Ann. Inst. Statist. Math.
Vol. 52, No. 3, 557-573 (2000)

CONVEX OPTIMAL DESIGNS FOR COMPOUND POLYNOMIAL EXTRAPOLATION

HOLGER DETTE[1] AND MONG-NA LO HUANG[2]

[1] *Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany*
[2] *Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, R.O.C.*

(Received May 18, 1998; revised May 13, 1999)

Abstract. The extrapolation design problem for polynomial regression model on the design space $[-1, 1]$ is considered when the degree of the underlying polynomial model is with uncertainty. We investigate compound optimal extrapolation designs with two specific polynomial models, that is those with degrees $\{m, 2m\}$. We prove that to extrapolate at a point z, $|z| > 1$, the optimal convex combination of the two optimal extrapolation designs $\{\xi_m^*(z), \xi_{2m}^*(z)\}$ for each model separately is a compound optimal extrapolation design to extrapolate at z. The results are applied to find the compound optimal discriminating designs for the two polynomial models with degrees $\{m, 2m\}$, i.e., discriminating models by estimating the highest coefficient in each model. Finally, the relations between the compound optimal extrapolation design problem and certain nonlinear extremal problems for polynomials are worked out. It is shown that the solution of the compound optimal extrapolation design problem can be obtained by maximizing a (weighted) sum of two squared polynomials with degree m and $2m$ evaluated at the point z, $|z| > 1$, subject to the restriction that the sup-norm of the sum of squared polynomials is bounded.

Key words and phrases: Chebyshev polynomials, convex combination, extremal problems for polynomials, Lagrange interpolation polynomial, optimal discrimination designs.

1. Introduction

When analyzing data obtained from scientific experiments, regression models have been used extensively to illustrate the relationships that may exist between variables. Among the numerous regression models, polynomial regression may be one of the most frequently used. However in many cases the specification of the exact degree of the polynomial model may still be a problem and need to be taken care of in the stage of designing the experiments.

The literature about designing experiments with concerns about uncertainty of the models dates back to Box and Draper (1959), where it was demonstrated that large bias may be introduced in estimation when the model is assumed to be linear and the true regression function may be quadratic. Many papers thereafter have addressed this issue and provided different design strategies, for some examples see Stigler (1971), Atkinson and Cox (1974), Läuter (1974*a*, 1974*b*), Huber (1975), Studden (1982), Sacks and Ylvisaker (1984), Huang and Studden (1988), Dette (1991, 1994, 1995*a*), Pukelsheim and Rosenberger (1993), and Dette and Studden (1995). On the other hand, there has also been appreciable previous work on optimal designs for extrapolation in a regression, most of them are for the one dimensional case. For example, Hoel and Levine (1964) provided optimal extrapolation design solution for the polynomial regression models. Kiefer and Wolfowitz (1965), Karlin and Studden (1966), Studden (1968) considered the models when the elements of the vector of regression functions form a Chebyshev system. Hoel (1965) and Studden (1971) investigated the optimal extrapolation designs with multi-dimensional design space. Huber (1975) dealt with the extrapolation design problem when the true regression function is in a class of possible candidates, such as regression functions with certain bounded derivatives, Kiefer (1980) discussed the extrapolation problems in the one dimensional case as well as in a ball of dimension q, $q > 1$, when model uncertainty induced bias in the estimation. Spruill (1984) investigated minimax extrapolation designs for a special class of regression functions. Chao (1995) studied the extrapolation design problems with discrimination of models in mind. Dette and Wong (1996) proposed a new class of model robust optimality criteria for extrapolation, which is motivated in part by Kiefer's L_p-class of optimality criteria.

In this work, the model robust extrapolation design problems proposed by the last-named authors are investigated in more details. It is usually not easy to derive the optimal designs analytically for general compound optimality criteria, and even numerical solutions are difficult to find for higher degree models. A natural thought is that for a compound type of criterion, a convex combination of the optimal designs for the individual criterion may be optimal. This is usually not true in the general case. We study this problem for those models where the set of support points of the optimal extrapolation designs for one model is a subset of the others. More precisely, we investigate compound optimal extrapolation designs for two specific polynomial models with degrees $\{m, 2m\}$. We prove in Section 3 that to extrapolate at a point z, $|z| > 1$, the optimal convex combination of the two optimal extrapolation designs $\{\xi_m^*(z), \xi_{2m}^*(z)\}$ for each model separately is a compound optimal extrapolation design to extrapolate at z. Then the asymptotic distribution (when the degrees of the two polynomials tend to infinity) is derived. The results are applied to find the compound optimal discriminating design for the two polynomial models with degrees $\{m, 2m\}$, i.e. discriminating models through estimation of the highest coefficient in each model. In Section 4 properties of the compound optimal extrapolation design are investigated in more details and it is shown that there is a dual problem in approximation theory. More precisely we prove that the compound optimal design problem is dual to the problem of maximizing a (weighted) sum of squared polynomials of degree m and $2m$ evaluated at the point z, $|z| > 1$, subject to the restriction that the sup-norm of the sum of the squared polynomials is bounded. Moreover the solution of one problem can be obtained from the other and vice versa. Consequently the results from design theory can be used to solve a nonlinear generalization of a classical extremal problem for polynomials. Finally all proofs are given in Section 5.

2. Hoel and Levine designs and robust optimality criteria

Assume that for each $x \in [-1, 1]$ an experiment can be performed, and the outcome is a random variable $y(x)$, with mean value $\Theta_k^T f_k(x)$, where $f_k(x) = (1, x, x^2, \ldots, x^k)^T$, and a common variance σ^2. The vector of parameters $\Theta_k^T = (\theta_0, \theta_1, \ldots, \theta_k)$ and the variance σ^2 are unknown. Suppose that n uncorrelated observations of the response $y(x)$ are to be obtained at levels $x_1, x_2, \ldots, x_n$. An exact design specifies a probability measure ξ on $[-1, 1]$ which concentrates mass q_i at distinct x_i, $i = 1, \ldots, r$, where $q_i n = n_i$ are integers. An approximate design takes away the restriction that $n_1, \ldots, n_r$ are integers. The moment matrix of a design ξ for a polynomial model of degree k is defined by

$$M_k(\xi) = \int_{-1}^{1} f_k(x) f_k^T(x) d\xi(x) = (c_{i+j})_{i,j=0}^{k}, \tag{2.1}$$

where $c_j = \int_{-1}^{1} x^j d\xi(x)$ denotes the ordinary j-th moment of ξ, $(j = 1, 2, \ldots)$. An optimal design maximizes or minimizes an appropriate function of the moment matrix. Since in the following our main interests are estimation of the "highest coefficient" θ_k or extrapolation at a point outside the design space, we will only consider designs ξ with a nonsingular moment matrix $M_k(\xi)$.

The optimal design problem for extrapolation at a point z, $|z| > 1$ in a single polynomial model with degree k is to seek an approximate design which minimizes the variance of the estimate at point z, which is proportional to

$$v_k(\xi, z) = f_k^T(z) M_k^{-1}(\xi) f_k(z). \tag{2.2}$$

Hoel and Levine (1964) proved that in a polynomial regression the optimal extrapolation design $\xi_k^* = \xi_k^*(z)$ is supported at the Chebyshev points, i.e. $\{s_{k,\nu} = \cos((k-\nu)\pi/k)\}_{\nu=0}^{k}$, which are the points in the interval $[-1, 1]$ such that $|T_k(x)| = 1$, where $T_k(x)$ is the k-th Chebyshev polynomial of the first kind, i.e. $T_k(x) = \cos(k \arccos x) = 2^{k-1} \prod_{j=1}^{k} (x - s_{2k,2j-1})$. The corresponding weights are given by $q_{k,\nu}$ with

$$q_{k,\nu} = \frac{|\ell_{k,\nu}(z)|}{\sum_{\nu=0}^{k} |\ell_{k,\nu}(z)|} = \frac{|\ell_{k,\nu}(z)|}{|T_k(z)|}, \qquad 0 \le \nu \le k, \tag{2.3}$$

where $|z| > 1$ and $\ell_{k,\nu}(x)$ denotes the ν-th Lagrange interpolation polynomial at the $k+1$ Chebyshev points $\{s_{k,\nu}\}_{\nu=0}^{k}$. It is also given there that

$$v_k(\xi_k^*(z), z) = \sum_{\nu=0}^{k} (\ell_{k,\nu}^2(z)/q_{k,\nu}) = \left(\sum_{\nu=0}^{k} |\ell_{k,\nu}(z)| \right)^2 = T_k^2(z).$$

In the following we will call $\xi_k^*(z)$ the k-th Hoel and Levine extrapolation design at point z (the k-th Hoel and Levine design in short) for convenience.

The robust optimality criterion proposed in Dette and Wong (1996) for extrapolation, when there is uncertainty in the degree of the polynomial, is defined as finding designs minimizing the following information function

$$\Pi_p(\xi, z) = \begin{cases} \left[\sum_{j=1}^{k} p_j \left\{ \dfrac{v_j(\xi, z)}{v_j(\xi_j^*(z), z)} \right\}^{-p} \right]^{-1/p} & (-\infty < p \le 1, p \ne 0), \\ \prod_{j=1}^{k} \left\{ \dfrac{v_j(\xi, z)}{v_j(\xi_j^*(z), z)} \right\}^{p_j} & (p = 0), \\ \max \left\{ \dfrac{v_j(\xi, z)}{v_j(\xi_j^*(z), z)} \,\middle|\, p_j > 0, j = 1, \ldots, k \right\} & (p = -\infty), \end{cases} \tag{2.4}$$

where $v_j(\xi, z)$ is as defined in (2.2), $p_j \ge 0$ for all j and $\sum_{j=1}^{k} p_j = 1$. The vector $(p_1, \ldots, p_k)$ is called prior for the class of polynomials of degree k and a robust optimal extrapolation design minimizes (2.4) in the class of all approximate designs on the interval $[-1,1]$. Note that $\Pi_p(\xi,z)$ is an information function of the efficiencies of the design ξ for extrapolation at the point z in the polynomial models of degree $1,\ldots,k$. The corresponding equivalence theorem for characterizing the optimal extrapolation designs with respect to the above criteria can be found in Pukelsheim ((1993), p. 290) and Dette and Wong (1996). To be more precise, define for a design ξ, $d_j(\xi,z,x) = f_j^T(z)M_j^{-1}(\xi)f_j(x)$. Let $\xi_j^*(z)$ be the j-th Hoel and Levine design and let $(p_1,\ldots,p_k)$ be a set of priors. When $-\infty < p \le 1$, then ξ^* minimizes $\Pi_p(\xi,z)$ if and only if for all x in $[-1,1]$,

$$\sum_{j=1}^{k} p_j \left\{ \frac{d_j^2(\xi^*,z,x)}{v_j(\xi^*,z)} - 1 \right\} \left\{ \frac{v_j(\xi^*,z)}{v_j(\xi_j^*(z),z)} \right\}^{-p} \le 0. \tag{2.5}$$

Similarly, for $p=-\infty$, define

$$A(\xi^*) = \left\{ i = 1,2,\ldots,k \mid p_i > 0, \frac{v_i(\xi^*,z)}{v_i(\xi_i^*(z),z)} = \max_{1\le j\le k} \frac{v_j(\xi^*,z)}{v_j(\xi_j^*(z),z)} \right\}$$

as the set of all indices where the maximum of the efficiencies is attained, then ξ^* minimizes $\Pi_{-\infty}(\xi,z)$ if and only if there exist nonnegative constants β_j such that $\sum_{j\in A(\xi^*)} \beta_j = 1$ and the inequality

$$\sum_{j\in A(\xi^*)} \beta_j \left\{ \frac{d_j^2(\xi^*,z,x)}{v_j(\xi^*,z)} - 1 \right\} \le 0, \tag{2.6}$$

holds for all $x \in [-1,1]$.

3. Optimality of convex Hoel and Levine designs for compound extrapolation

In this section, we will investigate the optimal extrapolation designs for two specific polynomial models, that is designs optimal in minimizing the information function $\Pi_{\lambda,p}(\xi,z)$ with $k=2m$, $m\ge 1$, and only p_m, p_{2m} are positive. In order to simplify the notation we put $p_m=\lambda$, $p_{2m}=1-\lambda$, $0<\lambda<1$ and obtain from (2.4)

$$\Pi_{\lambda,p}(\xi,z) = \begin{cases} \left[\lambda \left\{ \frac{v_m(\xi,z)}{v_m(\xi_m^*(z),z)} \right\}^{-p} + (1-\lambda) \left\{ \frac{v_{2m}(\xi,z)}{v_{2m}(\xi_{2m}^*(z),z)} \right\}^{-p} \right]^{-1/p} & (-\infty < p \le 1, p \ne 0), \\ \left\{ \frac{v_m(\xi,z)}{v_m(\xi_m^*(z),z)} \right\}^{\lambda} \left\{ \frac{v_{2m}(\xi,z)}{v_{2m}(\xi_{2m}^*(z),z)} \right\}^{1-\lambda} & (p=0), \\ \max \left\{ \frac{v_m(\xi,z)}{v_m(\xi_m^*(z),z)}, \frac{v_{2m}(\xi,z)}{v_{2m}(\xi_{2m}^*(z),z)} \right\} & (p=-\infty). \end{cases}$$

The design minimizing $\Pi_{\lambda,p}(\xi,z)$ will be called compound optimal extrapolation designs from now on. Note that as mentioned above $v_m(\xi_k^*(z),z) = T_m^2(z)$, and from the definition of $T_m(x)=\cos(m\arccos x)$ as well as the trigonometric identity, we know $T_{2m}(z) = 2T_m^2(z)-1$, and $v_{2m}(\xi_{2m}^*(z),z) = T_{2m}^2(z) = (2T_m^2(z)-1)^2$. It will be shown that an optimal convex combination of the two optimal extrapolation designs $\{\xi_m^*(z), \xi_{2m}^*(z)\}$ for the individual polynomial model of degree m and $2m$ is a compound optimal extrapolation design. The optimal convex combination constant $\alpha_{\lambda,p}^*$, $0 < \alpha_{\lambda,p}^* < 1$, can be obtained through solving a nonlinear equation. This result will be stated in Theorem 3.1. For the case $p = -\infty$, $\Pi_{\lambda,-\infty}$ is a nondifferentiable optimality criterion and it is usually harder to find the optimal designs. However using results from the proof in Theorem 3.1 we are able to prove a corresponding result in Theorem 3.2. Then from the proof of Theorem 3.2 and by verifying it with the equivalence theorem presented in Section 2, it is given in Corollary 3.1 that for a special value of λ, there is a design which is compound optimal extrapolating at a point z for all p, $-\infty < p \leq 1$. In Corollary 3.2 the asymptotic distribution (when the degrees of the two polynomials tend to infinity) is derived.

THEOREM 3.1. *For a given λ, $0 < \lambda < 1$, and $-\infty < p \leq 1$, the design*

$$\xi_{m,\alpha_{\lambda,p}^*}(z) = \alpha_{\lambda,p}^*\xi_m^*(z) + (1 - \alpha_{\lambda,p}^*)\xi_{2m}^*(z),$$

with $\alpha_{\lambda,p}^*$ being the unique solution in $(0, 1)$ of the equation

$$\frac{\lambda}{1-\lambda} + \frac{\alpha(1-\alpha)^{p-1}T_{2m}^p(z)(\alpha - 2T_m^2(z))}{(T_m^2(z) - \alpha)^{p+1}} = 0, \tag{3.1}$$

is a compound optimal extrapolation design.

THEOREM 3.2. *For $p = -\infty$, the design $\xi_{m,1/2}(z) = (1/2)\xi_m^*(z) + (1/2)\xi_{2m}^*(z)$ is a compound optimal extrapolation design.*

COROLLARY 3.1. *For* $\lambda = ((1/2)T_{2m}(z) + T_m^2(z))/(T_{2m}(z) + T_m^2(z))$ and all $-\infty < p \leq 1$, *the design $\xi_{m,1/2}(z) = (1/2)\xi_m^*(z) + (1/2)\xi_{2m}^*(z)$ is a compound optimal extrapolation design.*

COROLLARY 3.2. *The compound optimal extrapolation design $\xi_{m,\alpha_{\lambda,p}^*}(z)$ defined in Theorem 3.1 $(-\infty < p \leq 1)$ and 3.2 $(p = -\infty)$ converges (as $m \to \infty$) weakly to the design $\xi^*(z)$ on $(-1, 1)$ with density*

$$\frac{\sqrt{z^2-1}}{\pi|z-x|\sqrt{1-x^2}}. \tag{3.2}$$

Corollary 3.2 gives us some idea how the compound optimal extrapolation design behaves as the degrees $\{m, 2m\}$ of the polynomials become large. It is interesting to see that as $m \to \infty$, there is no difference on the asymptotic distribution for all p, $-\infty \leq p \leq 1$.

3.1. *Optimality of convex limiting Hoel and Levine designs on discrimination for polynomial models with degrees $\{m, 2m\}$*

In this subsection, we discuss the problem of designing experiments efficient for discriminating polynomial models, which is closely related to the problem of finding compound optimal extrapolation design discussed previously.

First recall that the design problem for the estimation of $\theta_k = e_k^T\Theta_k$, where $e_k = (0, \ldots, 0, 1)^T \in R^{k+1}$ denotes the $(k+1)$-th unit vector, is to minimize $e_k^T M_k^{-1}(\xi)e_k$ or to maximize

$$C_{e_k} = (e_k^T M_k^{-1}(\xi)e_k)^{-1} = \frac{|M_k(\xi)|}{|M_{k-1}(\xi)|}.$$

A design maximizing $|M_k(\xi)|/|M_{k-1}(\xi)|$ is called D_1-optimal and denoted by $\xi_k^{D_1}$. It is known (see Kiefer and Wolfowitz (1959)) that $\xi_k^{D_1}$ is also supported at the $k+1$ Chebyshev points $\{s_{k,\nu}\}_{\nu=0}^k$. The corresponding weights are $1/k$ at the interior points $\{s_{k,\nu}\}_{\nu=1}^{k-1}$ and $1/2k$ at the two end points -1 and 1. The parameter θ_k has often been used to discriminate polynomial models of degree k or $k-1$.

In order to discriminate between polynomials of different degree, Dette and Studden ((1997), p. 180) define a p-mean of D_1-efficiencies which for discriminating between polynomials of degree $(m-1, m)$ or $(2m-1, 2m)$ is reduced to

$$\left\{\lambda\left(2^{2m-2}\frac{|M_m(\xi)|}{|M_{m-1}(\xi)|}\right)^p+(1-\lambda)\left(2^{4m-2}\frac{|M_{2m}(\xi)|}{|M_{2m-1}(\xi)|}\right)^p\right\}^{-1/p} \tag{3.3}$$

$(-\infty \le p \le 1)$ where the cases $p=0, -\infty$ are interpreted as the corresponding limits (see (2.4)). The following two corollaries give the compound optimal discriminating design minimizing (3.3).

Corollary 3.3. *For a given λ, $0<\lambda<1$, and $-\infty<p\le 1$, the design $\xi^{D_1}_{m,\alpha^*_{\lambda,p}}$, where*

$$\xi^{D_1}_{m,\alpha^*_{\lambda,p}}=\alpha^*_{\lambda,p}\xi_m^{D_1}+(1-\alpha^*_{\lambda,p})\xi_{2m}^{D_1},$$

with $\alpha^*_{\lambda,p}$ being the unique solution in $(0,1)$ of the equation

$$\frac{\lambda}{1-\lambda}-\alpha(1-\alpha)^{p-1}2^{p+1}=0$$

is a compound optimal discriminating design.

Corollary 3.4. *For $p=-\infty$, the design $\xi^{D_1}_{m,1/2}=(1/2)\xi_m^{D_1}+(1/2)\xi_{2m}^{D_1}$ is a compound optimal discriminating design.*

Proofs of Theorem 3.1, 3.2 and Corollary 3.2 will be delayed to the Appendix, then from the proof of Theorem 3.2, Corollary 3.1 can be proved easily, and with the first two theorems, Corollary 3.3 and 3.4 can be obtained by observing that

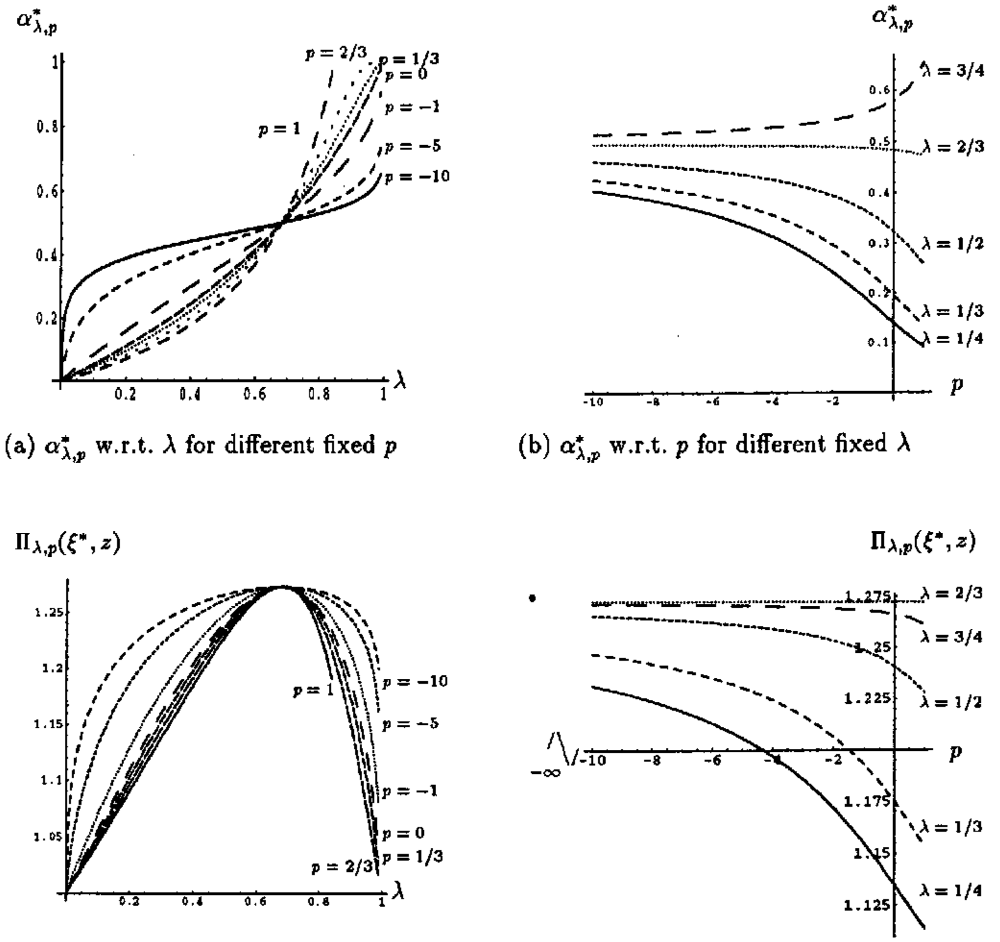
$$\lim_{|z|\to\infty}\frac{1}{z^{2j}}v_j(\xi,z)=e_j^TM_j^{-1}(\xi)e_j=\frac{|M_{j-1}(\xi)|}{|M_j(\xi)|}.$$

Consequently these results follow by considering the limits $z\to\infty$ in Theorems 3.1 and 3.2.

3.2. *Graphs of optimal weights and optimal information functions with respect to the prior λ and criterion parameter p*

Now we present some graphs of the optimal weights $\alpha^*_{\lambda,p}$ and optimal information functions $\Pi_{\lambda,p}(\xi_{m,\alpha^*_{\lambda,p}}(z),z)$ with models of degree 1 or 2 (i.e. $m=1$), for compound extrapolation at point $z=2$ to illustrate their behaviors with respect to the prior λ and criterion parameter p. These are given in Fig. 1.

From the graphs of the optimal weights and optimal information functions, for fixed value of p, both $\alpha^*_{\lambda,p}$ $\Pi_{\lambda,p}(\xi_{m,\alpha^*_{\lambda,p}}(z),z)$ become larger when λ gets larger and when $p\to-\infty$, $\alpha^*_{\lambda,p}\to 1/2$ as expected. But for fixed value of λ, it is interesting to note that there is a turning point that for $\lambda<((1/2)T_{2m}(z)+T_m^2(z))/(T_{2m}(z)+T_m^2(z))$, $\alpha^*_{\lambda,p}$



(a) $\alpha^*_{\lambda,p}$ w.r.t. λ for different fixed p

(b) $\alpha^*_{\lambda,p}$ w.r.t. p for different fixed λ

(c) $\Pi_{\lambda,p}(\xi^*, z)$ w.r.t. λ for different fixed p

(d) $\Pi_{\lambda,p}(\xi^*, z)$ w.r.t. p for different fixed λ

Fig. 1. Optimal weights $\alpha^*_{\lambda,p}$ and optimal information functions $\Pi_{\lambda,p}(\xi^*, z)$ for $\xi^* = \xi_{1,\alpha^*_{\lambda,p}}(z) = \alpha^*_{\lambda,p}\xi_1^*(z) + (1-\alpha^*_{\lambda,p})\xi_2^*(z)$ at $z = 2$ with models of degree 1 or 2.

is decreasing with respect to p, and for $\lambda > ((1/2)T_{2m}(z) + T_m^2(z))/(T_{2m}(z) + T_m^2(z))$, $\alpha^*_{\lambda,p}$ is increasing with respect to p. This phenomenon is also reflected in Corollary 3.1. Note also that the phenomenon is somewhat different from that of the optimal weights and optimal information functions with respect to p as obtained in Preitschopf and Pukelsheim (1987), where the optimal designs for estimating subsets of components of the parameter vector in a quadratic regression model were discussed.

4. Stieltjes transforms and extremal polynomials

In this section, we briefly indicate some extensions and applications of the results in Section 3. Our first result gives the Stieltjes transform

$$\Phi(x, \xi) = \int_{-1}^{1} \frac{d\xi(t)}{x - t}$$

of the convex combination $\xi_{m,\alpha}(z) = \alpha\xi_m^*(z) + (1-\alpha)\xi_{2m}^*(z)$. It has been pointed out in Lau (1983) and Dette and Wong (1996) that the Stieltjes transform is useful for obtaining the canonical moments and orthogonal polynomials with respect to the measure $\xi_{m,\alpha}(z)$. A further application is given in the proof of Corollary 3.2 below.

THEOREM 4.1. *The Stieltjes transform of the design* $\xi_{m,\alpha}(z) = \alpha\xi_m^*(z) + (1-\alpha)\xi_{2m}^*(z)$ *is given by*

$$\Phi(x, \xi_{m,\alpha}(z)) = \frac{(z^2-1)U_{2m-1}(z)}{T_{2m}(z)(x-z)} \cdot \left\{ \frac{T_{2m}(z)}{(z^2-1)U_{2m-1}(z)} - \frac{(1-\alpha)T_{2m}(x)T_m^2(z) + \alpha T_{2m}(z)T_m^2(x)}{T_m^2(z)U_{2m-1}(x)(x^2-1)} \right\}$$

where $U_j(x) = \sin((j+1)\arccos x)/\sin(\arccos x)$ *denotes the* j*-th Chebyshev polynomial of the second kind.*

Our second application refers to a generalization of a well known extremal property of the Chebyshev polynomials of the first kind (see Rivlin (1990), p. 93). To be precise let $P_m(x)$ and $P_{2m}(x)$ denote polynomial of degree m and $2m$ respectively and consider the extremal problem.

$$(P) \quad \max\left\{ \lambda P_m^2(z) + (1-\lambda)P_{2m}^2(z) \,\middle|\, \max_{x\in[-1,1]}(P_m^2(x) + P_{2m}^2(x)) \le 1 \right\}$$

where $\lambda \in [0,1]$ is a given constant and the maximum is taken over all possible polynomials $P_m(x)$ and $P_{2m}(x)$. The case $\lambda \in \{0,1\}$ is well known and yields the Chebyshev polynomial of the first kind as extremal polynomial. The following result gives a solution of the general problem. Its proof is based on a duality between the extremal problem (P) and a generalization of the minimax optimal extrapolation design criterion considered in Theorem 3.2.

THEOREM 4.2. *Let* $\lambda \in [0,1)$ *and* $\beta = \beta(\lambda) = \lambda/(1-\lambda)$.

(a) *If* $\beta(\lambda) \le T_{2m}(z)$, *then the solution of the extremal problem* (P) *is obtained by the polynomials* $P_m(x) \equiv 0$ *and* $P_{2m}(x) = \mp T_{2m}(x)$.

(b) *If* $\beta(\lambda) > T_{2m}(z)$, *then the solution of the extremal problem* (P) *is obtained by the polynomials*

$$P_m(x) = \mp\frac{\sqrt{(\beta-1)(\beta-T_{2m}(z))}}{\beta - T_m^2(z)} T_m(x),$$

$$P_{2m}(x) = \mp\left(\frac{(\beta-1)T_m^2(x)}{\beta - T_m^2(z)} - 1\right).$$

Note that for a uniform weight, $\lambda = \frac{1}{2}$, the optimal solution of (P) is given by the polynomials $P_m(x) \equiv 0$, $P_{2m}(x) \equiv T_{2m}(x)$. A more interesting situation occurs using weights proportional to $1/T_m^2(z)$ and $1/T_{2m}^2(z)$, which corresponds to some kind of standardization by the best values obtainable by the individual polynomials. In this case the condition $\beta(\lambda) \le T_{2m}(z)$ in Theorem 4.2 is equivalent to $T_m^2(z) \le 1$, and part (b) of the theorem applies. The dual problem of the extremal problem is the minimax optimal extrapolation design problem considered in Theorem 3.2 . This means that the optimum in (P) is always attained by two polynomials, namely

$$P_m(x) = \mp\frac{\sqrt{(4T_m^2(z)-1)(2T_m^2(z)-1)}}{3T_m^2(z)-1} T_m(x),$$

$$P_{2m}(x) = \mp\left(\frac{4T_m^2(z)-1}{3T_m^2(z)-1} T_m^2(x) - 1\right).$$

Acknowledgements

This work was done during a visit of the second author at the Department of Mathematics, Ruhr-Universität Bochum, and the second author would like to thank the department for the hospitality. The authors would also like to thank the DAAD of Federal Republic of Germany and the NSC of Taiwan, the Republic of China for financial support which made this visit possible.

Appendix: Proofs

Throughout this Appendix, we assume $z > 1$ (and as a consequence $T_k(z) > 0$, $k = m, 2m$); the remaining case is treated exactly in the same way.

PROOF OF THEOREM 3.1. For $0 \le \alpha < 1$, let $\xi_{m,\alpha}(z) = \alpha\xi^*_m(z) + (1-\alpha)\xi^*_{2m}(z)$ be a convex combination of the optimal extrapolation designs $\xi^*_m(z)$ and $\xi^*_{2m}(z)$ in the polynomial of degree m and $2m$, respectively. If $\{s_{k,\nu}\}_{\nu=0}^{k}$ denote the Chebyshev points, then

$$s_{m,\nu} = \cos((m-\nu)\pi/m) = s_{2m,2\nu}, \tag{A.1}$$

for $\nu = 0, 1, \ldots, m$, and by (2.2) and (2.3) $\xi_{m,\alpha}(z)$ is supported on $\{s_{2m,\nu}\}_{\nu=0}^{2m}$ with corresponding weights $\{w_\nu\}_{\nu=0}^{2m}$, given by

$$w_{2\nu} = \alpha\frac{|\ell_{m,\nu}(z)|}{|T_m(z)|} + (1-\alpha)\frac{|\ell_{2m,2\nu}(z)|}{|T_{2m}(z)|}, \qquad \nu = 0, 1, \ldots, m, \tag{A.2}$$

$$w_{2\nu-1} = (1-\alpha)\frac{|\ell_{2m,2\nu-1}(z)|}{|T_{2m}(z)|}, \qquad \nu = 1, 2, \ldots, m. \tag{A.3}$$

Note that for $\nu = 0, 1, \ldots, m$, we have

$$\begin{aligned}\ell_{2m,2\nu}(z) &= \prod_{j=1}^{m}\frac{(z - s_{2m,2j-1})}{(s_{2m,2\nu} - s_{2m,2j-1})}\prod_{\substack{j=0\\ j\neq\nu}}^{m}\frac{(z - s_{2m,2j})}{(s_{2m,2\nu} - s_{2m,2j})} \\ &= \frac{T_m(z)}{T_m(s_{2m,2\nu})}\ell_{m,\nu}(z) = (-1)^{m-\nu}T_m(z)\cdot\ell_{m,\nu}(z),\end{aligned} \tag{A.4}$$

where the second equality holds by the fact that the points $\{s_{2m,2j-1}\}_{j=1}^{m}$ are the m zeros for the m-th Chebyshev polynomial of the first kind $T_m(x) = \cos(m \arccos x)$, i.e. $T_m(x) = 2^{m-1}\prod_{j=1}^{m}(x - s_{2m,2j-1})$, and by (A.1). The last equality holds since $T_m(s_{2m,2\nu}) = (-1)^{m-\nu}, \nu = 0, 1, \ldots, m$. As a simple consequence we obtain the following Lemma which gives a simpler expression for the weights of the design $\xi_{m,\alpha}(z)$.

LEMMA A.1. *For $0 \le \alpha < 1$, the design $\xi_{m,\alpha}(z) = \alpha\xi^*_m(z) + (1-\alpha)\xi^*_{2m}(z)$ is supported at the Chebyshev points $\{s_{2m,\nu}\}_{\nu=0}^{2m}$, with corresponding weights*

$$w_{2\nu} = \frac{u|T_{2m}(z)|}{T_m^2(z)}\cdot\frac{|\ell_{2m,2\nu}(z)|}{|T_{2m}(z)|} = u\frac{|\ell_{m,\nu}(z)|}{|T_m(z)|}, \qquad \nu = 0, 1, \ldots, m, \quad \textit{and} \tag{A.5}$$

$$w_{2\nu-1} = (1-\alpha)\frac{|\ell_{2m,2\nu-1}(z)|}{|T_{2m}(z)|}, \qquad \nu = 1, 2, \ldots, m, \tag{A.6}$$

where

$$u = \sum_{\nu=0}^{m} w_{2\nu} = \alpha + (1-\alpha)\frac{T_m^2(z)}{|T_{2m}(z)|}.$$

In a second step we will evaluate the covariance functions $d_k(\xi_{m,\alpha}(z), z, x)$ for $k = m$, $2m$, which are needed in the equivalence theorems (2.5) and (2.6).

LEMMA A.2. *For u as defined in Lemma A.1 we have for every x and z with $|z| > 1$*

$$d_m(\xi_{m,\alpha}(z), z, x) = f_m^T(z) M_m^{-1}(\xi_{m,\alpha}(z)) f_m(x) = \frac{T_m(z)T_m(x)}{u}, \tag{A.7}$$

$$\begin{aligned} d_{2m}(\xi_{m,\alpha}(z), z, x) &= f_{2m}^T(z) M_{2m}^{-1}(\xi_{m,\alpha}(z)) f_{2m}(x) \\ &= \frac{T_m^2(z)T_m^2(x)}{u} + \frac{T_{2m}(z)(T_m^2(x)-1)}{(1-\alpha)}. \end{aligned} \tag{A.8}$$

PROOF. Note that from Lemma A.1, $M_m(\xi_{m,\alpha}(z))$ can be written as

$$\begin{aligned} M_m(\xi_{m,\alpha}(z)) &= \sum_{\nu=0}^{2m} w_\nu f_m(s_{2m,\nu}) f_m^T(s_{2m,\nu}) \\ &= \sum_{\nu=0}^{m} w_{2\nu} f_m(s_{2m,2\nu}) f_m^T(s_{2m,2\nu}) + \sum_{\nu=1}^{m} w_{2\nu-1} f_m(s_{2m,2\nu-1}) f_m^T(s_{2m,2\nu-1}) \\ &= \sum_{\nu=0}^{m} \frac{u|\ell_{m,\nu}(z)|}{|T_m(z)|} f_m(s_{m,\nu}) f_m^T(s_{m,\nu}) + \sum_{\nu=1}^{m} w_{2\nu-1} f_m(s_{2m,2\nu-1}) f_m^T(s_{2m,2\nu-1}) \\ &= u M_m(\xi_m^*(z)) + BWB^T, \end{aligned}$$

where $B = (f_m(s_{2m,1}), f_m(s_{2m,3}), \ldots, f_m(s_{2m,2m-1}))$ is a $(m+1)\times m$ matrix, and W is a diagonal matrix with elements $\{w_{2\nu-1}\}_{\nu=1}^m$. Then using a formula as in Rao ((1973), p. 33) we obtain the inverse of this matrix

$$\begin{aligned} M_m^{-1}(\xi_{m,\alpha}(z)) &= \frac{1}{u} M_m^{-1}(\xi_m^*(z)) \\ &\quad - \frac{1}{u^2} M_m^{-1}(\xi_m^*(z)) B \left(\frac{1}{u} B^T M_m^{-1}(\xi_m^*(z)) B + W^{-1}\right)^{-1} \\ &\quad \cdot B^T M_m^{-1}(\xi_m^*(z)). \end{aligned} \tag{A.9}$$

Now from Hoel and Levine (1964) or formula (2.13.10) of Fedorov (1972) and the fact that $T_m(z) = \sum_{\nu=0}^m |\ell_{m,\nu}(z)| > 0$, it yields

$$\begin{aligned} d_m(\xi_m^*(z), z, x) &= f_m^T(z) M_m^{-1}(\xi_m^*(z)) f_m(x) \\ &= \sum_{\nu=0}^{m} \frac{\ell_{m,\nu}(z)\ell_{m,\nu}(x)}{q_{m,\nu}} \\ &= \sum_{\nu=0}^{m} |\ell_{m,\nu}(z)| \sum_{\nu=0}^{m} (-1)^{m-\nu} \ell_{m,\nu}(x) \\ &= T_m(z)T_m(x), \end{aligned} \tag{A.10}$$

which leads for $\nu = 1, \ldots, m$, $d_m(\xi_m^*(z), z, s_{2m,2\nu-1}) = 0$. This in turn implies that $f_m^T(z) M_m^{-1}(\xi_m^*(z)) B = 0$. That is after multiplying $f_m^T(z)$ in front of $M_m^{-1}(\xi_{m,\alpha}(z))$ the second term in the right hand side of (A.9) is dropped and we have a simpler form for expressing $d_m(\xi_{m,\alpha}(z), z, x)$ as

$$d_m(\xi_{m,\alpha}(z), z, x) = \frac{1}{u} f_m^T(z) M_m^{-1}(\xi_m^*(z)) f_m(x) = \frac{1}{u} T_m(z) T_m(x).$$

As far as expressing $d_{2m}(\xi_{m,\alpha}(z), z, x)$ in a more explicit form, we have

$$\text{(A.11)} \quad d_{2m}(\xi_{m,\alpha}(z), z, x) = \sum_{\nu=0}^{2m} \frac{\ell_{2m,\nu}(z)\ell_{2m,\nu}(x)}{w_\nu} = \sum_{\nu=0}^{m} \frac{\ell_{2m,2\nu}(z)\ell_{2m,2\nu}(x)}{w_{2\nu}} + \sum_{\nu=1}^{m} \frac{\ell_{2m,2\nu-1}(z)\ell_{2m,2\nu-1}(x)}{w_{2\nu-1}}.$$

The first term on the right hand side of the last equality can be simplified to be

$$\text{(A.12)} \quad \sum_{\nu=0}^{m} \frac{\ell_{2m,2\nu}(z)\ell_{2m,2\nu}(x)}{w_{2\nu}} = \frac{T_m(z)T_m(x)}{u} \sum_{\nu=0}^{m} \frac{\ell_{m,\nu}(z)\ell_{m,\nu}(x)}{|\ell_{m,\nu}(z)|/|T_m(z)|} = \frac{T_m^2(z)T_m(x)}{u} \sum_{\nu=0}^{m} (-1)^{m-\nu} \ell_{m,\nu}(x) = \frac{T_m^2(z)T_m^2(x)}{u},$$

where we used (A.4), Lemma A.1 and $z > 1$ in order to determine the sign of $T_m(z)$ and $\ell_{m,\nu}(z)$. Before evaluating the second term on the last equality, recall that $\sum_{\nu=0}^{2m} \ell_{2m,\nu}(x) = 1$, for all $x \in R$, and that by (A.4)

$$\sum_{\nu=0}^{m} \ell_{2m,2\nu}(x) = T_m(x) \sum_{\nu=0}^{m} (-1)^{m-\nu} \ell_{m,\nu}(x) = T_m^2(x).$$

Observing these identities and Lemma A.1 we obtain for the first term in (A.11) the following expression

$$\text{(A.13)} \quad \sum_{\nu=1}^{m} \frac{\ell_{2m,2\nu-1}(z)\ell_{2m,2\nu-1}(x)}{w_{2\nu-1}} = \frac{T_{2m}(z)}{(1-\alpha)} \sum_{\nu=1}^{m} \frac{\ell_{2m,2\nu-1}(z)}{|\ell_{2m,2\nu-1}(z)|} \ell_{2m,2\nu-1}(x) = \frac{T_{2m}(z)}{(1-\alpha)} \left(\sum_{\nu=0}^{m} \ell_{2m,2\nu}(x) - 1 \right) = \frac{T_{2m}(z)}{(1-\alpha)} (T_m^2(x) - 1).$$

Combining (A.11)–(A.13) proves the second part of Lemma A.2. □

From Lemma A.2, it can be easily obtained that

$$\text{(A.14)} \quad v_m(\xi_{m,\alpha}(z), z) = \frac{T_m^2(z)}{u},$$

and

(A.15)
$$v_{2m}(\xi_{m,\alpha}(z), z) = \frac{T_m^4(z)}{u} + \frac{T_{2m}(z)(T_m^2(z)-1)}{(1-\alpha)}.$$

Now we evaluate the term on the left hand side of the equivalence theorem (2.5) at the support points $\{s_{2m,\nu}\}_{\nu=0}^{2m}$. To this end define

(A.16)
$$\phi_k(\xi_{m,\alpha}(z), z, x) = \frac{d_k^2(\xi_{m,\alpha}(z), z, x)}{v_k(\xi_{m,\alpha}(z), z)} - 1, \qquad k = m,\, 2m,$$

and let

(A.17)
$$\psi_\lambda^p(\xi_{m,\alpha}(z), z, x) = \lambda\phi_m(\xi_{m,\alpha}(z), z, x)\left\{\frac{v_m(\xi_{m,\alpha}(z), z)}{v_m(\xi_m^*(z), z)}\right\}^{-p} + (1-\lambda)\phi_{2m}(\xi_{m,\alpha}(z), z, x)\left\{\frac{v_{2m}(\xi_{m,\alpha}(z), z)}{v_{2m}(\xi_{2m}^*(z), z)}\right\}^{-p}$$

denote the left hand side of the equivalence theorem (2.5). By the fact that $\{s_{2m,2\nu-1}\}_{\nu=1}^m$ are the zeros of $T_m(x)$, and $T_m(x)$ has derivative zero at the points $\{s_{2m,2\nu}\}_{\nu=0}^m$, it is easy to see from Lemma A.2 that $\psi_\lambda^p(\xi_{m,\alpha}(z), z, x)$ has derivative zero (with respect to x) at all interior support points $\{s_{2m,\nu}\}_{\nu=1}^{2m-1}$ of the design $\xi_{m,\alpha}(z)$.

The following two lemmas are needed to evaluate (A.17) at the support points and are immediate consequences of Lemma A.2 and the identities (A.14)–(A.16).

LEMMA A.3. *At the even support points* $\{s_{2m,2\nu}\}_{\nu=0}^m$ *we have*

$$\phi_m(\xi_{m,\alpha}(z), z, s_{2m,2\nu}) = \frac{(1-u)}{u} = \frac{(1-\alpha)(T_m^2(z)-1)}{uT_{2m}(z)}, \qquad \textit{and}$$
$$\phi_{2m}(\xi_{m,\alpha}(z), z, s_{2m,2\nu}) = \frac{(T_m^2(z)-1)(\alpha^2 - 2\alpha T_m^2(z))}{uT_{2m}(z)(T_m^2(z)-\alpha)}.$$

LEMMA A.4. *At the odd support points* $\{s_{2m,2\nu-1}\}_{\nu=1}^m$ *we have*

$$\phi_m(\xi_{m,\alpha}(z), z, s_{2m,2\nu-1}) = (-1), \qquad \textit{and}$$
$$\phi_{2m}(\xi_{m,\alpha}(z), z, s_{2m,2\nu-1}) = \frac{(-1)(\alpha^2 - 2\alpha T_m^2(z))}{(1-\alpha)(T_m^2(z)-\alpha)}.$$

The proof of Theorem 3.1 is now completed by showing that α can be determined such that there is equality in (2.5) for all support points of $\xi_{m,\alpha}(z)$ and by observing that the derivative of the left hand side vanishes at all interior support points of $\xi_{m,\alpha}(z)$ (see the discussion following equation (A. 17)). Because this term is a polynomial of degree $4m$ it must attain its maximum (over the interval $[-1,1]$) at the Chebyshev points $\{s_{2m,\nu}\}_{\nu=0}^{2m}$ and the assertion follows.

To prove the remaining assertion we note that by Lemmas A.3 and A.4, the ratio $\phi_{2m}(\xi_{m,\alpha}(z), z, x)/\phi_m(\xi_{m,\alpha}(z), z, x)$, has the same value at all support points $\{s_{2m,\nu}\}_{\nu=0}^{2m}$. Therefore, when it comes to evaluate $\psi_\lambda^p(\xi_{m,\alpha}(z), z, x) = 0$ at all support points we obtain the same equation for α. Observing Lemma A.3, A.4, (A.14) and (A.15) this gives equation (3.1) of Theorem 3.1. Moreover if we let

$$g_\lambda(\alpha) = \frac{\lambda}{(1-\lambda)}(T_m^2(z)-\alpha)^{p+1}(1-\alpha)^{1-p} + \alpha T_{2m}^p(z)(\alpha - 2T_m^2(z)),$$

then solving equation (3.1) within $(0,1)$ is equivalent to finding the zeros for $g_\lambda(\alpha)$ in $(0,1)$. Note that $g_\lambda(0) = (\lambda/(1-\lambda))T_m^{2p+2}(z) > 0$, and $g_\lambda(1) = T_{2m}^p(z)(1-2T_m^2(z)) < 0$, and

$$\frac{d}{d\alpha}g_\lambda(\alpha) = -\frac{\lambda}{(1-\lambda)}\left\{\frac{(T_m^2(z)-\alpha)}{(1-\alpha)}\right\}^p \{(1+p)(1-\alpha)+(1-p)(T_m^2(z)-\alpha)\} + 2T_{2m}^p(z)(\alpha - T_m^2(z)),$$

is less than zero for all $\alpha \in (0,1)$. Consequently there is a unique solution of $g_\lambda(\alpha) = 0$ in $(0,1)$, say α_p^*, and there is equality in (2.5) for all support points of the design $\xi_{\alpha_p^*}(z)$. □

PROOF OF THEOREM 3.2. We investigate the design $\xi_{m,\alpha}(z)$ for which both efficiencies in the criterion in Section 3 ($p = -\infty$) are equal, i.e.

$$\frac{v_m(\xi_{m,\alpha}(z), z)}{v_m(\xi_m^*(z), z)} = \frac{v_{2m}(\xi_{m,\alpha}(z), z)}{v_{2m}(\xi_{2m}^*(z), z)}.$$

Noting that $\xi_{2m}^*(z) = \xi_{m,0}(z)$ we obtain from (A.10), (A.14) and (A.15)

$$\frac{1}{u} = \frac{(T_m^2(z)-\alpha)}{u(1-\alpha)T_{2m}(z)},$$

which implies that $\alpha = 1/2$. We now verify conditions in the equivalence theorem (2.6) for $p = -\infty$ and discuss equality at the support points. By Lemma A.3, A.4, (A.14) and (A.15) we find that all equations yield the same constants β_1, β_2 given by $\beta_2 = 1 - \beta_1$,

$$\beta_1 = \frac{(1/2)T_{2m}(z) + T_m^2(z)}{T_{2m}(z) + T_m^2(z)}.$$

Finally, similar arguments as given in the proof of Theorem 3.1 show that $\xi_{m,1/2}(z)$ does satisfy the equivalence theorem. Consequently $\xi_{m,1/2}(z)$ minimizes $\Pi_{-\infty}(\xi, z)$ and Theorem 3.2 is proved. □

PROOF OF THEOREM 4.1. By the linearity of the Stieltjes transform we have

$$\Phi(x, \xi_{m,\alpha}(z)) = \alpha\Phi(x, \xi_m^*(z)) + (1-\alpha)\Phi(x, \xi_{2m}^*(z)) \tag{A.18}$$

where the two expressions on the right hand side were already determined by Lau ((1983), p. 129), i.e.

$$\Phi(x, \xi_{2m}^*(z)) = \frac{(z^2-1)U_{2m-1}(z)}{T_{2m}(z)(x-z)}\left\{\frac{T_{2m}(z)}{(z^2-1)U_{2m-1}(z)} - \frac{T_{2m}(x)}{(x^2-1)U_{2m-1}(x)}\right\}, \tag{A.19}$$

$$\Phi(x, \xi_m^*(z)) = \frac{(z^2-1)U_{m-1}(z)}{T_m(z)(x-z)}\left\{\frac{T_m(z)}{(z^2-1)U_{m-1}(z)} - \frac{T_m(x)}{(x^2-1)U_{m-1}(x)}\right\}.$$

Now from the identity $2T_m(x)U_{m-1}(x) = U_{2m-1}(x)$ we have

$$\Phi(x, \xi_m^*(z)) = \frac{(z^2-1)U_{2m-1}(z)}{T_m^2(z)(x-z)}\left\{\frac{T_m^2(z)}{(z^2-1)U_{2m-1}(z)} - \frac{T_m^2(x)}{(x^2-1)U_{2m-1}(x)}\right\} \tag{A.20}$$

and a combination of (A.18)–(A.20) yields the desired result. □

PROOF OF COROLLARY 3.2. Observing the relation

$$2T_k(x) = (x+\sqrt{x^2-1})^k + (x-\sqrt{x^2-1})^k \tag{A.21}$$

$$U_{k-1}(x) = \frac{(x+\sqrt{x^2-1})^k - (x-\sqrt{x^2-1})^k}{2\sqrt{x^2-1}} \tag{A.22}$$

(see Rivlin (1990), p. 5) and Theorem 4.1 ($\alpha = 0$), it is straightforward to show that

$$\lim_{m\to\infty} \Phi(x, \xi^*_{2m}(z)) = \frac{1}{x-z}\left\{1 - \frac{\sqrt{z^2-1}}{\sqrt{x^2-1}}\right\} = \Phi(x)$$

uniformly on closed subsets of $C \setminus [-1,1]$, where the sign of the square root is defined by the condition $|x+\sqrt{x^2-1}| > 1$. By Theorem A.1 in Van Assche (1987) the sequence $(\xi^*_{2m}(z))_{m\in\mathcal{N}}$ converges weakly to a measure $\xi^*(z)$ with Stieltjes transform $\Phi(x)$, and the inversion formula for the Stieltjes transform (see Dette and Studden (1997), Section 3.6)) shows that $\xi^*(z)$ has a continuous density on $(-1,1)$ given by (3.2). The same argument shows that $\xi^*_m(z) \Rightarrow \xi^*(z)$. Note that the weak convergence of the optimal extrapolation design (for a fixed degree) was already established by Kiefer and Studden (1976) by a different method.

Finally, the identity (A.21) shows that the solution $\alpha^*_{\lambda,p}$ of equation (3.1) converges (as $m \to \infty$) to the unique solution β^* of the equation

$$\begin{cases} \dfrac{\lambda}{1-\lambda} = 2^{p+1}\alpha(1-\alpha)^{p-1} & \text{if } -\infty < p \le 1 \\ \alpha = \dfrac{1}{2} & \text{if } p = -\infty. \end{cases} \tag{A.23}$$

Consequently we have $\xi^*_{m,\alpha^*_{\lambda,p}}(z) \Rightarrow \beta^*\xi^*(z) + (1-\beta^*)\xi^*(z) = \xi^*(z)$ which proves the assertion of Corollary 3.2. □

PROOF OF THEOREM 4.2. Using a similar argument as in Dette (1995*b*) we obtain the following duality for the extremal problem stated in Theorem 4.2.

$$\max\{\lambda P^2_m(z) + (1-\lambda)P^2_{2m}(z) \mid \max_{x\in[-1,1]}(P^2_m(x) + P^2_{2m}(x)) \le 1\} \tag{A.24}$$

$$= \min_{\xi} \max\{\lambda v_m(\xi,z), (1-\lambda)v_{2m}(\xi,z)\} \tag{A.25}$$

where the minimum in the second expressions varies over the class of all measures with nonsingular information matrix $M_{2m}(\xi)$. Moreover, the general equivalence theorem of Pukelsheim (1993) shows that ξ^* is a solution of the optimal design problem in (A.25) if and only if there exist constants γ_1, $\gamma_2 \ge 0$, $\gamma_1 + \gamma_2 = 1$ such that

$$\sum_{l=1}^{2} \gamma_l \frac{d^2_{ml}(\xi^*, z, x)}{v_{ml}(\xi^*, z)} \le 1, \qquad \forall x \in [-1,1]. \tag{A.26}$$

Whenever the maximum in (A.25) is not attained for $\lambda v_m(\xi^*, z)$ we have $\gamma_1 = 0$, otherwise the maximum is attained for both expressions, i.e. $\lambda v_m(\xi^*, z) = (1-\lambda)v_{2m}(\xi^*, z)$; see Pukelsheim (1993). The solution of the extremal problem (P) is now simply obtained by defining

$$P^2_l(x) = \gamma_l \frac{d^2_{ml}(\xi^*, z, x)}{v_{ml}(\xi^*, z)}, \qquad l = 1, 2, \tag{A.27}$$

because this choice yields equality between (A.24) and (A.25) in all cases.

If $\beta(\lambda) = \lambda/(1-\lambda) < T_{2m}(z)$ and $\xi^*_{2m}(z)$ denotes the optimal extrapolation design in the regression of degree $2m$, then it follows from (A.7) and (A.8) ($\alpha = 0, x = z$) that

$$\lambda v_m(\xi^*_{2m}(z), z) < T^2_{2m}(z)(1-\lambda) = (1-\lambda)v_{2m}(\xi^*_{2m}(z), z).$$

In this case we have $\gamma_1 = 0$, $\gamma_2 = 1$ in (A.26), which reduces to $T^2_{2m}(x) \leq 1$, by Lemma A.2 ($\alpha = 0$). Consequently $\xi^*_{2m}(z)$ minimizes (A.25) and by (A.27) the extremal polynomials are given by $P_m(x) \equiv 0$, $P_{2m}(x) = \mp T_{2m}(x)$.

On the other hand, if $\beta(\lambda) = \lambda/(1-\lambda) \geq T_{2m}(z)$, then the minimum in (A.25) must be attained in both expressions. By the results of the previous section it is conjectured that the minimizing measure of (A.25) is a convex combination of $\xi^*_m(z)$ and $\xi^*_{2m}(z)$, i.e. $\xi_{m,\alpha}(z) = \alpha\xi^*_m(z) + (1-\alpha)\xi^*_{2m}(z)$. The equality in the maximum of (A.25) and Lemma A.2 determine α, i.e.

$$\text{(A.28)} \qquad \alpha^* = \frac{T^2_m(z)(\beta - T_{2m}(z))}{\beta T^2_m(z) - T_{2m}(z)}.$$

Note that $\alpha^* \in [0,1]$, because $\beta > T_{2m}(z)$, $z > 1$. Now a straightforward calculation gives for the quantity u in Lemma A.2

$$\text{(A.29)} \qquad u = \frac{T^2_m(z)(\beta - T^2_m(z))}{\beta T^2_m(z) - T_{2m}(z)}$$

and (A.7) and (A.8) reduce to

$$\text{(A.30)} \qquad d_m(\xi_{m,\alpha^*}(z), z, x) = \frac{T_m(x)}{T_m(z)} \cdot \frac{\beta T^2_m(z) - T_{2m}(z)}{\beta - T^2_m(z)}$$

$$\text{(A.31)} \quad d_{2m}(\xi_{m,\alpha^*}(z), z, x) = \frac{\beta T^2_m(z) - T_{2m}(z)}{(T^2_m(z) - 1)(\beta - T^2_m(z))}\{(\beta - 1)T^2_m(x) - \beta + T^2_m(z)\}.$$

Consequently by Lemma A.2, (A.30), (A.31) the equivalence theorem (A.26) can be rewritten as

$$\text{(A.32)} \quad \gamma_1 \frac{T^2_m(x)}{T^2_m(z)} \frac{\beta T^2_m(z) - T_{2m}(z)}{\beta - T^2_m(z)} + \gamma_2 \frac{\beta T^2_m(z) - T_{2m}(z)}{\beta(T^2_m(z) - 1)^2(\beta - T^2_m(z))} \left\{(\beta - 1)T^2_m(x) - \beta + T^2_m(z)\right\}^2 \leq 1.$$

If $\xi_{m,\alpha^*}(z)$ is a solution of (A.25) we must have equality at $s_{2m,\nu}$ ($\nu = 0, \ldots, 2m$). It can be shown that all these conditions yield the same equation for $\gamma_2 = 1 - \gamma_1$, i.e.

$$\text{(A.33)} \qquad \gamma_2 = \frac{\beta(T^2_m(z) - 1)^2}{(\beta T^2_m(z) - T_{2m}(z))(\beta - T^2_m(z))}$$

$$\text{(A.34)} \qquad \gamma_1 = 1 - \gamma_2 = \frac{T^2_m(z)(\beta - 1)(\beta - T_{2m}(z))}{(\beta T^2_m(z) - T_{2m}(z))(\beta - T^2_m(z))}.$$

Finally a similar argument as given in the proof of Theorem 3.1 shows that $\xi_{m,\alpha^*}(z)$ satisfies (A.26) and consequently minimizes (A.25). The assertion of Theorem 4.2 now follows from (A.27), (A.30), (A.31), (A.33) and (A.34). □

References

Atkinson, A. C. and Cox, D. R. (1974). Planning experiments for discriminating between models (with discussion), *J. Roy. Statist. Soc. Ser. B*, **36**, 321–348.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design, *J. Amer. Statist. Assoc.*, **54**, 622–654.

Chao, M. T. (1995). Some robust type of extrapolation designs, *Proceedings, International Conference on Statistical Methods and Statistical Computing and Productivity Improvement*, **1**, 99–108.

Dette, H. (1991). A note on robust designs for polynomial regression, *J. Statist. Plann. Inference*, **28**, 223–232.

Dette, H. (1994). Discrimination designs for polynomial regression on a compact interval, *Ann. Statist.*, **22**, 890–904.

Dette, H. (1995*a*). Optimal designs for identifying the degree of a polynomial regression, *Ann. Statist.*, **23**, 1248–1267.

Dette, H. (1995*b*). A note on some peculiar extremal phenomena of the Chebyshev polynomials, *Proc. Edinburgh Math. Soc.*, **38**, 343–355.

Dette, H. and Studden, W. J. (1995). Optimal designs for polynomial regression when the degree of the polynomial is not known, *Statist. Sinica*, **5**, 459–474.

Dette, H. and Studden, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis*, New York, Wiley.

Dette, H. and Wong, W. K. (1996). Robust optimal extrapolation designs, *Biometrika*, **83**, 667–680.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

Hoel, P. G. (1965). Minimax designs in two dimensional regression, *Ann. Math. Statist.*, **36**, 1097–1106.

Hoel, P. G. and Levine, A. (1964). Optimal spacing and weighting in polynomial prediction, *Ann. Math. Statist.*, **35**, 1553–1560.

Huang, M.-N. L. and Studden, W. J. (1988). Model robust extrapolation designs, *J. Statist. Plann. Inference*, **13**, 1–24.

Huber, P. J. (1975). Robustness and designs, *A Survey of Statistical Design and Linear Models*, 287–303, North Holland, Amsterdam.

Karlin, S. and Studden, W. J. (1966). Optimal experimental designs, *Ann. Math. Statist.*, **37**, 783–815.

Kiefer, J. C. (1980). Designs for extrapolation when bias is present, *Multivariate Analysis-V: Proceedings of the Fifth International Symposium on Multivariate Analysis*, North-Holland, Amsterdam.

Kiefer, J. and Studden, W. J. (1976). Optimal designs for large degree polynomial regression, *Ann. Statist.*, **4**, 1113–1123.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems, *Ann. Math. Statist.*, **30**, 271–294.

Kiefer, J. and Wolfowitz, J. (1965). On a theorem of Hoel and Levine on extrapolation, *Ann. Math. Statist.*, **36**, 1627–1655.

Lau, T. S. (1983). Theory of canonical moments and its application in polynomial I and II, Technical Reports, 83–23, 83–24, Department of Statistics, Purdue University, Lafayette, Indiana.

Läuter, E. (1974*a*). Experimental design for a class of models, *Mathematische Operationsforschung und Statistik*, **5**, 379–398.

Läuter, E. (1974*b*). Optimal multipurpose designs for regression models, *Mathematische Operationsforschung und Statistik*, **7**, 51–68.

Preitschopf, F. and Pukelsheim, F. (1987). Optimal designs for quadratic regression, *J. Statist. Plann. Inference*, **16**, 213–218.

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

Pukelsheim, F. and Rosenberger, J. L. (1993). Experimental designs for model discrimination, *J. Amer. Statist. Assoc.*, **88**, 642–649.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Rivlin, T. J. (1990). *Chebyshev Polynomials*, Wiley, New York.

Sacks, J. and Ylvisaker, D. (1984). Some model robust designs in regression, *Ann. Statist.*, **12**, 1324–1348.

Spruill, M. C. (1984). Optimal designs for minimax extrapolation, *J. Multivariate Anal.*, **15**, 52–62.

Stigler, S. M. (1971). Optimal experimental design for polynomial regression, *J. Amer. Statist. Assoc.*, **66**, 311–318.

Studden, W. J. (1968). Optimal designs on Tchebyscheff points, *Ann. Math. Statist.*, **39**, 1435–1447.

Studden, W. J. (1971). Optimal designs for multivariate polynomial extrapolation, *Ann. Math. Statist.*, **42**, 828–832.

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression, *J. Amer. Statist. Assoc.*, **77**, 916–921.

Van Assche, W. (1987). *Asymptotics for Orthogonal Polynomials*, Lecture Notes in Mathematics, 1265, Springer, Berlin.