T(T) + (\lambda/\kappa)\bar{V}(T)]$. Therefore, a design maximizing simultaneously the determinants of the matrices $\bar{X}(T)\bar{X}^T(T)$ and $\bar{V}(T)$ is very likely nearly optimal for all values of λ. This heuristical argument is confirmed by our numerical calculations. A further nice property of the basis consists in the fact that the D-optimal design is also G-optimal in the case $\rho = \infty$ (see Theorems 3–5 in the next section).

Remark 1 Denote by $\bar{M}(T)$ the information matrix for the basis defined in (19). Throughout this paper a design is called D-optimal if it maximizes $\det \bar{M}(T)$ in the set of all (n-points) designs. Since with $t_i = t_{i+1}$ for some i we have $\det \bar{M}(T) = 0$ the solution of the problem is achieved in the subset of nondegenerated designs. This solution maximizes $\det M(T)$ for any basis φ which is constructed from the basis $\bar{\varphi}$ by multiplication on arbitrary nondegenerated $n \times n$ matrix which does not depend on $t_1, \ldots, t_n$.

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 in the previous section). We begin with an invariance property of the optimal designs, which reduces the design problem to the interval $[-1, 1]$.

Theorem 1 *Let $T = \{t_1, \ldots, t_n\}$ denote the D- (G-) optimal design for estimation with the smoothing spline on the interval $[-1, 1]$, 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 $[a, b]$.

A proof of this theorem and of the next one is deferred to the Appendix for convenience of the reader. Theorem 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 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. $t_{n-i+1} = -t_i$, $i = 1, \ldots, n$.*

Lemma 6 *If there exists a design $\bar{T}$ on the interval $[-1, 1]$, which satisfies*

$$\max_{t \in [-1,1]} \varphi^T(t, \bar{T}) M^{-1}(\bar{T}) \varphi(t, \bar{T}) = \operatorname{tr}\left(X^T(\bar{T}) X(\bar{T}) M^{-1}(\bar{T})\right), \tag{21}$$

$$\operatorname{tr}\left(X^T(\bar{T}) X(\bar{T}) M^{-1}(\bar{T})\right) = \min_T \operatorname{tr}\left(X^T(T) X(T) M^{-1}(T)\right), \tag{22}$$

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

A proof of Lemma 6 can be found in the Appendix. 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 (22) of Lemma 6 is trivially satisfied. Note also that the case $\lambda = 0$ corresponds to an interpolating spline which are usually 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, the optimal designs calculated for the case $\lambda = 0$ are also useful for moderate values of λ, and the choice $\lambda = 0$ will now be considered in more detail 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. 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 here 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.

$$Y = X(T, U)\theta + \varepsilon. \tag{23}$$

Here, $Y = (Y_1, \dots, Y_n)^T$ denotes the vector of observations in the nonparametric regression model (1), $\theta = (\theta_1, \dots, \theta_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ denotes a vector of centred and uncorrelated random variables,

$$X(T, U) = (\varphi_i(t_j, U))_{i,j=1}^n \tag{24}$$

and $\varphi_1(t, U), \dots, \varphi_n(t, U)$ are the basis functions of $N^{2m}(u_1, \dots, u_n)$ defined by (20), where the points t_j have been replaced by u_j $(j = 1, \dots, n)$. In what follows define

$$R(T, U) = ((t_i - u_j)_+^{2m-1})_{i,j=1}^n, \tag{25}$$

$$F(T) = (t_i^{j-1})_{i,j=1}^m, \quad F(U) = (u_i^{j-1})_{i,j=1}^m, \tag{26}$$

$$\begin{aligned} d_i(T) &= (\det F(T))(F^T(T))^{-1} e_{m+i}, \\ d_i(U) &= (\det F(U))(F^T(U))^{-1} e_{m+i}, \end{aligned} \tag{27}$$

$$D(T) = (d_{ij}(T))_{i=1,\dots,n-m}^{j=1,\dots,n}, \quad D(U) = (d_{ij}(U))_{i=1,\dots,n-m}^{j=1,\dots,n}, \tag{28}$$

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

Lemma 7 *We have*

$$\begin{aligned} [\det(D(T)R(T, U)D^T(U))]^2 \le\ & \det(D(T)R(T, T)D^T(T)) \\ & \times \det(D(U)R(U, U)D^T(U)) \end{aligned}$$

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

For the formulation of the following result we consider an approximate design ξ with weights $\xi_1, \dots, \xi_N$ at the points $t_1, \dots, t_N$ (see e.g. Kiefer 1974) and its information matrix

$$\tilde{M}(\xi, U) = \sum_{i=1}^N \varphi(t_i, U)\varphi^T(t_i, U)\xi_i. \tag{29}$$

Theorem 3 *Consider the smoothing spline defined by* (4) *with* $n > m$ *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 also G-optimal for estimation with the smoothing spline.*

We will conclude this section with a full solution for the case $m = 1$, $\rho = \infty$ and the proof of the hypothesis used in the formulation of Theorem 3 for $m = 2$, $\rho = \infty$ (see the Appendix for details).

Theorem 4 *Let* $m = 1$, $\rho = \infty$ *and* $n \in \mathbb{N}$.

(i) *There exists a unique D-optimal design; it consists of the points*

$$t_1 = a, \quad t_2 = a + \frac{(b-a)}{n-1}, \ldots, \quad t_{n-1} = a + \frac{(b-a)}{n-1}(n-2), \quad t_n = b; \quad (30)$$

(ii) *Any nondegenerated approximate design is a (nondegenerated) G-optimal design.*

Theorem 5 *Let* $m = 2$, $\rho = \infty$ *and* $n \in \mathbb{N}$*, then a design is D-optimal if and only if it is G-optimal.*

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 size. The second example refers to some numerical calculation and comparison of D- and G-optimal designs for estimation with quadratic ($m = 2$) and cubic ($m = 3$) smoothing splines. Note that the case of linear splines ($m = 1$) and $\rho = \infty$ was fully studied theoretically in Theorem 4.

Example 1 Consider the case $m = 2$ and $\rho = \infty$. Without loss of generality we can take $[a, b] = [-1, 1]$. Using the representation for the determinant from Lemma 5 we can find explicitly optimal designs for moderate size of sample. After simple transformations we obtain

$$\det M(T)^{1/2} = \det \begin{pmatrix} 0 & Q^T \\ Q & R \end{pmatrix}$$
$$= \det \begin{pmatrix} 0 & t_1 - t_n & t_2 - t_n & \ldots & t_{n-1} - t_n \\ t_2 - t_1 & t_2 - t_1 & 0 & \ldots & 0 \\ t_3 - t_1 & t_3 - t_1 & t_3 - t_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ t_n - t_1 & t_n - t_1 & t_n - t_2 & \ldots & t_n - t_{n-1} \end{pmatrix}.$$

Note that for D-optimal design $t_1 = -1$, $t_n = 1$ due to Theorem 2. For $n = 3$ we have

$$\det M(T) = [2(1 - t_2^2)]^2.$$

Therefore, the unique D-optimal (and G-optimal) design is given by $T^* = (-1, 0, 1)$. Note that this design was also identified as G-optimal design for a quintic smoothing spline (see Butler 2001, Theorem 4). If $n = 4$ we restrict ourselves to symmetric designs of the form $T_x = (-1, -x, x, 1)$, $x > 0$. A direct but tedious calculation shows that

$$\det M(T_x) = [16(1 - x)^4 x^2 (2x + 1)]^2.$$

With the notation $r(x) = \ln[\det M(T_x)]^{1/2}$ it follows

$$(\ln r(x))' = \frac{4}{x - 1} + \frac{2}{x} + \frac{2}{2x + 1} = \frac{4(1 - 7x^2)}{x(2x + 1)(1 - x)}.$$

Thus the unique design T_x which maximizes $\det M(T_x)$ is given by $(-1, -1/\sqrt{7}, 1/\sqrt{7}, 1)$. Using the Kiefer–Wolfowitz equivalence theorem it is easy to check that the design is unique D-(G-) optimal design in the class of all designs.

Example 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 $m = 2$ considered in Theorem 5 and the case $m = 3$.

Table 1 D-, G- and I-optimal designs for estimation with cubic smoothing spline, $m = 2$ (upper part), and 4-degree smoothing spline, $m = 3$ (lower part)

		D-optimal			G-optimal			I-optimal		
m	λ	0	0.01	0.100	0	0.01	0.100	0	0.01	0.100
2	t_1	−1	−1	−1	−1	−1	−1	−1	−1	−1
	t_2	−0.64	−0.64	−0.64	−0.64	−0.78	−0.85	−0.63	−0.58	−0.49
	t_3	−0.21	−0.21	−0.21	−0.21	−0.30	−0.23	−0.21	−0.18	−0.22
	t_4	0.21	0.21	0.21	0.21	0.30	0.23	0.21	0.18	0.22
	t_5	0.64	0.64	0.64	0.64	0.78	0.85	0.63	0.58	0.49
	t_6	1	1	1	1	1	1	1	1	1
3	t_1	−1	−1	−1	−1	−1	−1	−1	−1	−1
	t_2	−0.70	−0.70	−0.70	−0.70	−0.72	−0.76	−0.69	−0.68	−0.66
	t_3	−0.24	−0.24	−0.24	−0.24	−0.26	−0.29	−0.23	−0.23	−0.22
	t_4	0.24	0.24	0.24	0.24	0.26	0.29	0.23	0.23	0.22
	t_5	0.70	0.70	0.70	0.70	0.72	0.76	0.69	0.68	0.66
	t_6	1	1	1	1	1	1	1	1	1

Number of knots: $n = 6$

Table 2 *D*-, *G*- and *I*-optimal designs for estimation with cubic smoothing spline, $m = 2$ (upper part), and 4-degree smoothing spline, $m = 3$ (lower part)

		D-optimal			*G*-optimal			*I*-optimal		
m	λ	0	0.01	0.100	0	0.01	0.100	0	0.01	0.100
2	t_1	−1	−1	−1	−1	−1	−1	−1	−1	−1
	t_2	−0.81	−0.81	−0.82	−0.81	−0.95	−0.99	−0.79	−0.72	−0.94
	t_3	−0.58	−0.58	−0.58	−0.58	−0.66	−0.89	−0.56	−0.55	−0.46
	t_4	−0.35	−0.35	−0.35	−0.35	−0.43	−0.33	−0.33	−0.32	−0.45
	t_5	−0.12	−0.12	−0.12	−0.12	−0.12	−0.31	−0.11	−0.11	−0.01
	t_6	0.12	0.12	0.12	0.12	0.12	0.31	0.11	0.11	0.01
	t_7	0.35	0.35	0.35	0.35	0.43	0.33	0.33	0.32	0.45
	t_8	0.58	0.58	0.58	0.58	0.66	0.89	0.56	0.55	0.46
	t_9	0.81	0.81	0.82	0.81	0.95	0.99	0.79	0.72	0.94
	t_{10}	1	1	1	1	1	1	1	1	1
3	t_1	−1	−1	−1	−1	−1	−1	−1	−1	−1
	t_2	−0.85	−0.84	−0.84	−0.85	−0.93	−0.95	−0.84	−0.79	−0.76
	t_3	−0.61	−0.61	−0.61	−0.61	−0.71	−0.73	−0.60	−0.55	−0.60
	t_4	−0.37	−0.36	−0.37	−0.37	−0.43	−0.47	−0.36	−0.33	−0.33
	t_5	−0.12	−0.12	−0.12	−0.12	−0.15	−0.15	−0.12	−0.11	−0.12
	t_6	0.12	0.12	0.12	0.12	0.15	0.15	0.12	0.11	0.12
	t_7	0.37	0.36	0.37	0.37	0.43	0.47	0.36	0.33	0.33
	t_8	0.61	0.61	0.61	0.61	0.71	0.73	0.60	0.55	0.60
	t_9	0.85	0.84	0.84	0.85	0.93	0.95	0.84	0.79	0.76
	t_{10}	1	1	1	1	1	1	1	1	1

Number of knots: $n = 10$

In Tables 1 and 2 we present the *D*- and *G*-optimal designs for the cases $m = 2$ and $m = 3$ for $n = 6$ and $n = 10$ knots. Further results for $n = 20$ knots can be obtained from the authors. It is interesting to note that the *D*-optimal designs do not change substantially for different values of the parameter λ. In the case of *G*-optimal designs the situation is completely different. Here a larger value of the smoothing parameter λ yields to a *G*-optimal design which is more concentrated at the boundary. This corresponds to intuition, because a larger value of λ 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 coincide with the *D*-optimal designs. On the other hand, if $\lambda > 0$, the *D*-optimal designs do not change substantially, 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

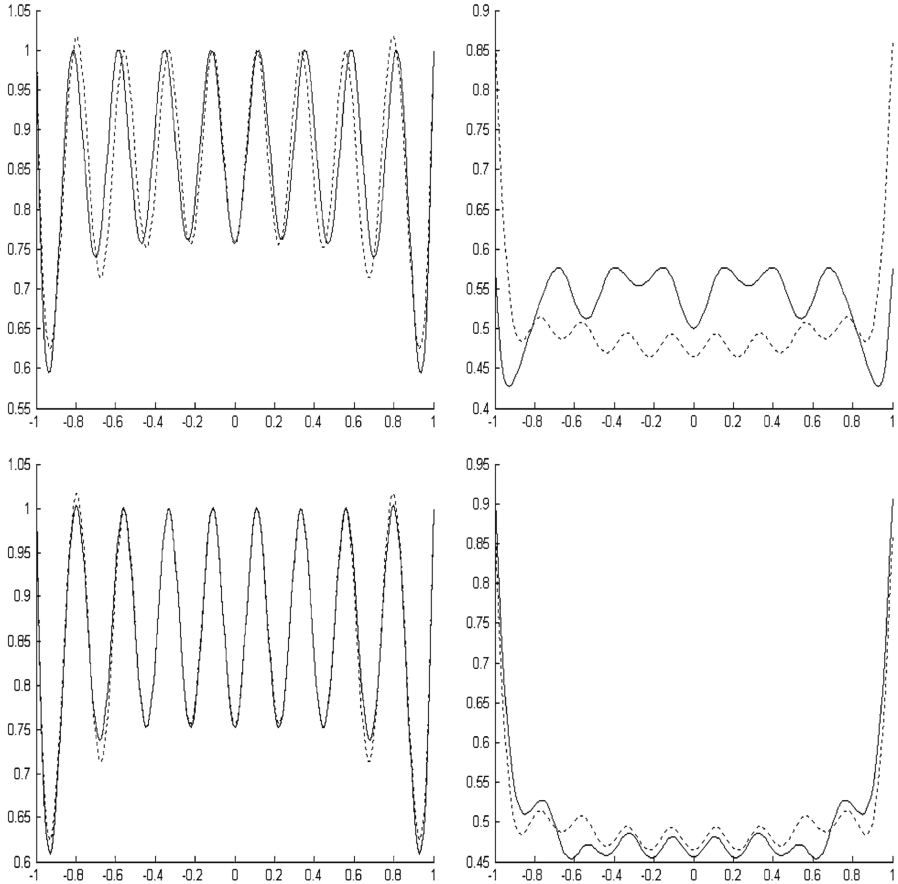

Fig. 1 The function $d(t, \xi)$ defined in (31) for various optimal designs and different values of the parameter λ (*left part*: $\lambda = 0$, *right part*: $\lambda = 0.01$. *Upper panel*: comparison of G-optimal (*solid line*) and uniform design (*dashed line*). *Lower panel*: comparison of I-optimal (*solid line*) and uniform design (*dashed line*). The number of observations is $n = 10$

visible. In particular, in the case $m = 2$ the G-optimal designs are more concentrated at the boundary of the design space.

It might also be of interest to investigate the function

$$d(t) = \varphi^T(t)(X^T X + \lambda B)^{-1}\varphi(t) \tag{31}$$

for the different designs. In the upper panel of Fig. 1 we show 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 $t \to d(t)$. 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"

$$\int_{-1}^{1} \varphi^T(t)(X^T X + \lambda B)^{-1} \varphi(t)\,\mathrm{d}t$$

and are called I-optimal designs for estimation with smoothing splines. For this criterion results similar to Theorems 1 and 2 can be obtained by the same arguments. For the sake of brevity we restrict ourselves to a discussion of numerical results.

Some I-optimal designs are shown for the cases $m = 2$ and $m = 3$ in Tables 1 and 2. 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 Fig. 1 that in the case $\lambda > 0$ the uniform design yields smaller values for the function $d(t)$ defined in (31) 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 for estimating with the smoothing spline and the uniform design.

Acknowledgments The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen", SFB 823 "Statistik nichtlinearer dynamischer Prozesse") is gratefully acknowledged. The work by V.B.Melas was partly supported by the Russian Foundation of Basic Research (project 09-01-00508). The work of H. Dette was supported in part by a NIH grant award IR01GM072876:01A1. The authors would also like to thank the reviewers for their constructive comments on an earlier version of this manuscript.

Appendix: Proofs

Proof of Lemma 1 For fixed $k \in \{1, \ldots, n\}$ we define for any vector $a \in \mathbb{R}^n$ the function

$$H(a) = \sup_{g \in \mathcal{F}_\rho} E[(g(t_k) - a^T Y)^2] \tag{32}$$

and note that this function depends on the class $\mathcal{F}_\rho$ only through the points $t_1, \ldots, t_n$. Consequently, it follows from well known properties of natural splines (see e.g. Karlin and Studden 1966, Section 11.9) that the supremum in (32) is attained at a function $g^* \in N^{2m}(t_1, \ldots, t_n) \cap \mathcal{F}_\rho$, that is

$$H(a) = \sup_{g \in N^{2m}(t_1,\ldots,t_n) \cap \mathcal{F}_\rho} E[(g(t_k) - a^T Y)^2]. \tag{33}$$

Note that all functions $g \in N^{2m}(t_1, \ldots, t_n) \cap \mathcal{F}_\rho$ are of the form (7) for some $\theta \in \Omega$, and define

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

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

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

where λ 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 (11). □

Proof of Lemma 3 Formula (17) is a particular case of the theorem on minors of inverse matrices (see, e.g. Gantmacher 1998). For a proof of assertion (b) we note that the matrix D is of the form

$$D = h(T)\left(\mathbf{O} \vdots I\right)F^{-1}, \tag{34}$$

where I is the unit matrix of size $(n-m)\times(n-m)$. From this formula the assertion follows by a direct calculation. □

Proof of Lemma 4 It is known (see e.g. Eubank 1999, Chapter 5) that the functions $1, t_1, \ldots, t^{m-1}, \psi_1(t), \ldots, \psi_n(t)$ are linearly independent on the interval $[a, b]$. It follows from (34) that the matrix D has full rank. Therefore, the functions defined by (19) are linearly independent.

It remains to prove that the function $g(t) = \sum_{i=1}^n a_i\bar{\varphi}_i(t)$ is a polynomial of degree $\leq m-1$ for $t < a$ and $t > b$, where $\bar{\varphi}(t) = t^{i-1}$, $i = 1, 2, \ldots, m$, $\bar{\varphi}_{i+m}(t) = \sum_{j=1}^n d_{ij}(t-t_j)_+^{2m-1}$, $i = 1, \ldots, n-m$, $a_1, \ldots, a_n$ are arbitrary constants.

For $t < a$ we have $(t-t_j)_+^{2m-1} = 0$, $j = 1, \ldots, n$ and, therefore $g(x) = \sum_{i=1}^m a_it^{i-1}$ for $t < a$. For $t > b$ we have $(t-t_j)_+^{2m-1} = (t-t_j)^{2m-1}$, $j = 1, \ldots, n$. Note that we obtain from (17) $\sum_{j=1}^n d_{ij}t_j^s = 0$, $i = 1, 2, \ldots, n-m$, $s = 0, 1, \ldots, m-1$. Therefore we have for $t > b$

$$\sum_{j=1}^n d_{ij}(t-t_j)_+^{2m-1} = \sum_{j=1}^n d_{ij}(t-t_j)^{2m-1} = \sum_{j=0}^{n-1} b_{ij}t^j,$$

where

$$b_{ij} = \sum_{k=1}^{n} d_{ik} t_k^{2m-1-j} C_{2m-1}^{j} (-1)^j.$$

It follows that $g(t) = \sum_{j=1}^{m} b_j t^{j-1}$ for $t > b$ for some constants $b_1, \ldots, b_m \in \mathbb{R}$. □

Proof of Lemma 5 Note that $\bar{X} = (Q \vdots RD^T)$, where $Q = Q(T)$, $R = R(T)$ are defined in (16). By Laplace rule we obtain

$$\det \bar{X} = \sum \det \left(t_{j_k}^{i-1}\right)_{i,k=1}^{m} \det \left(b_{is_k}\right)_{i,k=1}^{n-m} (-1)^{\pi(J)},$$

where $b_{ij} = [RD^T]_{(i,j)}$, and the sum is taken over all sets of indexes

$$\begin{aligned} &1 \le j_1 < \cdots < j_m \le n, \quad J = (j_1, \ldots, j_m), \\ &1 \le s_1 < s_2 < \cdots < s_{n-m} \le n, \\ &s_i \in \{1, 2, \ldots, n\} \backslash \{j_1, \ldots, j_m\}, \quad i = 1, 2, \ldots, n-m \\ &\pi(J) = j_1 + \cdots + j_m. \end{aligned}$$

Therefore, it follows

$$\det(\bar{X}) = \det(DRD^T) = (-1)^{(n+1)} \det \begin{pmatrix} 0 & X^T \\ X & R \end{pmatrix}.$$

Lemma 5.1 from (Eubank 1997, Chapter 5) finally yields

$$\bar{V} = \kappa(DRD^T), \quad \det \bar{V} = \kappa^{(n-m)} \det \bar{X}(T).$$

□

Proof of Theorem 1 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 φ_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 (19) satisfies

$$\varphi(ct + d, \bar{T}) = S\varphi(t, T), \tag{35}$$

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

$$S = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}, \tag{36}$$

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

$$J_2 = \operatorname{diag}\Big(c^{m-2}(-1)^s, c^{m-3}(-1)^{s+1}, \ldots, c^{2m-n-1}(-1)^{s+n-m-1}\Big) \in \mathbb{R}^{n-m\times n-m} \tag{37}$$

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

$$\begin{aligned} M(\bar{T}) &= X^T(\bar{T})X(\bar{T}) + \lambda B \\ &= \sum_{i=1}^n \varphi(\bar{t}_i, \bar{T})\varphi^T(\bar{t}_i, \bar{T}) + \lambda \int_a^b \varphi^{(m)}(t, \bar{T})(\varphi^{(m)}(t, \bar{T}))^T \, \mathrm{d}t \\ &= \sum_{i=1}^n S\varphi(t_i, T)\varphi^T(t_i, T)S^T + \lambda \int_{-1}^1 S\varphi^{(m)}(t, T)(\varphi^{(m)}(t, rT))^T S^T \, \mathrm{d}t \\ &= SM(T)S. \end{aligned}$$

Because the matrix S does not depend on the design T, the D-optimal design on the interval $[a, b]$ (maximizing det $M(\bar{T})$ with respect to $\bar{T}$) can be obtained from the D-optimal design on the interval $[-1, 1]$ (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

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

and it remains to prove the identity (35). For this purpose recall the definition of the matrix $F(T)$ and the vector $d_i(T)$ in (6) and (7); then it is easy to see that

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

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 j component $d_{ij}(\bar{T})$ of the vector $d_i(\bar{T})$ yields

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

This proves the representation

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

where the matrix J_2 is defined in (37). Finally, the representation

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

with a lower triangular matrix J_1 is obvious and the assertion of Theorem 1 follows. □

Proof of Theorem 2 (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 1 we have for the design $\bar{T}=\{\bar{t}_1,\ldots,\bar{t}_n\}$ on the interval $[-1,1]$

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

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

(ii) This follows similarly observing the identity

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

(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. □

Proof of Lemma 6 We have for any design $T=\{t_1,\ldots,t_n\}$ on the interval $[-1,1]$

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

Consequently, we obtain from (21) and (22)

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

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

Proof of Lemma 7 With the same arguments as presented in the proof of Lemma 5 it follows

$$\kappa \det(D(T)R(T,U)D^T(U)) = \det V(T,U), \tag{38}$$

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

$$V(T,U) = \left(\int_a^b \varphi_{m+i}^{(m)}(t,T)\varphi_{m+j}^{(m)}(t,U)\,\mathrm{d}t\right)_{i,j=1,\ldots,n-m}$$

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

$$\begin{aligned}
\det V(T,U) &= \det\left(\int_a^b \varphi_{m+i}^{(m)}(t,T)\varphi_{m+j}^{(m)}(t,U)\,\mathrm{d}t\right)_{i,j=1,\ldots,n-m}\\
&= \int_a^b\cdots\int_a^b \det\left(\varphi_{m+i}^{(m)}(z_j,T)\right)_{i,j=1}^{n-m}\\
&\quad\times\det\left(\varphi_{m+j}^{(m)}(z_i,U)\right)_{i,j=1}^{n-m}\,\mathrm{d}z_1\ldots\mathrm{d}z_{n-m}\\
&\le\left\{\int_a^b\cdots\int_a^b\left[\det\left(\varphi_{m+i}^{(m)}(z_j,T)\right)_{i,j=1}^{n-m}\right]^2\mathrm{d}z_1\ldots\mathrm{d}z_{n-m}\right.\\
&\quad\left.\times\int_a^b\cdots\int_a^b\left[\det\ \left(\varphi_{m+i}^{(m)}(z_i,U)\right)_{i,j=1}^{n-m}\right]^2\mathrm{d}z_1\ldots\mathrm{d}z_{n-m}\right\}^{1/2}\\
&= \{\det V(T,T)\det V(U,U)\}^{1/2},
\end{aligned}$$

and the assertion of Lemma 7 follows. □

Proof of Theorem 3 Let $T^*=\{t_1^*,\ldots,t_n^*\}$ denote a D-optimal design for the smoothing spline and $\tilde{\xi}$ be an approximate design maximizing $\det\tilde{M}(\xi,T^*)$ with $N=n$ support points. 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). If $\tilde{T}$ denotes the support of $\tilde{\xi}$ we obtain from Lemma 7 and the formula given in Lemma 5

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

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 ξ^* is also an approximate G-optimal design, which means that it minimizes

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

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

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

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

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

which proves that the design T^* is also G-optimal. □

Proof of Theorem 4 Due to Lemma 5 we have

$$\det \bar{X}(T) = (-1)^{n+1} \det \begin{pmatrix} 0 & Q^T \\ Q & R \end{pmatrix},$$

where in the case $m = 1$ the matrix on the right-hand side reduces to

$$\begin{pmatrix} 0 & Q^T \\ Q & R \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & t_1 - t_2 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & t_n - t_1 & t_n - t_2 & \ldots & t_n - t_{n-1} & 0 \end{pmatrix}.$$

Subtracting the last column from 2-nd,..., $(n-1)$-th column and then subtracting the second line from 3-rd,..., n-th lines we obtain

$$\begin{aligned}\det \begin{pmatrix} 0 & Q^T \\ Q & R \end{pmatrix} &= \det \begin{pmatrix} 0 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & t_1 - t_2 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & t_n - t_1 & t_n - t_2 & \ldots & t_n - t_{n-1} & 0 \end{pmatrix}\\ &= (-1)^{n+1} \prod_{i=1}^{n-1} (t_{i+1} - t_i).\end{aligned}$$

Therefore, det $\bar{M}(T) = \prod_{i=1}^{n-1}(t_{i+1} - t_i)^2$. Using the inequality between arithmetic and geometric means we obtain that the maximum is achieved if and only if the design has the form (30).

In order to prove assertion (ii) we consider the function

$$g(t) = \varphi^T(t, U)M^{-1}(T, U)\varphi(t, U),$$

where $\varphi(t, U) = \bar{\varphi}(t, U)$ with $U = T$. A direct calculation shows that

$$g(t_i) = 1, \quad i = 1, 2, \ldots, n.$$

Moreover, $g(t)$ is a quadratic polynomial at each of the intervals (t_i, t_{i+1}), $i = 1, 2, \ldots, n-1$, while the highest coefficient is strictly positive. Therefore it follows that $g(t) < 1$, whenever $t \in (t_i, t_{i+1})$, $i = 1, \ldots, n-1$. An application of the usual equivalence theorem shows that T is a G-optimal design. □

Proof of Theorem 5 Let $T^* = (t_1^*, \ldots, t_n^*)$ denote a D-optimal design (in the sense explained in Remark 1). Define ξ^* the corresponding n-point approximate design, i.e.

$$\xi^* = \begin{pmatrix} t_1^* & \ldots & t_n^* \\ 1/n & \ldots & 1/n \end{pmatrix},$$

and consider model (4.6) with $U = T^*$, $\varphi(t, U) = \bar{\varphi}(t, U)$. Let

$$g(t) = \varphi^T(t, U)\tilde{M}^{-1}(\xi^*, U)\varphi(t, U) - n,$$

where $\tilde{M}(\xi^*, U) = \frac{1}{n}\bar{M}(T^*, U)$. A direct calculation shows that

$$\begin{aligned} g(t_i^*) &= 0, \quad i = 1, \ldots, n, \\ g'(t_i^*) &= 0, \quad i = 2, \ldots, n-1, \\ g(t) &\to \infty, \quad |\, t \,| \to +\infty. \end{aligned}$$

Note that at each of the interval (t_i, t_{i+1}), $i = 1, 2, \ldots, n-1$ the function $g(t)$ is a polynomial of degree $2(2m-1) = 6$. Suppose that there exists a point t such that $t \in (t_i, t_{i+1})$ for some $i \in \{1, 2, \ldots, n-1\}$ and $g(t) > 0$. Then, obviously, the function $g(t)$ has on the interval $[t_i, t_{i+1}]$ at least 8 zeros (counted with multiplicity), which is impossible. Thus $g(t) \leq 0, \quad t \in [a, b]$, which means that ξ^* is an approximate G-optimal design. Therefore, it follows from Theorem 3 that T^* is a G-optimal design. The converse assertion can be proved by a similar argument using the Kiefer–Wolfowitz equivalence theorem. □

References

Butler, N. A. (2001). Properties of optimal polynomial designs under smoothing spline models. *Journal of Statistical Planning and Inference, 92*, 259–266.

De Boor, C. (1978). A practical guide to splines. In *Applied mathematical sciences* (Vol. 27, XXIV). New York: Springer-Verlag.

Dette, H., Melas, V. B., Pepelyshev, A. N. (2008). Optimal designs for free knot least squares splines. *Statistica Sinica, 18*, 1047–1062.

Dierckx, P. (1995). Curve and surface fitting with splines. In *Monographs on numerical analysis*. Oxford: Clarendon Press.
Efromovich, S. (1999). *Nonparametric curve estimation: Methods, theory and applications*. New York: Springer.
Eubank, R. L. (1999). Nonparametric regression and spline smoothing (2nd ed.). In *Statistics: Textbooks and monographs* (Vol. 157). New York: Marcel Dekker.
Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.
Gantmacher, F. R. (1998). *The theory of matrices*. Providence: Chelsea.
Green, P., Silverman, B. (1994). *Nonparametric regression and generalized linear models*. London: Chapman and Hall.
Heiligers, B. (1998). E-optimal designs for polynomial spline regression. *Journal of Statistical Planning and Inference, 75*, 159–172.
Heiligers, B. (1999). Experimental design for polynomial spline regression. *Tatra Mountains Mathematical Publications, 17*, 157–165.
Jupp, D. L. B. (1978). Approximation to data by splines with free knots. *SIAM Journal on Numerical Analysis, 15*, 328–343.
Kaishev, V. K. (1989). Optimal experimental designs for the B-spline regression. *Computational Statistics & Data Analysis, 8*, 39–47.
Karlin, S., Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics*. New York: Wiley.
Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of statistics, 2*, 849–879.
Kiefer, J., Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics, 12*, 363–366.
Kuks, J., Olman, W. (1971). Minimaxnaja linejnaja ozenka coefficientov regressii II. *Izvestiya Akademii Nauk Ehstonskoj SSR 21*, 66–72.
Mao, W., Zhao, L. H. (2003). Free-knot polynomial splines with confidence intervals. *Journal of the Royal Statistical Society Series B, 65*(4), 901–919.
Murty, V. N. (1971a). Optimal designs with a Tchebycheffian spline regression function. *Annals of Mathematical Statistics, 42*, 643–649.
Murty, V. N. (1971b). Optimal designs with a polynomial spline regression with a single multiple knot at the center. *Annals of Mathematical Statistics, 42*, 952–960.
Park, S. H. (1978). Experimental designs for fitting segmented polynomial regression models. *Technometrics, 20*, 151–154.
Pukelsheim, F. (1993). *Optimal design of experiments*. New York: Wiley.
Reinsch, C. (1967). Smoothing by spline functions. *Numerische Mathematik, 10*, 177–183.
Schoenberg, I. J. (1964). Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences, 52*, 947–950.
Silverman, B. W. (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.
Silvey, S. D. (1980). *Optimal design*. London: Chapman and Hall.
Studden, W. J. (1971). Optimal designs and spline regression. Optimizing methods in statistics. In *Proceedings of the Symposium*, Ohio State University, Columbus, Ohio (pp. 63–76). New York: Academic Press.
Studden, W. J., VanArman, D. J. (1969). Admissible designs for polynomial spline regression. *Annals of Mathematical Statistics, 40*, 1557–1569.
Toutenburg, H. (1982). *Prior information in linear models*. New York: Wiley.
Wahba, G. (1990). Spline models for observational data. In *CBMS-NSF Regional Conference Series in Applied Mathematics* (Vol. 59). Philadelphia: SIAM.
Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society, 41*, 63–75.
Woods, D. C. (2005). Designing experiments under random contamination with application to polynomial spline regression. *Statistica Sinica, 15*, 619–635.
Woods, D., Lewis, S. (2006). All-bias designs for polynomial spline regression models. *Australian & New Zealand Journal of Statistics, 48*, 49–58.