Tatra Mt. Math. Publ. **17** (1999), 157–165

TATRA MOUNTAINS
Mathematical Publications

EXPERIMENTAL DESIGN FOR POLYNOMIAL SPLINE REGRESSION

Berthold Heiligers

ABSTRACT. We discuss design aspects in dth degree polynomial spline regression with prescribed knots over a compact regression interval. The B-basis provides an advantageous parameterization, ensuring equivariance of designs under rescaling of the setup. Numerical results for Kiefer's φ_p-criteria indicate that highly efficient designs are among those supported by the alternation points of the equi-oscillating spline.

1. Introduction

Polynomial models for analyzing regression functions of real-valued influence variables are well accepted in practice. Nevertheless, alternatives become desirable when good approximations to the true regression require high-degree polynomials, coming along with an unfavorable large number of parameters. Then polynomial splines, i.e, piecewise polynomials, might be attractive, since they combine excellent approximation properties with the advantages of low-degree polynomials. Interest in splines satisfying some smoothness conditions may also arise when the regression function can change its form at certain points. For example, the soot emission of a diesel engine depends (smoothly) on its turning speed, but it drastically changes when the turbocharger is activated at a particular speed.

In the present paper we discuss design aspects in a dth degree spline setup with fixed knots $a < \kappa_1 < \cdots < \kappa_{\ell-1} < b$ and associated smoothnesses; (for some results on spline designs we refer to Gaffke & Heiligers [5] and the references given there). That is, the regression function y is supposed to be a piecewise dth degree polynomial over the compact interval $[a, b]$, possessing s_i continuous derivatives at κ_i, $1 \leq i \leq \ell - 1$. Both, knots and smoothnesses are assumed to be known. Consequently, the associated spline space $\mathcal{S}$, say, is linear,

2000 Mathematics Subject Classification: 62K05.
Keywords: A-optimality, B-basis, Chebyshev spline, D-optimality E-optimality, equioscillation, Kiefer's φ_p-criteria, Remez exchange algorithm, total positivity, truncated powers.

and $\dim(\mathcal{S}) = \ell d + 1 - \sum_{i=1}^{\ell-1} s_i = k$. Hence, fixing a basis $\beta = (\beta_1, \ldots, \beta_k)'$ of $\mathcal{S}$,

$$y(x) = \sum_{\mu=1}^{k} \vartheta_\mu \, \beta_\mu(x) \,, \qquad x \in [a,b] \,. \tag{1.1}$$

Regression (1.1) is embedded in the usual statistical context: We assume that $\theta = \big(\vartheta_\mu\big)_{1 \le \mu \le k}$ is an unknown parameter, and that for all choices of x-values one can observe unbiased, homoscedastic and pairwise uncorrelated observations of y at x with variance σ^2. An approximate design ξ is a probability measure on $[a,b]$ with finite support, thus assigning $\xi(x) \cdot 100\,\%$ of the total sample size to the experimental conditions $x \in \mathrm{supp}(\xi)$. With a design ξ we associate its $k \times k$ information matrix,

$$M(\xi) = \sum_{x \in \mathrm{supp}(\xi)} \xi(x)\, \beta(x) \beta'(x) \,. \tag{1.2}$$

If ξ is an exact design of sample size n, i.e., if all weights $\xi(x)$ are integer multiples of $1/n$, then $M(\xi)$ is proportional to the inverse of the covariance matrix of the least squares estimator of θ based on $n \cdot \xi(x)$ observations of y at $x \in \mathrm{supp}(\xi)$. In practice, for moderate n, good exact designs are obtained by rounding the optimal approximate ones, (cf., e.g., P u k e l s h e i m [9], Chapter 12).

Design optimality refers to minimizing an optimality criterion, i.e., a convex function φ of (nonsingular) information matrices $M(\xi)$, which is nonincreasing w.r.t. the Loewner partial ordering. Among then, the most popular are the D-, A- and E-criteria, all of which fit into the class of Kiefer's φ_p-criteria, $-\infty \le p \le 1$, (with $p = 0, -1$, and $p = -\infty$, resp.),

$$\varphi_p\big(M(\xi)\big) = \begin{cases} \Big(\frac{1}{k}\,\mathrm{trace}\big[M^p(\xi)\big]\Big)^{-1/p} & \text{if } M(\xi) \text{ is nonsingular;} \\ \infty & \text{otherwise.} \end{cases}$$

Since the φ_p-criteria are functions of the eigenvalues of $M(\xi)$, they — apart from the D-criterion — sensitively depend on the parameterization of the setup, thus raising the question for an appropriate basis β in (1.1).

A natural choice of β is the truncated power basis,

$$\begin{aligned} \beta'(x) = \big(1, x, \ldots, x^d, (x-\kappa_1)_+^{s_1+1}, \ldots, (x-\kappa_1)_+^d, \ldots \\ \ldots, (x-\kappa_{\ell-1})_+^{s_{\ell-1}+1}, \ldots, (x-\kappa_{\ell-1})_+^d\big) \,, \end{aligned} \tag{1.3}$$

with $(x-a)_+^\mu = (x-a)^\mu$ if $x \ge a$, and $(x-a)_+^\mu = 0$ otherwise ($a \in \mathbb{R}$, $\mu \in \mathbb{N}_0$). That basis, however, shares the unfavorable property of ordinary monomials in polynomial setups, cf. Dette [4]: The entries of associated information matrices may be scaled quite differently, and thus most criteria do not satisfy the desirable "equivariance property". That is, when rescaling the regression interval, an optimal design is not obtainable from that in the original setup by rescaling the support points, keeping the weights fixed. As a possible way out of the problem, Dette [4] recommends the use of "standardized" criteria.

In Section 2 we address the same problem, but we propose a more rigorous solution: Instead of utilizing the truncated powers we recommend the (normalized) B-basis of $\mathcal{S}$ as a special local parameterization, ensuring equivariance under rescaling. Of course, the coefficients of y w.r.t. any local basis might not be directly interpretable, as, e.g., those of the x- or x^2-terms in (1.3), as linear or quadratic effects of x in the regression. This additional lack of interpretability, however, might be only virtual: The different ranges of the truncated monomials in the effective domain may lead to quite different scalings of estimates of the corresponding coefficients, thus overemphasizing (or neglecting) individual basis functions.

In Section 3 we present a Remez type exchange algorithm for obtaining the Chebyshev points, i.e., the alternation points of the (equi-oscillating) Chebyshev spline. Numerical examples indicate that they provide the support of highly efficient designs for all φ_p-criteria.

2. B-bases

For both reasons, ease of computation and well-conditioned design matrices, it is desirable that the basis β of $\mathcal{S}$ consists of splines with "small" supports. Among the spline bases, the "least supported" is the (normalized) B-basis $\beta = (\beta_1, \dots, \beta_k)$, defined next. For a discussion on B-bases and their properties we refer to Carnicer & Peña [1], [2].

Extend the knot vector κ by repeating $(d - s_i)$-times the interior knots κ_i, $1 \le i \le \ell - 1$, and $(d+1)$-times the boundary points a and b on the left and on the right, resp.,

$$\big(\underbrace{a, \dots, a}_{d+1}, \underbrace{\kappa_1, \dots, \kappa_1}_{d-s_1}, \dots, \underbrace{\kappa_{\ell-1}, \dots, \kappa_{\ell-1}}_{d-s_{\ell-1}}, \underbrace{b, \dots, b}_{d+1}\big) = (z_\nu)_{1 \le \nu \le k+d+1},$$

and define a family $\beta_{\mu,q}$, $1 \le \mu \le k + d - q$, $0 \le q \le d$, of functions on $[a, b]$ recursively as follows. Let, for $1 \le \mu \le k + d$,

$$\beta_{\mu,0}(x) = \begin{cases} 1 & \text{if } z_\mu \le x < z_{\mu+1}; \\ 0 & \text{else}; \end{cases}$$

and, for $1 \le q \le d,\ 1 \le \mu \le k + d - q$,

$$\beta_{\mu,q}(x) = w_{\mu,q}(x)\,\beta_{\mu,q-1}(x) \;+\; \bigl(1 - w_{\mu+1,q}(x)\bigr)\,\beta_{\mu+1,q-1}(x)\,, \quad \text{where}$$

$$w_{\mu,q}(x) = \begin{cases} (x - z_\mu)\big/(z_{q+\mu} - z_\mu) & \text{if } z_\mu < z_{\mu+q}\,; \\ 0 & \text{else}\,. \end{cases}$$

Then the dth degree B-splines β_μ are (cf., e.g., M ø r k e n [8], Theorems 10 and 11)

$$\beta_\mu(x) = \beta_{\mu,d}\,, \qquad 1 \le \mu \le k\,. \tag{2.1}$$

It is easily verified that

$$0 \le \beta_\mu \le 1 \text{ for all } \mu\,, \text{ and } \sum_\mu \beta_\mu \equiv 1\,; \tag{2.2}$$

moreover, β_μ vanishes outside the interval $[z_\mu, z_{\mu+d+1}]$ — more precisely we have

$$\bigl\{x\colon \beta_\mu(x) > 0\bigr\} = \begin{cases} [a, z_{d+2}) & \text{if } \mu = 1\,; \\ (z_\mu, z_{\mu+d+1}) & \text{if } 2 \le \mu \le k-1\,; \\ (z_k, b] & \text{if } \mu = k\,. \end{cases} \tag{2.3}$$

For example, in dth degree polynomial regression ($\ell = 1$), the B-basis consists of the Bernstein polynomials over $[a,b]$; Figure 2.1 shows the B-splines for another special case.

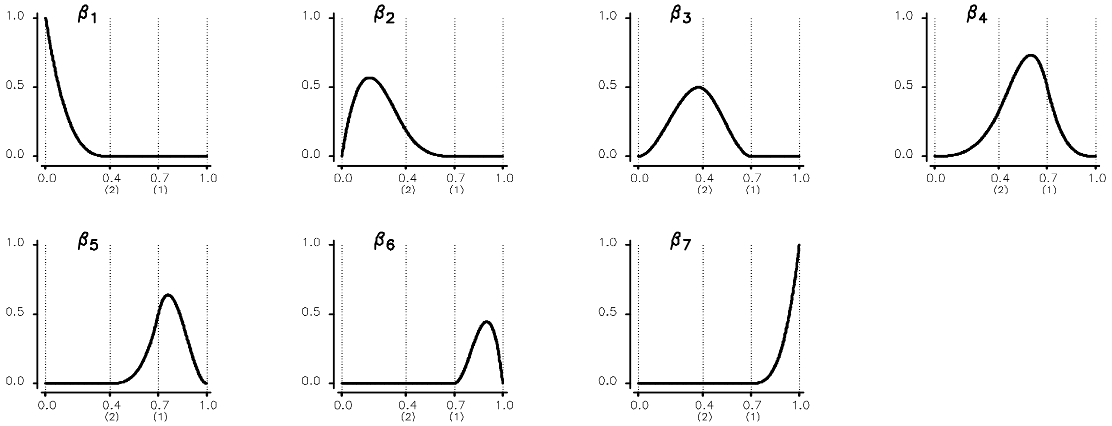

FIGURE 2.1. B-splines over $[a,b] = [0,1]$ for $d = 3$, $\ell = 3$, $\kappa_1 = .4$, $\kappa_2 = .7$, $s_1 = 2$ and $s_2 = 1$.

Regarding numerical aspects, a favorable consequence from (2.3) is that the associated information matrices $M(\xi)$ are $(2d+1)$-banded, i.e., the (i,j)-entries vanish whenever $|\,i - j\,| > d$. Moreover, that basis is totally positive, see, e.g., Mørken [8], ensuring total positivity of $M(\xi)$ for all designs ξ. That property is known to be very advantageous in many (numerical) problems of approximation and optimization.

Finally, the B-basis in (1.1) guarantees the desired equivariance: Let $L(x) = ux + v$, $x \in [a,b]$ $(u \neq 0)$, be an affine-linear mapping, transforming the regression interval $[a,b]$ (and its knots κ_i) into $[\widetilde{a}, \widetilde{b}]$ (with knots $\widetilde{\kappa}_i = L(\kappa_i)$, $1 \leq i \leq \ell - 1$); the smoothnesses are not affected. Let $\widetilde{\beta}$ be the B-basis of the associated spline space $\widetilde{\mathcal{S}}$, and denote, for a design η over $[\widetilde{a}, \widetilde{b}]$, by $\widetilde{M}(\eta)$ its information matrix from (1.2) with β replaced by $\widetilde{\beta}$.

THEOREM 1. *For all designs ξ over $[a,b]$ we have*

$$M(\xi) = \widetilde{M}(\xi^L),$$

where ξ^L is the image of ξ under L, given by $\xi^L\big(L(x)\big) = \xi(x)$, $x \in [a,b]$.

Proof. From the definition of β and $\widetilde{\beta}$ one straightforwardly gets

$$\widetilde{\beta}(x) = \beta\big(L^{-1}(x)\big) \qquad \text{for all } x \in [\widetilde{a}, \widetilde{b}],$$

and the assertion follows immediately. □

If $d = 1$, the information matrices are of a particular simple form, and many design problems can be solved explicitly.

THEOREM 2. *In linear spline regression $(d = 1)$ the uniform distribution on $\kappa_0, \dots, \kappa_\ell$ is uniformly optimal w.r.t. all orthogonally invariant optimality criteria.*

Proof. From (2.2) and (2.3) it follows $\beta_\mu(x) \leq 1 = \beta_\mu(\kappa_{\mu-1})$ for all μ and all x, and Kiefer's famous Proposition 1 proves the lemma. □

3. Nearly optimal designs

In Gaffke & Heiligers [5] we propose a (quasi-Newton type) algorithm for solving numerically φ_p-optimal spline design problems. The algorithm actually runs fast and accurately, finding (if $d \leq 8$ and $k \leq 30$) in reasonable computation times the designs with guaranteed efficiencies $\geq 1 - 10^{-7}$. However, it shares a common feature of second order methods, as its convergence behavior sensitively depends on the starting point.

Our proposal for choosing the starting point is motivated by the well known convergence of the φ_p-criteria: $\varphi_{-\infty}$ (the E-criterion) is the pointwise limit of these criteria when p tends to $-\infty$. So one might expect that the φ_p-optimal designs among those concentrated on the E-optimal support will serve as good starting points, at least for small values of p.

The E-optimal support is shown in Heiligers [6] to consist of the $k = \dim(\mathcal{S})$ *Chebyshev points,* i.e., the alternation points $a = \rho_1 < \cdots < \rho_k = b$ of the *Chebyshev spline* $\mathrm{C}^* \in \mathcal{S}$. That is, the unique (normalized) equi-oscillating spline in $\mathcal{S}$ satisfying

$$(-1)^\nu \, \mathrm{C}^*(\rho_\nu) = \max_{x\in[a,b]} |\mathrm{C}^*(x)| = 1 \qquad \text{for all } 1 \le \nu \le k\,. \tag{3.1}$$

We remark that the collocation matrix $\big(\beta_\mu(\rho_\nu)\big)_{1\le\mu,\nu\le k}$ in the Chebyshev points is nonsingular, as follows from Lemma 15 in Mørken [8] and the Schoenberg-Whitney Theorem. Thus all designs ξ with $\operatorname{supp}(\xi) = \{\rho_1, \ldots, \rho_k\}$ have nonsingular information matrices $M(\xi)$.

Of course, in most setups the Chebyshev points are not know explicitly, but have to be computed numerically. Regarding this problem, note that C^* is (up to normalization) the error function when approximating in the Chebyshev norm over $[a,b]$ the B-spline β_1 by $\beta_2, \ldots, \beta_k$. Hence, algorithms from approximation theory can be utilized for computation. Here we found in our numerical examples that an adopted version of Remez's second exchange algorithm, see, e.g., Cheney [3], is remarkably fast and stable, showing an excellent convergence behavior to the optimum. For convenience, we give a brief description of that procedure.

> At the lth cycle, k points $a = x_1^{(l)} < \cdots < x_k^{(l)} = b$ are given such that $\big(\beta_\nu(x_\mu^{(l)})\big)_{\nu,\mu}$ is nonsingular. (A good choice for the first cycle is $x_\mu^{(0)} = d^{-1} \sum\limits_{\nu=1}^{d} z_{\nu+\mu}$.) Then real numbers $q_2^{(l)}, \ldots, q_k^{(l)}$ and $m^{(l)}$ are obtained by solving the linear equations
>
> $$\sum_{\mu=2}^{k} q_\mu \, \beta_\mu\big(x_\nu^{(l)}\big) - (-1)^\nu \, m^{(l)} = \beta_1\big(x_\nu^{(l)}\big)\,, \qquad 1 \le \nu \le k\,. \tag{3.2}$$
>
> Define $Q^{(l)} = -\beta_1 + \sum\limits_{\mu=2}^{k} q_\mu^{(l)} \beta_\mu \in \mathcal{S}$. If $\max\limits_{x\in[a,b]} \big|Q^{(l)}(x)\big| = \big|m^{(l)}\big|$ then $\{x_1^{(l)}, \ldots, x_k^{(l)}\} = \{\rho_1, \ldots, \rho_k\}$. Otherwise, choose k points $a = x_1^{(l+1)} < \cdots < x_k^{(l+1)} = b$ such that
>
> $$\operatorname{sgn}\big(Q^{(l)}(x_\nu^{(l+1)})\big) = \operatorname{sgn}\big(Q^{(l)}(x_\nu^{(l)})\big) \quad \text{for all } \nu\,, \tag{3.3a}$$
> $$\big|Q^{(l)}(x_\nu^{(l+1)})\big| \ge \big|m^{(l)}\big| \quad \text{for all } \nu\,, \text{ and} \tag{3.3b}$$

$$\left|Q^{(l)}\left(x_{\nu_0}^{(l+1)}\right)\right| = \max_{x\in[a,b]} \left|Q^{(l)}(x)\right| \quad \text{for some } \nu_0\,, \tag{3.3c}$$

and the next cycle starts with $x_\nu^{(l)}$ replaced by $x_\nu^{(l+1)}$, $1 \le \nu \le k$.

Remark 3. Total positivity of β and nonsingularity of $\left(\beta_\nu(x_\mu^{(l)})\right)_{\nu,\mu}$ ensure

$$\det\begin{pmatrix} -(-1)^1 & \beta_2(x_1^{(l)}) & \cdots & \beta_k(x_1^{(l)}) \\ \vdots & \vdots & & \vdots \\ -(-1)^k & \beta_2(x_k^{(l)}) & \cdots & \beta_k(x_k^{(l)}) \end{pmatrix} \ge \det\begin{pmatrix} \beta_2(x_2^{(l)}) & \cdots & \beta_k(x_2^{(l)}) \\ \vdots & & \vdots \\ \beta_2(x_k^{(l)}) & \cdots & \beta_k(x_k^{(l)}) \end{pmatrix} > 0,$$

and equations (3.2) possess the unique solution. Moreover, $1 = -Q^{(l)}(a) = m^{(l)}$, thus, observing $Q^{(l)}(x_\nu^{(l)}) = (-1)^\nu$ for all ν, the spline $Q^{(l)}$ has $k-1$ sign changes. Consequently, by Lemma 15 in Mørken [8], its k local extrema $a = x_1^{(l+1)} < \cdots < x_k^{(l+1)} = b$ satisfy conditions (3.3a) – (3.3c); the same lemma, combined with the Schoenberg–Whitney Theorem, ensures that the corresponding collocation matrix is again nonsingular.

We emphasize that, although the procedure found the optimum in all our numerical tests, its theoretic convergence in the spline space is not guaranteed up to now.

The φ_p-optimal design on the Chebyshev points can be determined numerically utilizing roughly the same method as in Gaffke & Heiligers [5]: For the fixed-support problem, the search direction in each cycle is obtained by minimizing a quadratic approximation (being updated in the iterations) to the objective function at the current matrix over the polyhedron $\operatorname{conv}\left(\{M(\xi)\colon \operatorname{supp}(\xi) \subset \{\rho_1,\ldots,\rho_k\}\}\right)$. In all examples from Table 3.1a, below, this took not more than 25 iterations.

It actually turned out that the φ_p-optimal designs on the ρ_ν's are highly efficient, even for large values of p. Their efficiencies relative to the optimal designs usually exceed .98 for the D-criterion; higher values are obtained for A-optimality. In any case, the optimal Chebyshev support design is an excellent starting point for solving the original problem. In our examples, a design with guaranteed efficiency $\ge 1 - 10^{-7}$ was found after not more than 20 additional iterations.

TABLE 3.1A. A- and D-efficiencies of optimal B-spline maxima (**B**), and of Chebyshev points designs (**C**).

setup	1		2		3		4	
design	**B**	**C**	**B**	**C**	**B**	**C**	**B**	**C**
A-criterion	.9454	.9953	.9175	.9981	.9623	.9940	.5153	.9969
D-criterion	.9863	.9939	.9799	.9945	.9908	9955	.8639	.9850

TABLE 3.1B. Setups from Table 3.1a; the regression interval is $[a, b] = [0, 1]$.

setup	d	knot (smoothness)					
1	3	.400 (2)	.700 (1)				
2	3	.325 (2)	.525 (1)	.850 (2)			
3	3	.310 (2)	.425 (2)	.535 (2)	.680 (2)	.750 (1)	.850 (1)
4	5	.250 (0)	.600 (4)	.700 (2)			

Tables 3.1a/b give some numerical results on D- and A-optimality in different setups; setup 1 corresponds to Figure 2.1; setups 2 and 3 are from Kaishev [7], Table 1. For reference, Table 3.1a contains also the efficiencies of the optimal designs concentrated on the individual B-spline maxima, which are recommended — for D-optimality — by Kaishev [7]. A typical feature of these designs is that they, compared to the optimal ones, tend to assign too much mass to the interior of the regression interval. This effect becomes more significant for higher degrees d and larger numbers ℓ of knots. Figure 3.1 illustrates the competing designs for the setup from Figure 2.1 (and the A-criterion).

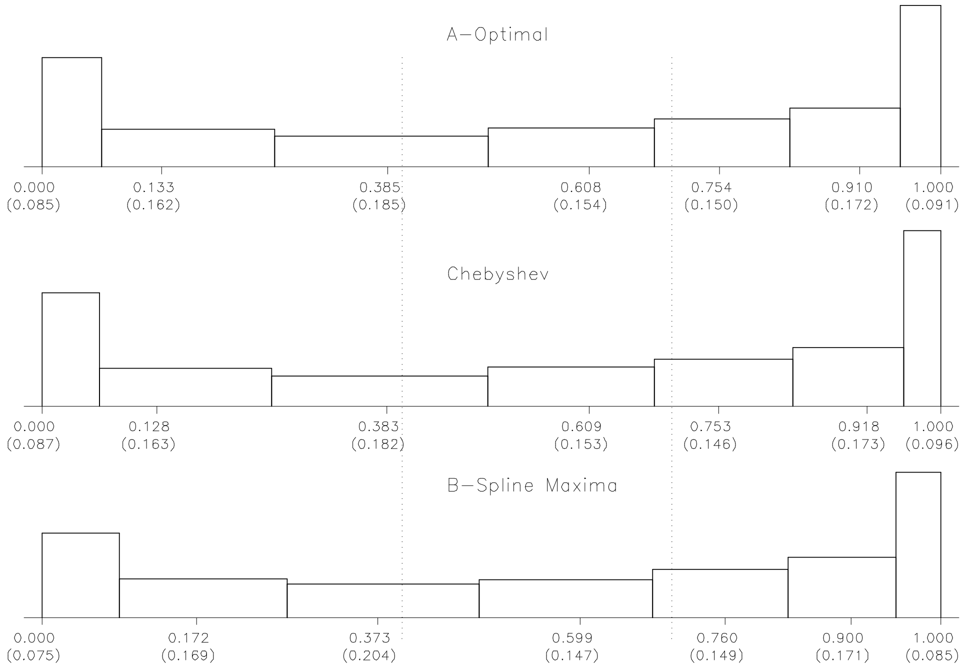

FIGURE 3.1. Histogram representations of the A-optimal, the A-optimal on Chebyshev points, and the A-optimal design on B-spline maxima for the setup from Figure 2.1; the first rows give the support points, the numbers in parenthesis the weights. The dotted lines indicate the location of knots.

Due to the excellent efficiencies of the optimal designs on the Chebyshev points, one might expect that they also provide the support of good designs when interest is in mixture criteria, aiming at minimizing the average deficiency of a design w.r.t. several individual φ_p-criteria. An optimal design under a mixture may be expected to be fairly efficient w.r.t. all involved criteria, and may thus be viewed to be "robust" against particular choices of criteria. In fact, first numerical tests show that again excellent efficiencies are achieved among the designs supported by the Chebyshev points.

REFERENCES

[1] CARNICER, J. M.—PEÑA, J. M.: *Totally positive bases for shape preserving curve design and optimality of B-splines*, Comput. Aided Geom. Design **11** (1994), 633–654.

[2] CARNICER, J. M.—PEÑA, J. M.: *Least supported bases and local linear independence*, Numer. Math. **67** (1994), 289–301.

[3] CHENEY, E. W.: *Introduction to Approximation Theory.* (*2nd. edition*), Chelsea Publ. Company, New York, 1982.

[4] DETTE, H.: *Designing experiments with respect to 'standardized' optimality criteria*, J. Roy Statist. Soc. Ser. **B 59** (1997), 97–110.

[5] GAFFKE, N.—HEILIGERS, B.: *Optimal approximate designs for B-spline regression with multiple knots*, in: Statistical Process Monitoring and Optimization (S .H. Park and G. G. Vining, eds.), Marcel Dekker, 1999 (to appear).

[6] HEILIGERS, B.: *E-optimal designs for polynomial spline regression*, J. Statist. Plann. Inference **75** (1998), 159–172.

[7] KAISHEV, V. K.: *Optimal experimental design for B-spline regression*, Comput. Statist. Data Anal. **8**, 39–47.

[8] MØRKEN, K.: *Total positivity and splines*, in: Total Positivity and its Applications (M. Gasca and C. M. Micchelli, eds.), Kluwer Academic Publ., Dordrecht, 1996, pp. 47–84.

[9] PUKELSHEIM, F.: *Optimal Design of Experiments*, J. Wiley, New York, 1993.

Received May 15, 1998

Fakultät für Mathematik
Institut für Mathematische Stochastik
Universität Magdeburg
D–39106 Magdeburg
GERMANY

E-mail: berthold.heiligers@mathematik.uni-magdeburg.de