Optimal and Admissible Designs for Polynomial Monospline Regression*

by

Norman T. Bruvold

Department of Statistics

Division of Mathematical Science

Mimeograph Series #254

May 14, 1971

*This research was supported in part by the National Science Foundation Grant GP 20306 and the Office of Naval Research Contract N0014-67-A-0226-0014 at Purdue University. Reproduction is permitted in whole or in part for any purposes of the United States Government.

Abstract - Optimal and Admissible

Designs for Polynomial Monospline Regression

by Norman T. Bruvold

We consider regression of the form $\sum_{i=0}^{n} a_i x^i + \sum_{i=1}^{h} \sum_{j=\ell_i}^{k_i} b_{ij}(x-\xi_i)_+^{n-j}$ where $n-1 \geq k_i \geq \ell_i \geq 0$, $a<\xi_1<\ldots<\xi_h<b$ and $x \varepsilon [a,b]$. We define admissibility in terms of a positive semi-definite difference of information matrices. Some sufficient conditions for admissibility on the spectrum of a design is given.

When $\ell_1=1$, $h=1$ and ξ_1 lies in the center of the interval $[a,b]$, optimal experimental designs for the individual regression coefficients are given. Some of the optimal designs are not unique but are convex combinations of two probability measures. Three distinct bases are considered.

Extrapolation and minimax extrapolation designs are given for the centered knot situation along with some other special cases.

CHAPTER I

INTRODUCTION

1.1. Introduction to Admissible and Optimal Designs.

Let $\overline{f(x)} = (f_0(x), f_1(x),\ldots,f_n(x))$ denote a vector of n+1 linearly independent continuous functions on a compact space $\mathcal{X}$. The points of $\mathcal{X}$ are referred to as the possible levels of feasible experiments. For each level $x \in \mathcal{X}$ some experiment can be performed whose outcome is a random variable Y(x). We assume that Y(x) has a mean of the explicit form

1.1.1. $$E\,Y(x) = \sum_{j=0}^{n} \theta_j f_j(x)$$

and a common variance σ^2 independent of x which is unknown. In most instances we will assume for convenience that the variance is normalized = 1. The functions $f_0(x),\ldots,f_n(x)$ are called the regression functions and are assumed known to the experimenter. The parameters $\theta_0,\ldots,\theta_n$ are unknown. The problem concerned with here is the estimation of functions of the vector $\overline{\theta} = (\theta_0,\ldots,\theta_n)$ by means of a finite number N of uncorrelated observations $\{Y(x_i)\}_{i=1}^{N}$.

An experimental design is a probability measure μ concentrating mass $p_1,\ldots,p_r$ on the points $x_1,\ldots,x_r$ where the values

$$p_i N = n_i \qquad i = 1,2,\ldots,r$$

are integers. The associated experiment involves taking n_i uncorrelated observations of the random variable $Y(x_i)$, $i = 1,2,\ldots,r$. An experimental design determines the points at which the experiment takes place, namely the x_i, $i = 1,\ldots,r$ and the number n_i of experiments at each level x_i. Given a criterion of what a good estimate of a certain $h(\overline{\theta})$ is, the problem confronting the experimenter is to choose the design possessing certain optimality properties.

Definition 1.1.1. Let μ be an arbitrary probability measure on the Borel sets $\mathcal{B}$ of $\mathcal{X}$ where $\mathcal{B}$ includes all one point sets. $M(\mu)$, the information matrix of μ, is defined as $||m_{ij}(\mu)||_{i,j=0}^{n}$, where

1.1.2. $$m_{ij}(\mu) = \int_{\mathcal{X}} f_i(x) f_j(x) \mu(dx).$$

The information matrix plays an important role in the following chapters in determining the accuracy of estimates to various $h(\overline{\theta})$.

If the unknown parameter vector $\overline{\theta}$ is estimated by the method of least squares thus securing a best linear unbiased estimate, say $\hat{\theta}$, then the covariance matrix of $\hat{\theta}$ is given by

1.1.3. $$E(\hat{\theta}-\overline{\theta})(\hat{\theta}-\overline{\theta})' = \frac{\sigma^2}{N} M^{-1}(\mu)$$

where μ assigns mass $p_i = n_i/N$ to the points x_i, $i = 1,2,\ldots,r$.

If the matrix $M^{-1}(\mu)$ is "small" according to some criterion, or $M(\mu)$ is "large", then roughly speaking $\hat{\theta}$ is close to $\overline{\theta}$. Most criteria for discerning optimality of an experimental design are based on maximizing some appropriate functional of the matrix $M(\mu)$.

Definition 1.1.2. A linear form

1.1.4 $$(\overline{c},\overline{\theta}) = \sum_{i=0}^{n} c_i\theta_i$$

is called estimable with respect to μ if $\overline{c} = (c_0,\ldots,c_n)$ is contained in the range of the matrix $M(\mu)$.

A criterion for optimality, formalized and interpreted in Kiefer (1959), is as follows:

If $\overline{c}$ is estimable with respect to μ let

1.1.5. $$V(\overline{c},\mu) = \sup \frac{(\overline{c},\overline{d})^2}{(\overline{d},M(\mu)\overline{d})}$$

where the sup is taken over the set of vectors $\overline{d}$ such that the denominator is non-zero. If $\overline{c}$ is not estimable with respect to μ we define $V(\overline{c},\mu) = \infty$. If μ is an experimental design and if we take n uncorrelated observations according to μ then the variance of the best linear unbiased estimate of $(\overline{c},\overline{\theta})$ is given by

$$\frac{\sigma^2}{n} V(\overline{c},\mu).$$

With this in mind we are able to define the concepts of admissibility and optimality as used in this thesis.

1.2. Introduction to Admissible Designs.

If we have an experimental design μ' such that $M(\mu')-M(\mu)$ is non-negative definite it follows that if $\overline{c}$ is estimable with respect to μ

then $\bar{c}$ is also estimable with respect to μ'. Karlin and Studden (1966 b, page 788). Since the set of vectors $\bar{d}$ for which the denominator of $V(\bar{c},\mu')$ is non-zero, say $D_{\mu'}$, is contained in D_μ we have that $V(\bar{c},\mu')$ is at least smaller than $V(\bar{c},\mu)$. With this in mind we may think of μ' as giving a better best variance than μ for linear unbiased estimates of $(\bar{c},\bar{\theta})$. This motivates the definition of admissibility.

Definition 1.2.1. Let μ and ν be probability measures on $\mathcal{X}$. We say $\mu \geq \nu$ or $M(\mu) \geq M(\nu)$ if the matrix $M(\mu)-M(\nu)$ is non-negative definite and unequal to the zero matrix.

Definition 1.2.2. A probability measure or design μ is said to be admissible if there is no design ν such that $\nu \geq \mu$. Otherwise μ is inadmissible.

Because inadmissible designs give at least larger variances than their dominating designs and because every inadmissible design is dominated by an admissible design, Van Arman (1968), we are interested in the class of admissible designs.

Definition 1.2.3. Let μ be a probability measure on $\mathcal{X}$ concentrated on $\{x_1,\ldots,x_r\}$ such that

$$\mu(x) \quad > 0 \qquad \text{for } x=x_i \quad i=1,\ldots,r$$
$$\qquad = 0 \qquad \text{otherwise}$$

$$\text{and} \quad \sum_{i=1}^{r} \mu(x_i) = 1$$

then the set $\{x_1,\ldots,x_r\}$ is called the spectrum of μ (also the support of μ) and is written as $S(\mu)$. When we mention that μ is supported by the full set A we mean that $S(\mu) = A$.

The concept of admissibility of a design is essentially a property of the spectrum. In other words if ν is admissible and μ is an experimental design such that $S(\mu) \subset S(\nu)$ then μ is admissible. Elfving (1959). It is clear from this that if two experimental designs have the same spectrum they are either both admissible or both inadmissible. Thus we may classify admissible or inadmissible designs by properties of their spectra.

When $\bar{f}(x) = (1,x,\ldots,x^n)$ the class of admissible designs for $\mathcal{X} = [a,b]$ have been completely characterized by Kiefer (1959). His results show that a spectrum in [a,b] is admissible if it contains no more than n-1 points on the open interval (a,b).

When we consider the interval [a,b] and choose h fixed points or "knots" $\xi_1,\ldots,\xi_h$ such that $a < \xi_1 < \xi_2 < \ldots < \xi_h < b$, and the vector of regression functions $\bar{f}(x)$ is in the following form

1.2.1.
$$1,x,\ldots,x^n$$
$$(x-\xi_i)_+^{n-k_i},\ (x-\xi_i)_+^{n-k_i+1},\ldots,(x-\xi_i)_+^{n}\quad i = 1,2,\ldots,h$$

where

1.2.2.
$$(x-\xi)_+^m = \begin{cases} 0 & x < \xi \\ (x-\xi)^m, & x \geq \xi \end{cases},\quad m = 1,2,\ldots\ ;$$

the class of admissible designs have been completely characterized by Studden and Van Arman (1970). Their results show that a design μ is admissible if and only if the spectrum of μ, $S(\mu)$, has less than or equal to

1.2.3. $$n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]$$

points on the open interval $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$; $\ell=0,1,\ldots,h$. (Here we let $\xi_0=a$, $\xi_{n+1}=b$ and $[x]$ denotes the greatest integer in x.) A polynomial in the component functions of $\bar{f}(x)$ (1.2.1) is called a polynomial spline function of degree n. Spline functions have received considerable attention from mathematicians working in numerical analysis, interpolation and approximation theory. (See Karlin (1968), Rice (1969), and Shoenberg (1964), for further references.)

1.3. Introduction to Optimality

When estimating the linear form $(\bar{c},\bar{\theta})$ where $\sum_{i=1}^{n} c_i^2 > 0$ we are interested in those designs that minimize the variance of the best linear unbiased estimate of $(\bar{c},\bar{\theta})$.

Definition 1.3.1. A probability measure or design μ is said to be optimal with respect to the estimation of $(\bar{c},\bar{\theta})$ if μ minimizes $V(\bar{c},\mu)$. We will also refer to the above designs as $\bar{c}$-optimal.

If $\bar{c}=\bar{f}(x)$ for some fixed value of $x \in \mathcal{X}$ we shall write $V(x,\mu)$ for $V(\bar{c},\mu)$. In the following discussions we will be mainly concerned with the determination of $\bar{c}_p$-optimal designs where

$$\bar{c}_p = (0,\ldots,0,1,0,\ldots,0)$$

with a 1 only in the (p+1)st component. If we consider polynomial regression where

1.3.1 $$\bar{f}(x) = (1,x,\ldots,x^n) \quad \text{for } x\epsilon\ [-1,1]$$

the c_n-optimal design was originally given in Kiefer and Wolfowitz (1959) while the remaining $\bar{c}_p$-optimal designs are given in Studden (1968). If n-p is even the unique c_p-optimal design is supported by the full set of Tchebycheff points $s_0,s_1,\ldots,s_n$ associated with the functions (1.3.1). If n-p is odd the unique $\bar{c}_p$-optimal design is supported by the full set of Tchebycheff points $t_0,t_1,\ldots,t_{n-1}$ associated with the functions $(1,x,\ldots,x^{n-1})$. Hoel and Levine (1964) showed that if $\bar{f}(x)$ is as in (1.3.1) then the $\bar{c}$-optimal design for $\bar{c}=\bar{f}(x_0)$ with $|x_0|>1$ is supported on the Tchebycheff points $s_0,s_1,\ldots,s_n$ associated with the functions (1.3.1).

Murty (1969) gives the c_p-optimal designs for the set of regression functions

1.3.2. $$(1,x,\ldots,x^n,x_+^k,\ldots,x_+^n) \quad \text{for } x\epsilon\ [-1,1].$$

Essentially the regression parameters are separated into two groups dependent on their relation to n and k. In both cases the $\bar{c}_p$-optimal designs are unique while one group is supported by the same set of (2n-k+2) points and the other by a set of (2n-k+1) points.

Chapter II will begin with some statements of known results on which the discussions following are dependent. Section 2.1 will present these background lemmas along with discussions that permit us to classify admissible experimental designs according to their finite spectra. The remaining part of the chapter, section 2.2, is concerned with determining the admissible designs for regression in the functions

$$1.3.3.\quad \begin{cases} 1, x, \ldots, x^n \\ (x-\xi_i)_+^{n-k_i}, \ldots, (x-\xi_i)_+^{n-1} \qquad x \varepsilon\ [a,b] \\ i = 1,2,\ldots,h:\ n-1 \geq k_i \geq 1,\ a < \xi_1 < \xi_2 < \ldots < \xi_h < b. \end{cases}$$

In chapter III we are concerned with polynomial monospline regression with a single multiple knot in the center. The optimal designs for the individual regression coefficients are obtained for the regression function expressed in three different bases. Each of the bases are handled in a different section. There are many similarities in the treatments but each is distinct. Several examples are presented in each of the sections.

In chapter IV we consider some special cases of monospline regression with non-centered knots. In section 4.2 we treat the almost centered knot that corresponds to the non-unique designs of chapter III. Section 4.3 defines the Johnson monosplines and works with the optimal designs for monospline regression with these special knots.

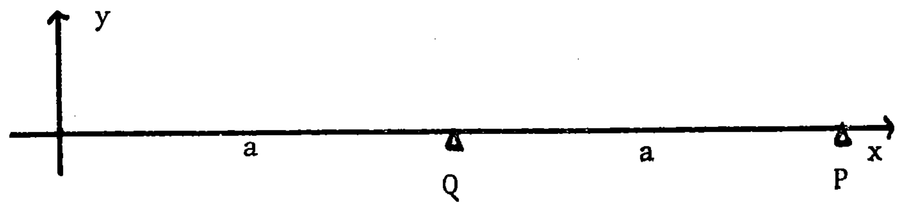
Chapter V treats extrapolation designs for the basis of section 3.3. (Extrapolation designs are independent of the basis). Minimax extrapolation designs are discussed and are found to be a particular extrapolation design.

We conclude this introductory chapter with a discussion of an early application of monosplines.

1.4. Thin Beam Monospline

A uniform heavy beam OP (see 1.4.1) of length 2a and weight W is hinged at O and rests on two smooth supports, one at P and the other at its middle point Q.

1.4.1.



Let us suppose that the beam is subjected to a force ω per unit length in the negative sense either due to its own weight or to a load placed on it. Let the x and y axis be as positioned in (1.4.1). The fundamental differential equation in the theory of thin elastic beams applied in this situation is

1.4.2. $$k \frac{d^4y}{dx^4} = -\omega$$

where k is a constant. Care must be taken in integrating (1.4.2) because discontinuities in d^3y/dx^3 occur when we pass an isolated load on a support of the beam. There are no discontinuities in the bending moment hence no discontinuities in y, dy/dx, or d^2y/dx^2. Such discontinuities would imply sudden changes in the height of the beam, in its direction, or in the bending moment.

Integrating (1.4.2) and considering the initial conditions and continuity requirements the equation of the two portions of the beam can be written in the monospline polynomial form as

1.4.3. $$ky(x) = -1/48\ \omega a^3 x + 1/16\ \omega a x^3 - 1/24\ \omega x^4 + 5/24\ \omega a(x-a)^3_+.$$

A more complete discussion of the above can be found in Synge and Griffith (1942, pages 92-98). If an experimenter were interested in the estimation of a particular parameter in (1.4.3) then the designs considered in examples (3.2.3) and (3.3.3) would be strong candidates for his consideration.

Piecewise cubic polynomial functions of one variable with continuous slope and curvature have long been used by draftsmen and engineers. For practical design work mechanical splines have been used: thin beams carrying loads ω_i concentrating at points ξ_i, according to the classical Euler-Bernoulli theory. Such mechanical splines (thin beams) have been used as analog computers to fair curves through given sets of points. Birkhoff and DeBoor (1965, pages 165-166).

By using clamped splines one can represent very accurately horizontal plane sections of ship hulls. Typically ship hulls have long straight midsections onto which a smooth pointed bow and stern are appended. The types of curves required to represent a water line must be continuous and have continuous slope and curvature or, what amounts to the same, have continuous first and second derivatives. To some extent a batten or spline held in place by so-called ducks, as it is used in the drawing of ship lines, can be approximated by a thin beam supported at a finite number of points.

The analogy between a spline and a thin beam gave rise to the name "spline curve". One of the reasons for choosing spline curves as typical ship lines was the fact that ship lines often contain straight portions. A polynomial or any other analytic function cannot contain straight portions as well as curved ones. For further discussion of fitting ship lines by splines see Theilheimer and Starkweather (1961).

CHAPTER II

CHARACTERIZATION OF ADMISSIBLE DESIGNS FOR POLYNOMIAL MONOSPLINE REGRESSION

2.1. Introduction with Background Lemmas

In this section we present some known lemmas that will be used in the remaining work in the chapter. The first lemma describes some basic properties of information matrices.

Lemma 2.1.1. Let $\overline{f}(x) = (f_0(x),\ldots,f_n(x))$ be a vector valued function composed of n+1 linearly independent continuous functions defined on a compact space $\mathcal{X}$. Let $M(\mu)$ be as in definition 1.1.1. Then

(1) for each μ, $M(\mu)$ is positive semi-definite;

(2) $\det M(\mu)=0$ whenever $S(\mu)$ contains less than n+1 points;

(3) the family of matrices $M(\mu)$, as μ ranges over the class of probability measures, is a convex compact set;

(4) for each μ there is a probability measure μ' concentrated on r points, $r \leq \frac{(n+1)(n+2)}{2} + 1$, such that $M(\mu) = M(\mu')$.

For a proof of these familiar properties of the matrices $M(\mu)$ in the above setting see Karlin and Studden (1966 b, page 787).

Part (4) of this lemma allows us to restrict our attention to probability measures concentrating their mass on a finite number of points. If a probability measure in part (4) is not an experimental design, then it can still be viewed as an approximate experimental design for large

N. Since we may classify admissible experimental designs by their spectra, we will restrict our consideration to those spectra with a finite set of points. This next lemma allows us to consider admissible or inadmissible spectra.

Lemma 2.1.2. Let μ be an admissible experimental design concentrated on $\{x_1,\ldots,x_r\}$ with weight $p_i>0$ at x_i such that $\sum_{i=1}^{r} p_i=1$. Then the experimental design ν with weight $q_i \geq 0$ at x_i such that $\sum_{i=1}^{r} q_i=1$ is also admissible. Elfving (1959, page 71).

This lemma tells us that any measure concentrated on a subset of the spectrum of an admissible spectrum is admissible, or that a subspectrum of an admissible spectrum is admissible. It also tells us that if μ is inadmissible, a measure whose spectrum contains that of μ is inadmissible.

The next lemma guarantees that for any inadmissible design we can find an admissible design that will be dominant.

Lemma 2.1.3. Let μ be an inadmissible design. Then there is an admissible design ν such that $\nu \geq \mu$. Van Arman (1968).

This lemma also tells us that we get best linear unbiased estimation results by staying in the admissible design class.

The following lemma gives a characterization of the type of regression function we are considering.

Lemma 2.1.4. A function B(x) on [a,b] can be expressed in the form

2.1.1. $$B(x) = \sum_{i=0}^{n} a_i x^i + \sum_{i=1}^{h} \sum_{j=\ell_i}^{k_i} b_{ij}(x-\xi_i)_+^{n-j}$$

where $$n-1 \geq k_i \geq \ell_i \geq 0$$

if and only if

(1) $B(x)$ is an ordinary polynomial of degree at most n in each of the intervals:

$$[a,\xi_1),(\xi_1,\xi_2),\ldots,(\xi_{h-1},\xi_h),(\xi_h,b];$$

(2) $B(x)$ has $n-k_i-1$ continuous derivatives at ξ_i, $i=1,2,\ldots,h$; and

(3) the coefficients of x^m in (ξ_i,ξ_{i+1}) are the same as those in (ξ_{i-1},ξ_i) for $m=n-\ell_i+1,\ldots,n$.

Proof: For (1) and (2) see Karlin and Ziegler (1966, page 518). This implies that (1) and (2) hold if and only if $B(x)$ is of the form

$$B(x) = \sum_{i=0}^{n} a_i x^i + \sum_{i=1}^{h} \sum_{j=0}^{k_i} b_{ij} \; (x-\xi_i)_+^{n-j} \quad .$$

(3) implies that $b_{ij}=0$ for $j=0,1,\ldots,\ell_i-1$. Otherwise the coefficient of x^m would change going across intervals.

Next we consider an important result of Karlin and Ziegler (1966, pages 519-522) paraphrased for polynomial splines.

Definition 2.1.1. For any vector of functions $\overline{f}(x)=(f_1(x),\ldots,f_h(x))$ and vector of constants $\overline{t} = (t_1,\ldots,t_h)$ where $t_1 \leq t_2 \leq \ldots \leq t_h$, we define $M(\overline{t},\overline{f})$ to be the matrix with the vector $\overline{f}(t_i)$ in the ith row. If t_i values coincide then the successive rows are replaced by successive derivatives taken from the right.

Lemma 2.1.5. Let $\overline{f}$ denote the vector of functions

$$2.1.2. \quad \begin{cases} 1,x,\ldots,x^s \\ \\ (x-\xi_i)_+^{s-\lambda_i},\ldots,(x-\xi_i)_+^s \qquad i=1,\ldots,h. \end{cases}$$

Let $\overline{t}=(t_1,\ldots,t_r)$ where $r=s+1+h+\sum_{j=1}^{h}\lambda_j$, no more than $(s-\lambda_i+1)$ t_j values are ξ_i, and no more than $(s+1)$ t_j values coincide. Then $M(\overline{t},\overline{f})$ is non-singular if and only if

$$2.1.3. \qquad t_{\gamma_i} < \xi_i < t_{s+2+\gamma_{i-1}} \qquad i=1,2,\ldots,h,$$

where $\gamma_i = \sum_{j=1}^{i} (\lambda_j+1)$, $i=1,2,\ldots,h$, $\gamma_0=0$. For some further discussion of this application see Studden and Van Arman (1969, pages 1561-1562).

The proofs of many statements in this thesis will require a somewhat delicate analysis of the zeros of polynomials in the functions (2.1.2). This is due mainly to the fact that spline polynomials are not infinitely differentiable and non-trivial spline polynomials may vanish identically on intervals between knots. All systems of functions we shall use will be linearly independent so that a linear combination of these functions will be trivial or identically zero on $(-\infty,\infty)$ if and only if all the coefficients vanish.

We shall use the following conventions when counting the zeros of a spline polynomial $P(x)$ (See Karlin and Schumaker (1967).):

(a) No zeros are counted on any open interval (ξ_i,ξ_j) if $P(x)\equiv 0$ there.

(b) The multiplicity of a zero $z\neq\xi_i$, $i=1,2,\ldots h$, is counted in the usual manner, i.e., z is a zero of order r if

$$P^{(j)}(z)=0,\ j=1,\ldots,r-1,\ P^{(r)}(z)\neq 0.$$

(c) If $P(x)\equiv 0$ on (ξ_{i-1},ξ_i) and $\neq 0$ on (ξ_i,ξ_{i+1}) the zero at ξ_i is counted as in (b) using right hand derivatives. Similarly we use left hand derivatives for $P(x)\not\equiv 0$ on (ξ_{i-1},ξ_i) and $\equiv 0$ on (ξ_i,ξ_{i+1}).

(d) If $P(x)\not\equiv 0$ on (ξ_{i-1},ξ_i) or (ξ_i,ξ_{i+1}) and

$$P^{(j)}(\xi_i-) = P^{(j)}(\xi_i+)=0 \quad j=0,1,\ldots,r-1,$$

and if $A=P^{(r)}(\xi_i-)\neq P^{(r)}(\xi_i+)=B$,

then ξ_i is a zero of order

(i) r if $AB>0$;

(ii) r+1 if $AB<0$;

(iii) r+1 if $AB=0$ and $B-A>0$;

r+2 if $AB=0$ and $B-A<0$.

It is easily seen that a zero of order r of P(x) is a zero of order r-1 of P'. We let Z(P) denote the number of zeros of P according to the above conventions.

<u>Lemma 2.1.6.</u> A non-trivial polynomial P in the functions

$$1,x,\ldots,x^s$$

2.1.4.

$$(x-\xi_j)_+^{p_j},\ldots,(x-\xi_j)_+^{s},\ j=1,2,\ldots,h,$$

where $1 \le p_j \le s$, has

$$Z(p) \le s + \sum_{j=1}^{h} (s-p_j+1).$$

For a proof see Studden and Van Arman (1969, page 1563).

2.2. Admissible Designs

In this section of the chapter we will be concerned with classifying the admissible experimental designs relative to regression of the form

$$B(x) = \sum_{i=0}^{n} a_i x^i + \sum_{i=1}^{h} \sum_{j=1}^{k_i} b_{ij}(x-\xi_i)_+^{n-j} \tag{2.2.1}$$

where $n-1 \ge k_i \ge 1$, $i = 1,2,\ldots,h$; $a < \xi_1 <\ldots< \xi_h < b$ and $x \varepsilon [a,b]$. We call regression of the form (2.2.1) monospline regression. We will first establish a moment condition for admissibility. Next we will restrict the class of admissible designs for (2.2.2) as a proper subclass of the admissible designs for (1.2.1), and then we will give a sufficient condition for admissibility. Following this will be several examples illustrating the delicacy of the problem.

The next two lemmas are needed in the proof of theorem(2.2.1) and can be found in Studden and Van Arman (1969,pages 1559 and 1560).

Lemma 2.2.1. Let A be a matrix of the form

$$A = \begin{bmatrix} A_{r_0} & A_{r_1} & \cdots & A_{r_k} \\ A_{r_1} & A_{r_1} & \cdots & A_{r_k} \\ \vdots & \vdots & & \vdots \\ A_{r_k} & A_{r_k} & & A_{r_k} \end{bmatrix}.$$

Then $A \geq 0$ (non-negative definite) if and only if

$$0 \neq A_{r_0} \geq A_{r_1} \geq \dots \geq A_{r_k} \geq 0;$$

$A > 0$ (positive definite) if and only if we have that one of the inequalities above is strict, i.e., $A_{r_i} > A_{r_{i+1}}$ for some i, $i = 0,\dots,k-1$.

Proof: We need only notice that

$$x'Ax = \sum_{ij} x_i x_j a_{ij} = \sum_{i=0}^{k} (A_{r_i} - A_{r_{i+1}}) (x_{r_0} + x_{r_1} + \dots + x_{r_i})^2$$

where $A_{r_{k+1}} = 0$.

Lemma 2.2.2. If $M = (m_{ij})$ is a symmetric non-negative definite matrix and a diagonal element $m_{ii} = 0$ for some i, then $m_{ij} = 0$ and $m_{ji} = 0$ for all j. ($m_{ji} = 0$ since M is symmetric).

The next theorem gives moment conditions for admissibility. It was motivated by and gives a slight generalization of a theorem of Studden and Van Arman (1969, page 1559). All the integrals in the following will be over $[a,b]$ unless specified otherwise.

Theorem 2.2.1. Let $f(x)$ consist of the vector of regression functions

2.2.2. $$f(x) = \begin{cases} 1, x, \dots, x^n \\ (x-\xi_p)_+^{n-k_p}, \dots, (x-\xi_p)_+^{n-\ell_p} & x \in [a,b] \\ p = 1,2,\dots,h;\ n-1 \geq k_p \geq \ell_p \geq 0,\ 0 < \xi_1 < \dots < \xi_h < b \end{cases}$$

and let g(x) consist of the vector of regression functions

$$2.2.3.\quad g(x)=\begin{cases} 1,x,\dots,x^{2n-1} \\ (x-\xi_p)_+^{n-k_p},\dots,(x-\xi_p)_+^{2n-\delta_p} & \text{where } \delta_p=\begin{matrix}\ell_p \text{ when } \ell_p\neq 0 \\ 1 \text{ when } \ell_p=0\end{matrix} \\ p=1,2,\dots,h;\ k_p,\ell_p,h,\xi_p \text{ same as in } f(x). \end{cases}$$

Then $\nu \geq \mu$ (or $M(\nu) \geq M(\mu)$), (ν and μ designs for f(x))

if and only if

$$(1)\qquad \int g(x)d(\nu-\mu) = 0 \quad \text{and}$$

$$(2)\quad 0\neq\int x^{2n}d(\nu-\mu) \geq \int(x-\xi_{r_1})_+^{2n} d(\nu-\mu) \geq\dots\geq\int(x-\xi_{r_m})_+^{2n} d(\nu-\mu) \geq 0$$

where r_j, $j = 1,\dots,m$, is the ordered set of i's for which $\ell_i=0$, $0\leq m\leq h$.

Proof: We prove sufficiency first.

Let $M = M(\nu)-M(\mu)$. Since ν and μ are both probability measures, the first row and column of M has zero elements. This gives the following:

$$(a)\qquad \int x^i d(\nu-\mu) = 0 \quad i = 1,2,\dots,n \quad \text{and}$$

$$(b)\qquad \int (x-\xi_p)_+^j d(\nu-\mu) = 0 \quad j = n-k_p,\dots,n-\ell_p;\ p=1,\dots,h$$

by lemma (2.2.2).

From (a) with i=2, the second row and column have all zeros. Continuing in this manner, we obtain

$$\int x^i d(\nu-\mu) = 0 \qquad i = 0,\ldots, 2n-1 \text{ and}$$

$$\int x^i (x-\xi_p)_+^j \, d(\nu-\mu) = 0 \qquad i = 0,1,\ldots,n$$

$$j = n-k_p,\ldots,n-\delta_p$$

$$p = 1,2,\ldots,h.$$

Note that $\int x^n (x-\xi_p)_+^j \, d(\nu-\mu)=0$, $j=n-k_p,\ldots,n-\delta_p$, since the column with diagonal term $\int (x-\xi_p)_+^{2j} \, d(\nu-\mu)=0$, $j=n-k_p,\ldots,n-\delta_p$. Now for $r \leq n$ and any $p=1,2,\ldots,h$, we have that

$$\int (x-\xi_p)_+^{n-\delta_p+r} d(\nu-\mu) = \int (x-\xi_p)_+^r (x-\xi_p)_+^{n-\delta_p} d(\nu-\mu) =$$

$$\sum_{i=0}^{r} a_i \int x^i (x-\xi_p)_+^{n-\delta_p} \, d(\nu-\mu) = 0.$$

Therefore $\int g(x) \, d(\nu-\mu) = 0$,

which means condition (1) holds. At this point we know that M has all diagonal elements = 0 except possibly the elements

$$\int x^{2n} \, d(\nu-\mu) \qquad \text{and}$$

$$\int (x-\xi_p)_+^{2n} \, d(\nu-\mu) \quad \text{when } \ell_p = 0,\ p = 1,\ldots,h.$$

Let r_1 = smallest p for which $\ell_p = 0$,

r_2 = next smallest p for which $\ell_p = 0$,

⋮

r_m = largest p for which $\ell_p = 0$,

and define A_{r_i} as

$$A_{r_i} = \int (x-\xi_{r_i})_+^{2n} \, d(\nu-\mu), \quad i = 1,\ldots,m,$$

$$A_{r_0} = \int x^{2n} \, d(\nu-\mu), \; i = 0.$$

The element corresponding to the r_s row and r_t column, $s<t$, is

$$\int (x-\xi_{r_s})_+^n (x-\xi_{r_t})_+^n \, d(\nu-\mu) = \int x^n (x-\xi_{r_t})_+^n \, d(\nu-\mu) =$$

$\int (x-\xi_{r_t})_+^{2n} \, d(\nu-\mu) = A_{r_t}$. So the conditions of lemma (2.2.1) are satisfied and this implies condition (2).

In order to prove necessity, we note that if conditions (1) and (2) hold, and $M=M(\nu)-M(\mu)$, we see that $M \geq 0$ by lemma (2.2.1).

The following lemma restricts the class of admissible designs for (1.2.1) to the class of admissible designs for (1.1.6).

<u>Lemma 2.2.3.</u> If μ is admissible for

$$b(x) = \begin{cases} 1, x, \ldots, x^n \\ (x-\xi_i)_+^{n-k_i}, \ldots, (x-\xi_i)_+^{n-\ell_i}; \quad i = 1,2,\ldots,h; \; x\varepsilon \, [a,b] \\ n-1 \geq k_i \geq \ell_i \geq 0; \quad a < \xi_1 < \ldots < \xi_h < b \end{cases}$$

then μ is admissible for

2.2.4. $$f(x) = \begin{cases} 1, x, \ldots, x^n \\ (x-\xi_i)_+^{n-k_i}, \ldots, (x-\xi_i)_+^{n} \quad i = 1,2,\ldots,h; \; x\varepsilon \, [a,b] \\ k_i, h, \xi_i, a, b \text{ same as in (2.2.4).} \end{cases}$$

Proof: Assume μ is admissible $b(x)$ and inadmissible $f(x)$. Since μ is inadmissible $f(x)$, there exists a ν admissible $f(x)$ such that $M(\nu) \geq M(\mu)$. See lemma (2.1.3). Let $M'(\nu)$ and $M'(\mu)$ represent the submatrices of $M(\nu)$ and $M(\mu)$ corresponding to $b(x)$. Since μ is admissible $b(x)$, we have that $M'(\nu) \equiv M'(\mu)$. By theorem (2.2.1) this implies that $\int x^{2n}\, d\nu = \int x^{2n}\, d\mu$ which in turn implies that $M(\nu) \equiv M(\mu)$, the desired contradiction.

This lemma tells us that if μ is inadmissible for $f(x)$ then μ is also inadmissible for $b(x)$. In order to completely classify the admissible designs for $b(x)$, we need only list those designs that are (i) admissible $f(x)$ and (ii) inadmissible $b(x)$ since the admissible designs for $f(x)$ are given by (1.2.3). This task appears somewhat formidable as the remainder of the section is devoted to the solution for several general cases.

Lemma 2.2.4. Given a design μ such that

(1) $S(\mu)$ has $\leq n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]$ points on the open interval $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$, $\ell = 0,1,\ldots,h$, we can always add a set B of points in $[a,b]$ such that

(2) $B \cap S(\mu) = \phi$

and

(3) $S(\mu) \cup B$ has $\leq n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]$ points on the open interval $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,.,\ldots,h-\ell$, $\ell = 0,1,\ldots,h$, where equality holds for $\ell=h$ when $i=0$.

Proof: The proof will be by induction on the number of knots. Let μ be a design satisfying (1) for which the number of knots $h=1$.

If there were $\leq \left[\frac{n+k_1+1}{2}\right]$ points in $[\xi_1,b)$, we would add distinct points to $[\xi_1,b)$ until equality would hold. If $k_1=n-1$, then one of the points in $[\xi_1,b)$ either contributed or present in $S(\mu)$ would be ξ_1. If in the remaining piece (a,ξ_1) there were less than $n-1$ points, we would add distinct points until there were exactly $n-1$ points in (a,ξ_1). Let B be the set of points added. It is easily seen that (2) and (3) hold.

If there were $r > \left[\frac{n+k_1+1}{2}\right]$ points in $[\xi_1,b)$, we would let $s=r - \left[\frac{n+k_1+1}{2}\right]$ and note that (1) requires that we have $\leq n-1-s$ points in (a,ξ_1). If there were $\leq (n-1-s)$ points in (a,ξ_1), we would add distinct points until equality held. Let B be the set of points added. We have now shown (2) and (3) for the case of one knot.

Let μ be a design for which the number of knots $h=m+1$. If there were $\leq \left[\frac{n+k_{m+1}+1}{2}\right]$ points in $[\xi_{m+1},b)$, we would use the induction hypothesis to require $S(\mu) \cup B'$ to satisfy (2) and (3) for the interval (a,ξ_{m+1}) and add necessary points to the interval $[\xi_{m+1},b)$ to have the interval total $= \left[\frac{n+k_{m+1}+1}{2}\right]$. If $k_{m+1}=n-1$, then ξ_{m+1} would be a counted point. Let B be the set of all points added. $B' \subset B$ and again (2) and (3) hold.

If there were $r > \left[\frac{n+k_{m+1}+1}{2}\right]$ points on (ξ_{m+1}, b), we would use the induction hypothesis to require $S(\mu) \cup B'$ to satisfy (2) and (3) on (a, ξ_{m+1}). Let $s = r - \left[\frac{n+k_{m+1}+1}{2}\right]$ and note that B' has at least s points, otherwise assumption (1) would be contradicted. We now remove the largest s points of B' and call the remaining set B. All that remains is to check the requirement (2) on subintervals that contain $[\xi_{m+1}, b)$. Let (ξ_t, b) be any interval that contains points in B. Since (ξ_t, ξ_{m+1}) has

$$\leq n-1 + \sum_{j=t+1}^{m} \left[\frac{n+k_j+1}{2}\right] - s$$

points, we have that (ξ_t, b) has

$$\leq n-1 + \sum_{j=t+1}^{m+1} \left[\frac{n+k_j+1}{2}\right] \text{ points.}$$

If (ξ_t, b) does not contain points of B, the subinterval requirement is a part of our assumption (1).

This completes the discussion since (2) and (3) hold.

Remark: We can delete any number of points from B and condition (1) would hold for $S(\mu) \cup$ (B deleted).

In the next two lemmas we develop properties of spectra that when used with the preceding lemmas and the moment theorem will give a large class of admissible designs. Essentially we will be able to classify as admissible those designs for which the moments $\int g(x)d\mu$ prohibit the existance of a ν admissible f(x) such that $\nu \geq \mu$. The results will be stated in theorems (2.2.2) and (2.2.3).

Lemma 2.2.5. If a design μ is such that

$$S(\mu) \text{ has } \leq n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]$$

points on the open interval $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$,

$\ell = 0,1,\ldots,h$, where equality holds for $\ell=h$ when $i=0$ and p is such that $k_p=n-1$, $1 \leq p \leq h$, then $\xi_p \varepsilon S(\mu)$.

Proof: The number of points in (a,ξ_p) is $\leq n-1 + \sum_{j=1}^{p-1} \left[\frac{n+k_j+1}{2}\right]$

The number of points in (ξ_p,b) is $\leq n-1 + \sum_{j=p+1}^{h} \left[\frac{n+k_j+1}{2}\right]$. The number of points in $(a,\xi_p) \cup (\xi_p,b)$ is

$$\leq 2(n-1)+ \sum_{j=1}^{h} \left[\frac{n+k_j+1}{2}\right] - \left[\frac{n+k_p+1}{2}\right] = n-2 + \sum_{j=1}^{h} \left[\frac{n+k_j+1}{2}\right]$$

since $\left[\frac{n+k_p+1}{2}\right] = n$. The number of points in $(a,b)-[(a,\xi_p)\cup(\xi_p,b)]=1$. This implies that $\xi_p \varepsilon S(\mu)$.

Lemma 2.2.6. Let $f^1(x)$ consist of the vector of regression functions

$$f^1(x) = \begin{cases} 1,x,\ldots,x^n \\ (x-\xi_i)_+^{n-k_i},\ldots,(x-\xi_i)_+^{n} \quad i = 1,2,\ldots,h \\ \text{where for each } i,\ k_i \text{ is such that} \\ n+k_i \text{ is even or } k_i=n-1. \end{cases}$$

Let $g^1(x)$ consist of the vector of regression functions

$$g^1(x) = \begin{cases} 1, x, \ldots, x^{2n-1} \\ (x-\xi_i)_+^{n-k_i}, \ldots, (x-\xi_i)_+^{2n-1} \quad i = 1,\ldots,h \\ \text{where the } \xi_i \text{ and } k_i \text{ are the same for } f^1(x) \text{ above.} \end{cases}$$

If μ and ν are admissible designs relative to $f^1(x)$ with supports $S(\mu)$ and $S(\nu)$, then any design relative to $g^1(x)$ with support $S(\mu) \cup S(\nu)$ is admissible $g(x)$. (In applying this lemma, we are more concerned with the placement of points in their spectra than with admissibility with respect to $g^1(x)$.)

Proof: $S(\mu)$ and $S(\nu)$ each have

$$\leq n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right] \text{ points in the interval}$$

$(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$; $\ell = 0,1,\ldots,h$, where we may assume equality holds for $\ell=h$ when $i=0$ for $S'(\mu)$ and $S'(\nu)$. $S'(\mu) \equiv S(\mu) \cup B$ from lemma (2.2.4) and $S'(\nu)$ is defined similarly. An admissible design for $g^1(x)$ would have

2.2.5. $$\leq 2n-2 + \sum_{j=i+1}^{i+\ell} \left[\frac{2n-1+n+k+1}{2}\right] \text{ points in the interval}$$

$(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$; $\ell = 0,1.,,,.h$. $S'(\mu) \cup S'(\nu)$ has

2.2.6. $$\leq 2(n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]) - r_{i\ell} \text{ distinct points in}$$

$(\xi_i, \xi_{i+\ell+1})$. $r_{i\ell}$ is the number of indexes j such that $i+1 \leq j \leq i+\ell$ for which $k_j=n-1$. To see this we note that by lemma (2.2.5), $\xi_j \varepsilon S'(\mu)$ and $\xi_j \varepsilon S'(\nu)$ when $k_j=n-1$. The subtraction of $r_{i\ell}$

eliminates the counting of ξ_j twice in $S(\mu) \cup S(\nu)$. It is easily seen that (2.2.6) $= 2n-2 + \sum_{j=i+1}^{i+\ell} (n+k_j)$ with the restrictions on k_i. Since (2.2.6) $\le 2n-2 + \sum_{j=i+1}^{i+\ell} \left(n + \left[\frac{n+k_i}{2}\right]\right)$=(2.2.5), we have that $S(\mu) \cup S(\nu)$ is admissible $g^1(x)$.

Theorem 2.2.2. Let b(x) consist of the vector of regression functions

2.2.7. $$b(x) = \begin{cases} 1, x, \ldots, x^n \\ (x-\xi_i)_+^{n-k_i}, \ldots, (x-\xi_i)_+^{n-\ell_i} & \ell_i = 0 \text{ or } 1 \text{ for each } i, \\ i = 1,2,\ldots,h;\ n-1 \ge k_i \ge \ell_i \ge 0 & x \in [a,b]. \end{cases}$$

A design μ is admissible b(x) if $S(\mu)$ has $\le n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_i}{2}\right]$ points on $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$, $\ell = 0,1,\ldots,h$.

Proof: Assume μ is inadmissible b(x). Then after consideration of lemmas (2.2.3) and (2.1.3), there exists a ν admissible with respect to f(x) (as defined in (2.2.5) with the same k_i, ξ_i, h, a and b as in b(x) above) such that $\nu \ge \mu$.

Now $S(\nu)$ has $\le n-1 + \sum_{j=i+1}^{i+\ell} \left[\frac{n+k_j+1}{2}\right]$ points on $(\xi_i, \xi_{i+\ell+1})$ for $i = 0,\ldots,h-\ell$; $\ell = 0,1,\ldots,h$. And $S(\nu) \cup S(\mu)$ has

2.2.8. $$\le 2\left(n-1 + \sum_{\substack{j\in[i+1,i+\ell] \\ \text{such that} \\ n+k_j \text{ is even}}} (n-k_j)\right)$$

(continued)

$$+ \frac{1}{2} \sum_{\substack{j\varepsilon[i+1,i+\ell] \\ \text{such that} \\ n+k_j \text{ is odd}}} (n+k_j-1)$$

$$+ \frac{1}{2} \sum_{\substack{j\varepsilon[i+1,i+\ell] \\ \text{such that} \\ n+k_j \text{ is odd}}} (n+k_j+1)$$

$$= 2n-2 + \sum_{j=i+1}^{i+\ell} (n+k_j)$$

points on $(\xi_i,\xi_{i+\ell+1})$ for $i = 0,1,\ldots,h-\ell$; $\ell = 0,1,\ldots,h$.

Now $S(\mu) \cup S(\nu)$ is admissible with respect to

$$g(x)= \begin{cases} 1,x,\ldots,x^{2n-1} \\ (x-\xi_i)_+^{n-k_i},\ldots,(x-\xi_i)_+^{2n-1} \qquad i = 1,\ldots,h \\ n,\xi_i,\ k_i \text{ the same as in } b(x) \text{ above,} \end{cases}$$

by the previous lemma. Without loss of generality, we may assume the equality holds in (2.2.8) for ℓ=h when i=0 by lemma (2.2.4). Note that the exact number of functions in g(x) is $2n + \sum_{j=1}^{h} (n+k_j)$.

Since $\nu \geq \mu$ we have by theorem (2.2.1) that $\int g(x)d(\nu-\mu)=0$. This can be written as

$$M'(\overline{t},\overline{g})\overline{\nu} = M'(\overline{t},\overline{g})\overline{\mu}$$

where $\nu(t_p)=\nu_p$, $\mu(t_p)=\mu_p$ are the weights assigned to the vector $\overline{t}$ of the $m=2n + \sum_{j=1}^{h} (n+k_j)$ ordered points of $S(\mu) \cup S(\nu) \cup \{a\} \cup \{b\}$.

$M'(\overline{t},\overline{g})$ is the transpose of the matrix $M(\overline{t},\overline{f})$ given in definition (2.1.1). $M'(\overline{t},\overline{f})$ is nonsingular by lemma (2.1.5) since

$$t_{\gamma_i} < \xi_i < t_{2n+1+\gamma_{i-1}}$$

where

$$\gamma_i = \sum_{j=1}^{i} (n+k_i).$$

$M(\overline{t},\overline{g})$ being invertable implies $\nu\equiv\mu$, and we have the desired contradiction.

Example 2.2.1. Let $b_1(x)$ consist of the vector of regression functions

$$b_1(x) = \begin{cases} 1,x,x^2,x^3,x^4 \\ x_+^3 & x\varepsilon\ [-1,1]. \end{cases}$$

The following designs are admissible. (We classify them by their spectra.)

(1) The points {-1} and {1} with three points in (-1,0) and two in (0,1).

(2) The points {-1}, {1} and {0} with two points in both (-1,0) and (0,1).

It is possible to add the function x_+^4 to those in $b_1(x)$ and place an extra point in (0,1) in (1) and one point to either (-1,0) or (0,1) in (2) and retain admissibility.

Example 2.2.2. Let $b_2(x)$ consist of the vector of regression functions

$$b_2(x) = \begin{cases} 1, x, x^2, x^3, x^4 \\ (x-1)_+^2, (x-1)_+^3, (x-2)_+^4 \end{cases} \quad x \varepsilon [0,3].$$

The following designs are admissible:

(1) The points {0} and {3} with three points in both (0,1) and (1,2) and two points in (2,3).

(2) The points {0} and {3} with three points in both (0,1) and (2,3) with two points in (1,2).

If one adds the function $(x-1)_+^4$ to those in $b_2(x)$, the above designs are admissible, and if any point is added in either case the designs would be inadmissible.

The following theorem is closely related to theorem (2.2.2) but does describe some additional admissible designs.

Theorem 2.2.3. Let $b_1(x)$ consist of the vector of regression functions in (2.2.7) with the restriction that $n+k_i$ is even or $k_i=n-1$ for each $i = 1,2,...,h$. If $S(\mu)$ has $\leq n-1 + \sum_{j=i+1}^{i+p} \left[\frac{n+k_i+1}{2}\right]$ points in $(\xi_i, \xi_{i+p+1}) i = 0,1,...,h-p$; $p = 0,1,...,h_i$, then μ is admissible $b_1(x)$.

Proof: Note that $n+k_i$ is even if and only if $n-k_i$ is even. The "only if" part follows from lemma (2.2.3). The "if" part follows that of theorem (2.2.2) with some modification. We would have the ν and μ with similar assumptions and notice that (2.2.6) for this theorem equals

$2n-2+\sum_{j=i+1}^{i+\ell}(n+k_j)$ which is the case in theorem (2.2.2) for $S(\mu)\cup S(\nu)$. The remainder of the proof follows that of theorem (2.2.2) word for word.

The following example of an admissible design is covered by theorem (2.2.3) and not by theorem (2.2.2).

Example 2.2.3. Let $b_3(x)$ consist of the vector of regression functions

$$b_3(x) = \begin{cases} 1,x,x^2,x^3,x^4 \\ x_+,x_+^2,x_+^3 & x\varepsilon\ [-1,1]. \end{cases}$$

The following design is admissible: the points {-1}, {0} and {1} with three points in both (-1,0) and (0,1).

Let $\varphi(x)$ denote the set of functions

$$2.2.9.\quad \begin{cases} 1,x,\dots,x^{2n} \\ (x-\xi_i)_+^{n-k_i},\dots,(x-\xi_i)_+^{2n-1} & i=1,\dots,h \\ \xi_i \text{ and } k_i \text{ same as in (2.2.7)}, & x\varepsilon\ [a,b], \end{cases}$$

and let

$$\varphi_0(x)=1,\ \varphi_1(x)=x,\dots,\ \varphi_{2n}(x)=x^{2n},\ \varphi_{2n+1}(x)=(x-\xi_i)_+^{n-k_1},\dots,$$

$$\varphi_m(x)=(x-\xi_h)_+^{2n-1};\ \text{where } m=2n+\sum_{i=1}^{h}(n+k_i).$$

Let

$$\mathcal{M}=\{\overline{c}=(c_1,\dots,c_m)\mid c_t=\int_a^b \phi_t(x)\,d\mu(x),\mu\varepsilon\,\mathcal{P},t=1,\dots,m\}$$

where

$\mathcal{P}$ is the set of probability measures on [a,b]. $\mathcal{M}$ is a closed convex set in m-space since the functions in $\varphi(x)$ are continuous and defined on a compact space. Theorem (2.2.1) states that a design μ is admissible $b(x)$ ((2.2.7) with $\ell_i=1$ for all i) if and only if, for fixed c_t, $t = 1,\ldots,m$, $t\neq 2n$, μ maximizes

$$c_{2n} = \int_a^b x^{2n} \, d\,\nu(x)$$

for all probability measures ν defined on [a,b] with

$$c_t = \int_a^b \phi_t(x)d\,\mu(x) = \int_a^b \phi_t(x)d\,\nu(x) \quad \text{for all } t\neq 2n.$$

Roughly speaking, μ is admissible if and only if it corresponds to an "upper" boundary point of $\mathcal{M}$. Since $\mathcal{M}$ is closed and convex, there must be a nontrivial supporting hyperplane at any boundary point of $\mathcal{M}$.

Lemma 2.2.7. Any admissible design μ for $b(x)$ ((2.2.7) with $\ell_i=1$ for all i) has an associated nontrivial polynomial $p(x)$ in the $\varphi(x)$ (2.2.9) such that: (1) $P(x)=0$ for $x\varepsilon\, S(\mu)$,

(2) $P(x)\geq 0$ for $x\varepsilon\, [a,b]$,

and

(3) the coefficient of x^{2n} in $P(x)$ is ≤ 0.

Proof: Let c^0 be the point $(c_1^0,\ldots,c_m^0)$ in $\mathcal{M}$ where

$$c_t^0 = \int_a^b \phi_t(x)d\,\mu(x) \text{ for } t = 1,\ldots,m.$$

In constructing a supporting hyperplane at c^0 there exists real constants $\{a_t\}_{t=0}^{m}$, not all zero, such that

$$\sum_{t=1}^{m} a_t c_t + a_0 \geq 0 \qquad \text{for all } c \varepsilon \mathcal{M}$$

2.2.10. and

$$\sum_{t=1}^{m} a_t c_t^0 + a_0 = 0.$$

We have that

$$\sum_{t=1}^{m} a_t c_t + a_0 = \sum_{t=0}^{m} a_t \int_a^b \phi_t(x) d\mu(x) = \int_a^b \left(\sum_{t=0}^{m} a_t \phi_t(x) \right) d\mu(x).$$

Let $P(x) = \sum_{t=0}^{m} a_t \phi_t(x)$. Note that $P(x) \geq 0$ for $x \varepsilon [a,b]$ and thus $P(x)=0$ for $x \varepsilon S(\mu)$. The point $c_\lambda = (c_1^0, \ldots, c_{2n-1}^0, c_{2n}^0 + \lambda, c_{2n+1}^0, \ldots, c_m^0)$ for all $\lambda > 0$ lies in the half space complementary to that of (2.2.10), so that

$$\sum_{t=0}^{m} a_t c_t^0 + \lambda a_{2n} \leq 0 \qquad \text{for all } \lambda > 0. \text{ This requires}$$

that $a_{2n} \leq 0$.

A lemma which is a partial converse of the preceding follows.

Lemma 2.2.8. A design μ is admissible for $b(x)$ ((2.2.7) with $\ell_i=1$ for all i) if there exists a nontrivial polynomial $P(x)$ in the $\varphi(x)$ such that:

(1) $P(x) \geq 0$ for $x \varepsilon [a,b]$,

(2) $P(x) = 0$ for $x \varepsilon S(\mu)$,

and

(3) the coefficient of x^{2n} in $P(x)$ is negative.

Proof: Let ν be a probability measure on $[a,b]$ such that $M(\nu) \geq M(\mu)$. By theorem (2.2.1) we have that

$$\int x^{2n}\, d(\nu-\mu) > 0.$$

Also by theorem (2.2.1) we have that

$$\int P(x)d(\nu-\mu) = \int a_{2n}x^{2n}\, d(\nu-\mu)$$

where a_{2n} is the coefficient of x^{2n} in $P(x)$.

$\int P(x)\ d(\nu-\mu) = \int P(x)d\nu \geq 0$ by conditions (1) and (2) of the lemmas.

Combining the above inequalities, we have

$$\int a_{2n}x^{2n}d(\nu-\mu) = \int P(x)d(\nu-\mu) = \int P(x)d\nu \geq 0.$$

This implies that

$$\int x^{2n}\, d(\nu-\mu) \leq 0$$

by condition (3). This is the desired contradiction.

We will use the preceding two lemmas to construct some examples for the regression functions

2.2.11. $$\begin{cases} 1, x, x^2, x^3, x^4 \\ \\ x_+^3 & x\ \ [-1,1]. \end{cases}$$

The examples will show that the converse of theorem (2.2.2) does not hold. A more complete discussion will follow.

Assume we have a polynomial in the form

2.2.12. $$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_8 x^8.$$

Let $a_{k\ell}$ denote the sum of all possible products of k prescribed roots taken ℓ at a time. If we have a polynomial of degree n with n real

$$\int x^{2n}\, d(\nu-\mu) > 0.$$

Also by theorem (2.2.1) we have that

$$\int P(x) d(\nu-\mu) = \int a_{2n} x^{2n}\, d(\nu-\mu)$$

where a_{2n} is the coefficient of x^{2n} in $P(x)$.

$\int P(x)\, d(\nu-\mu) = \int P(x) d\nu \geq 0$ by conditions (1) and (2) of the lemmas.

Combining the above inequalities, we have

$$\int a_{2n} x^{2n} d(\nu-\mu) = \int P(x) d(\nu-\mu) = \int P(x) d\nu \geq 0.$$

This implies that

$$\int x^{2n}\, d(\nu-\mu) \leq 0$$

by condition (3). This is the desired contradiction.

We will use the preceding two lemmas to construct some examples for the regression functions

2.2.11. $$\begin{cases} 1, x, x^2, x^3, x^4 \\ \\ x_+^3 & x\ [-1,1]. \end{cases}$$

The examples will show that the converse of theorem (2.2.2) does not hold. A more complete discussion will follow.

Assume we have a polynomial in the form

2.2.12. $$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_8 x^8.$$

Let $a_{k\ell}$ denote the sum of all possible products of k prescribed roots taken ℓ at a time. If we have a polynomial of degree n with n real roots, then $a_{n\ell}$ is the ℓth symmetric function of the roots. Let P(x) represent a polynomial in the functions

2.2.13. $(1,x,\ldots,x^8,x_+^3,x_+^4,\ldots,x_+^7) \quad x\varepsilon\ [-1,1]$

and denote by $P_1(x)$ the form of P(x) on [-1,0] but extended to the real line. Let $P_2(x)$ denote the similar polynomial determined by P(x) on [0,1]. $P_1(x)$ and $P_2(x)$ have the form (2.2.12). Let $a^1_{k\ell}$ denote the appropriate sum and product of the roots of the polynomial $P_1(x)$ and $a^2_{k\ell}$ denote them for $P_2(x)$.

Consider P(x), a polynomial in (2.2.13), which has the associated polynomials

$$P_1(x) = -1(x+1)(x+3/4)^2(x+2/4)^2(x+1/4)^2(x-r)$$

and

$$P_2(x) = -1(x-1)(x-3/4)^2(x-2/4)^2(x-1/4)^2(x-s).$$

For P(x) to be of the correct form $\sum_{j=0}^{8} \alpha_j x^j + \sum_{j=3}^{7} \beta_i x_+^j$,

we must have the coefficients of 1, x and x^2 identical in $P_1(x)$ and $P_2(x)$. In the following we will be considering the 7 roots of $P_1(x)$ and $P_2(x)$ that exclude r and s. Using the equality of the coefficients of 1, x and x^2 to solve for r and s, we obtain

2.2.14. $$\begin{bmatrix} a^1_{77} & -a^2_{77} \\ a^1_{76} & -a^2_{76} \\ a^1_{75} & a^2_{75} \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ a^2_{77} - a^1_{77} \\ a^2_{76} - a^1_{76} \end{bmatrix}.$$

We will have a solution in r and s of the above system of equations if and only if

$$2.2.15. \quad \det \begin{bmatrix} a^1_{77} & -a^2_{77} & 0 \\ a^1_{76} & -a^2_{76} & a^2_{77} - a^1_{77} \\ a^1_{75} & -a^2_{75} & a^2_{76} - a^1_{76} \end{bmatrix} = 0.$$

For this particular problem we know that the rank of the coefficient matrix in (2.2.14) is 2 and (2.2.15) holds. So the system has a unique solution $r=-s=3/47$. Thus

$$2.2.16. \quad P(x) = \begin{cases} -1(x+1)(x+3/4)^2(x+2/4)^2(x+1/4)^2(x-3/47) & x\varepsilon\ [-1,0] \\ -1(x-1)(x-3/4)^2(x-2/4)^2(x-1/4)^2(x+3/47) & x\varepsilon\ [0,1], \end{cases}$$

which can be written in the form $\sum_{j=0}^{8} a_j x^j + \sum_{i=3}^{7} b_j x_+^j$ with $a_8=-1$. Note that $P(x) \geq 0$ for $x\varepsilon\ [-1,1]$, $P(x)=0$ for $x\varepsilon\ \{-1,-3/4,-1/2,-1/4,1/4,1/2,3/4,1\}$, and $P(x) > 0$ otherwise.

Example 2.2.4. Let $b_4(x)$ consist of the vector of regression functions (2.2.11). If μ is such that $S(\mu)=\{-1,-3/4,-1/2,-1/4,1/4,1/2,3/4,1\}$, then μ is admissible $b_4(x)$ by lemma (2.2.8) and (2.2.16).

Example 2.2.5. Let $b_5(x)$ consist of the vector of regression functions (2.2.11). In this example it is shown that a nontrivial polynomial does not exist for which $P(x) \geq 0$ on $[-1,1]$ and $P(x)=0$ for $x\varepsilon\ \{-1,-3/4,-2/4,-1/4,2/5,3/5,4/5,1\}$. Thus if μ is such that $S(\mu)=\{-1,-3/4,-2/4,-1/4,2/5,3/5,4/5,1\}$, then μ is inadmissible by lemma (2.2.7). First let us assume that the coefficient of x^8 in any nontrivial polynomial $P(x)$ in the functions (2.2.13) is non-zero. Proceeding as in example (2.2.4) we find that

$$\det\begin{bmatrix} a^1_{77} & -a^2_{77} & 0 \\ a^1_{76} & -a^2_{76} & a^2_{77} - a^1_{77} \\ a^1_{75} & -a^2_{75} & a^2_{76} - a^1_{76} \end{bmatrix} = \det\begin{bmatrix} -\ 9/45 & -\ 36/56 & 0 \\ 141/45 & -\ 696/56 & 36/56 - 9/45 \\ -226/44 & -1097/55 & 696/56 - 141/45 \end{bmatrix}$$

is non-zero. Thus there is no solution for r and s and $P(x)$ does not exist. If the coefficient of x^8 is zero, then lemma (2.1.5) implies that any $P(x)$ must be trivial ($\equiv 0$).

Theorem (2.2.2) states for the proceeding two examples that a design is admissible if $S(\mu)$ has at most three points in $(-1,0)$ and two in $(0,1)$. The examples show that a simple counting argument will not give a completely general sufficient condition for admissibility. One must consider the placement of the points in the admissible spectra as well as their number.

Example 2.2.6. Let $b_6(x)$ consist of the vector of regression functions (2.2.11). If μ is such that

$$S(\mu)=\{-1,-x_1,-x_2,-x_3,x_3,x_2,x_1,1\}\ ,\ 0<x_i<1,\ i = 1,2,3,$$

then μ is admissible. This states that for $b_6(x)$ symmetric designs (three points in both $(-1,0)$ and $(0,1)$) are admissible. We proceed again as in example (2.2.4) and note that for the symmetric case

$$\det \begin{bmatrix} a^1_{77} & -a^2_{77} & 0 \\ a^1_{76} & -a^2_{76} & a^2_{77} - a^1_{77} \\ a^1_{75} & -a^2_{75} & a^2_{76} - a^1_{76} \end{bmatrix} = \det \begin{bmatrix} -a & -a & 0 \\ b & -b & 2a \\ -c & -c & 0 \end{bmatrix} = 0.$$

Thus, when solving for r and s, the system (2.2.14) reduces to

$$\begin{bmatrix} -a & -a \\ b & -b \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ 2a \end{bmatrix}.$$

This implies that there is a unique solution with r and s real where $r=-s$ and $r>0$. (a and b are both positive). Thus there exists a polynomial P(x) in the functions (2.2.13) that satisfies the conditions of lemma (2.2.8) so that μ is admissible.

At this point one would like to know if there exists nonsymmetric admissible designs for the regression functions (2.2.11). Let μ be an experimental design such that $S(\mu)=\{-1,-3/4,-2/4,-1/4,1/4,(2-\varepsilon)/4,t,1\}$ where $t\varepsilon\ (0,1)$ and $|\varepsilon|$ is small. Evaluating the determinant (2.2.15) as was done in earlier examples, we find that for this case

$$\begin{aligned} K \det = {} & t^4[10774(2-\varepsilon)^4 + 7104(2-\varepsilon)^3 + 1728(2-\varepsilon)^2] \\ & + t^3[1776(2-\varepsilon)^4 + 576(2-\varepsilon)^3] \\ & + t^2[108(2-\varepsilon)^4 - 25110(2-\varepsilon)^2 - 15984(2-\varepsilon) - 1296] \\ & + t\ [-3996(2-\varepsilon)^2 - 1296] \\ & -81\ (2-\varepsilon)^2, \end{aligned}$$ where K is a positive constant.

If ε is near zero, the coefficients of t^4 and t^3 are positive while that of t^2, t and 1 are negative. If ε is near zero and t=0 the determinant is negative. If ε is near zero and t=1 the determinant is positive. By Descartes' Rule of Signs, there is only one $t^1\varepsilon(0,1)$ for which the above determinant is zero. We are now assured of a solution with r and s real where $r>0$ and $s<0$. This gives rise to a polynomial P(x) satisfying the conditions of lemma (2.2.8) so that μ is admissible. We have the following:

Example 2.2.7. If μ is a design such that $S(\mu)=\{-1,-3/4,-2/4,-1/4,1/4,(2-\varepsilon)/4,t^1,1\}$ for $|\varepsilon|$ small where t^1 is mentioned above, then μ is admissible. If $\varepsilon=0$ then $t^1=3/4$, and this would be a symmetric design. However, with $|\varepsilon|$ small, the example includes many nonsymmetric admissible designs.

Example 2.2.8. If μ is a design such that $S(\mu)=\{-1,-3/4,-2/4,-1/4,2/5,3/5,4/5,1,4/3,5/3\}$, then μ is inadmissible with respect to the regression functions $(1,x,\ldots,x^4,x_+^3,(x-1)_+^3)$. This follows from example (2.2.5) since μ is not sub-admissible on [-1,1]. (See Studden and Van Arman (1969, page 1561).) However, if we consider the polynomial

$$P(x) = 4/81(x-1)_+^3-4/9(x-1)_+^4+13/9(x-1)_+^5-2(x-1)_+^6+(x-1)_+^7,$$

which can be written as

$$(x-1)^3(x-4/3)^2(x-5/3)^2 \quad \text{for } x\varepsilon\ [1,2],$$

we see that it satisfies the conditions of lemma (2.2.7). Thus the converse of lemma (2.2.7) does not hold.

CHAPTER III

POLYNOMIAL MONOSPLINE REGRESSION WITH A SINGLE MULTIPLE KNOT IN THE CENTER

3.1. Introduction with Background Lemmas

In this chapter we are concerned with monospline regression in the form

3.1.1. $$\sum_{i=0}^{n} \theta_i x^i + \sum_{i=n-1-k}^{n-1} \alpha_i x_+^i \quad \text{for } x \varepsilon [-1,1].$$

We will, however, consider different bases. But in each case the regression function could be expressed in the basis of (3.1.1). Lemma (2.1.4) describes the type of function we are considering. The following result due to Elfving (1952) characterizes optimal designs μ for the problem of estimating $(\bar{c},\theta)$. This result is geometric in nature and will be frequently employed throughout this chapter.

Theorem 3.1.1. Let

$$\mathcal{R}_+ = \{\bar{f}(x) = (f_0(x),\ldots,f_n(x)) \mid x \varepsilon \mathcal{X}\},$$

$$\mathcal{R}_- = \{-f(x) \mid x \varepsilon \mathcal{X}],$$

and

$$\mathcal{R} = \text{convex hull of } \mathcal{R}_+ \cup \mathcal{R}_- .$$

A design μ_0 is $\bar{c}$-optimum if and only if there exists a measurable function $\varphi(x)$, satisfying $|\varphi(x)| = 1$, such that

(i) $\int \varphi(x)f(x)\mu_0(dx) = \beta\overline{c}$ for some β

and

(ii) $\beta\overline{c}$ is a boundary point of $\mathfrak{R}$.

Moreover, $\beta\overline{c}$ lies on the boundary of $\mathfrak{R}$ if and only if

$$\beta^2 = v_0^{-1} \quad \text{where } v_0 = \min_\mu V(\overline{c},\mu).$$

A proof of this theorem in the above form may be found in Karlin and Studden (1966, pages 789-791). The following lemma and remark that aid in characterizing the boundary points of $\mathfrak{R}$ aredue to Studden (1968, page 1437).

Every vector $\overline{c}\in\mathfrak{R}$ can be put in the form

3.1.2. $$\overline{c} = \sum_{\nu=1}^{k} \varepsilon_\nu p_\nu f(x_\nu)$$

where $\varepsilon_\nu = \pm 1$, $p_\nu > 0$ and $\sum_{\nu=1}^{k} p_\nu = 1$. The integer k may always be taken to be at most n+2 or at most n+1 if c is a boundary point of $\mathfrak{R}$.

Lemma 3.1.1. A vector $\overline{c}$ of the form (3.1.2) lies on the boundary of $\mathfrak{R}$ if and only if there exists a nontrivial "polynomial" $u(x) = \sum_{\nu=0}^{n} a_\nu f_\nu(x)$ such that $|u(x)| \leq 1$ for $x\varepsilon [-1,1]$, $\varepsilon_\nu u(x_\nu) = 1$, $\nu = 1,2,\ldots,k$, and $\sum_{\nu=0}^{n} a_\nu c_\nu = 1$.

Before the following remark, we recall that

$$V(\overline{c},\mu) = \sup_{\overline{b}} \frac{(\overline{c},\overline{b})}{\int(\overline{b},f(x))^2\mu(dx)}.$$

Remark 3.1.1. For an arbitrary vector $c \neq (0,\ldots,0)$, $\beta\bar{c}$ lies on the boundary of $\mathfrak{R}$ for some $\beta > 0$ and hence

$$\beta c = \sum_{\nu=1}^{n+1} \varepsilon_\nu p_\nu f(x_\nu) \text{ for some } \{\varepsilon_\nu p_\nu\} \text{ and } \{x_\nu\}.$$

If $(\bar{a},f) = \sum_{i=0}^{n} a_i f_i$ denotes the polynomial of lemma (3.1.1), then the minimal value of $V(\bar{c},\mu)$ is $\beta^{-2} = (\sum_{i=0}^{n} a_i c_i)^2 = (\bar{a},\bar{c})^2$ since $(\beta\bar{c},\bar{a}) = 1$.

Moreover,

$$\begin{aligned} \inf_\mu V(c,\mu) &= \inf_\mu \sup_{\bar{b}} (\bar{c},\bar{b})^2[\int(\bar{b},f(x))^2\mu(dx)]^{-1} \\ &\geq \sup_{\bar{b}} \inf_\mu (\bar{c},\bar{b})^2[\ (\bar{b},f(x))^2\ (dx)]^{-1} \\ &\geq (\bar{c},\bar{a})^2. \end{aligned}$$

Since the first and last terms are equal,

$$\inf_\mu \sup_{\bar{b}} (\bar{c},\bar{b})^2[\int(\bar{b},f(x))^2\mu(dx)]^{-1} = \sup_{\bar{b}} \inf_\mu (\bar{c},\bar{b})^2[\int(b,f(x))^2\mu(dx)]^{-1}.$$

Using this last equality, we see that

$$\begin{aligned} \inf_\mu V(\bar{c}_p,\mu) &= \inf_\mu \sup_{\bar{b}} (\bar{c}_p,\bar{b})^2[\int(b,f(x))^2\mu(dx)]^{-1} \\ &= \sup_{\bar{b}} \inf_\mu (\bar{c}_p,\bar{b})^2[\int(b,f(x)^2\mu(dx)]^{-1} \\ &= \sup_{\bar{b}} (\bar{c}_p,\bar{b})^2[\sup_{-1\leq x\leq 1} (\bar{b},f(x))^2]^{-1}. \end{aligned}$$

If we normalize $\bar{b}$ so that $|b_p|=1$, then $(\bar{c}_p,\bar{b})^2=1$ and the last equality becomes

3.1.3. $$\inf_\mu V(\bar{c}_p,\mu) = \sup_{\bar{b} \ni b_p=1} [\sup_{-1\leq x\leq 1} (\bar{b},f(x))^2]^{-1}$$

(continued)

$$= \min_{\bar{a}} \sup_{-1\le x\le 1} [f_p(x)-(\bar{a},\bar{f}_\ell(x))]^2$$

where $\bar{f}_\ell(x)=(f_0(x),\ldots,f_{p-1}(x),f_{p+1}(x),\ldots,f_n(x))$. This last expression suggests that we find the n-vector a* such that $(a^*,\bar{f}_\ell(x))$ is a best approximation to $f_p(x)$ on [-1,1] in the sense of Tchebycheff, i.e., in the uniform norm. Throughout the remainder of this chapter we will be frequently concerned with such best approximations to $f_p(x)$.

<u>Definition 3.1.1</u>. Let $f_0(x)$, $f_1(x),\ldots,f_n(x)$ denote continuous real-valued functions defined on a closed finite interval [a,b]. These functions will be called a Tchebycheff system over [a,b], abbreviated T-system, provided the (n+1)<u>st</u> order determinants

$$\begin{vmatrix} f_0(x_1) & \cdots & f_0(x_{n+1}) \\ f_1(x_1) & \cdots & f_1(x_{n+1}) \\ \vdots & & \vdots \\ f_n(x_1) & \cdots & f_n(x_{n+1}) \end{vmatrix}$$

are of one strict sign for $a \le x_1 < x_2 <\ldots< x_{n+1} \le b$.

<u>Definition 3.1.2</u>. Any linear combination $\sum_{i=0}^{n} a_i f_i(x)$ of the functions $f_0(x),\ldots,f_n(x)$ will be called a polynomial. If any linear combination is identically zero in [a,b], then this linear combination will be called the zero polynomial. A T-system, $f_0(x),\ldots,f_n(x)$, is also such that every polynomial with some non-zero coefficients has at

most n distinct zeros on [a,b]. See Karlin and Studden (1966 a).

Another definition that provides terminology that will be used later is the following:

Definition 3.1.3. A continuous function s(x) is said to alternate r times on an interval [a,b] provided there exists r+1 alternating points $a \leq x_1 < \ldots < x_{r+1} \leq b$ such that

$$s(x_i) = (-1)^i \varepsilon \max_{a \leq x \leq b} |s(x)|$$

for $i = 1,\ldots,r+1$, where $\varepsilon = \pm 1$.

3.2. Optimal Designs for Basis I.

In this section we consider a random variable Y(x) with mean

$$\text{3.2.1.} \qquad \sum_{i=0}^{n} \theta_i x^i + \sum_{i=n-1-k}^{n-1} \beta_i (x^i - 2x_+^i)$$

where $x \varepsilon [-1,1]$ and $n-1-k \geq 1$. We note that in the vector of regressions functions

$$\text{3.2.2.} \quad b(x) = (1,x,\ldots,x^n, x^{n-1-k}-2x_+^{n-1-k},\ldots,x^{n-1}-2x_+^{n-1})$$

each $b_i(x)$ is either odd or even. In order to classify the c_p-optimal designs, our first few lemmas will be concerned with finding a best approximation of $b_p(x)$ by the remaining functions in (3.2.2) in the uniform norm. We will show that there are two monospline polynomials in normalized form that will have the desired properties, one for the even functions in (3.2.2) and the other for the odd. The next lemma concerns one of these polynomials.

Lemma 3.2.1. There exists a unique polynomial $W^1_{n,k}(x)$, a linear combination of the functions in (3.2.2), satisfying

(1) $|W^1_{n,k}(x)| \leq 1$ for $x \in [-1,1]$;

(2) The set $E^1_{n,k} = \{x \,\big|\, |W^1_{n,k}(x)| = 1\}$ contains exactly $n+2\,[\frac{k}{2}]+3$ points and is symmetric about zero;

(3) $W^1_{n,k}(x)$ attains its supremum at each of the points of the set $E^1_{n,k}$ with alternating signs; $W^1_{n,k}(x)$ is of the form (with non-zero coefficients)

$$3.2.3. \quad W^1_{n,k}(x) = \sum_{j=0}^{[\frac{n}{2}]} \beta_{n-2j}\, x^{n-2j} + \sum_{j=0}^{[\frac{k}{2}]} \beta_{n-2j-1}\,(x^{n-2j-1} - 2x_+^{n-2j-1}).$$

(Note that $W^1_{n,k}(x)$ is even or odd as n is even or odd.)

Proof: We consider first the case where n is even. Let V be the linear space spanned by the functions in (3.2.2) excluding x^n. If $g(x) \varepsilon V$ then $g(-x) \varepsilon V$ since each function in (3.2.2) is either odd or even. There exists a best approximation of $b_n(x)=x^n$, say $Q_n(x)$, with respect to V which is even. Meinardus (1967, pages 26 and 27). Thus $b_n(x) - Q_n(x)$ has the form

$$b_n(x) - Q_n(x) = \sum_{j=0}^{\frac{n}{2}} \beta_{2j} x^{2j} + \sum_{j=0}^{[\frac{k}{2}]} \beta_{n-2j-1}(x^{n-2j-1} - 2x_+^{n-2j-1}).$$

For $x \geq 0$

$$b_n(x) - Q_n(x) = \sum_{j=0}^{\frac{n}{2}} \beta_{2j} x^{2j} - \sum_{j=0}^{[\frac{k}{2}]} \beta_{n-2j-1}\, x^{n-2j-1}.$$

Let V_1 be the linear space spanned by the functions

3.2.4. $$\{x^{2j}\}_{j=0}^{\frac{n-2}{2}} \cup \{x^{n-2j-1}\}_{j=0}^{[\frac{k}{2}]} \text{ on } [0,1].$$

There exists a unique best approximation of x^n on [0,1] by functions in V_1 since the functions in (3.2.4) form a T-system. (For discussion of T-systems and verification of these statements see Karlin and Studden (1966 a, page 280).) Let $P(x) = x^n - s(x)$ where $s(x)$ is the minimizing polynomial in V_1. $x^n - s(x)$ alternates $\frac{n}{2} + [\frac{k}{2}] + 1$ times in [0,1] with the endpoints included in the set of $\frac{n}{2} + [\frac{k}{2}] + 2$ extremal points.

Let

$$H^1_{n,k}(x) = \begin{cases} P(x) & x \geq 0 \\ P(-x) & x < 0. \end{cases}$$

It is clear that $H^1_{n,k}(x)$ is a linear combination of the functions

$$\{x^{2j}\}_{j=0}^{\frac{n}{2}} + \{x^{n-2j-1} - 2x_+^{n-2j-1}\}_{j=0}^{[\frac{k}{2}]}.$$

Claim: $H^1_{n,k}(x) = b_n(x) - Q_n(x)$.

We need only check the equivalence on [0,1] since both functions are even. In this case P(x) was the unique minimizing polynomial so that equality holds.

$H^1_{n,k}(x)$ is characterized by the property that there exists $m=(n+2[\frac{k}{2}]+3)$ points $\{t_i\}_{i=1}^{m}$ symmetric about zero where

$-1 = t_1 < t_2 < \ldots < t_m = 1$ such that

$$(-1)^{m-i}(H^1_{n,k}(x_i)) = \max_{-1\leq x\leq 1} |H^1_{n,k}(x)| \ ; \ i = 1,2,\ldots,m.$$

$H^1_{n,k}(x)$ is unique in approximating x^n, for if $K_{n,k}(x)$ were better, we would have that the form of $H^1_{n,k}(x) - K_{n,k}(x)$ would be

3.2.5. $$\sum_{i=0}^{n-1} \alpha_i x_i - \sum_{i=n-1-k}^{n-1} \alpha_i x_+^i$$

and have at least $\frac{n}{2} + [\frac{k}{2}]$ zeros in both $[1,0)$ and $(0,1]$ and at least $n + 2[\frac{k}{2}] + 2$ zeros in $[-1,1]$. Lemma (2.1.5) would imply that $H^1_{n,k} - K_{n,k} \equiv 0$ since the only polynomial with the above zeros in the form (3.2.5) can only be the zero polynomial. Let

$$W^1_{n,k}(x) = \frac{H^1_{n,k}(x)}{||H^1_{n,k}(x)||}$$

where "$||\ ||$" denotes the sup norm for $x \varepsilon [-1,1]$. Assume we have a $K(x)$ satisfying (1), (2) and (3). In order for $K(x)$ to be nontrivial and have at least $(n+k+1)$ zeros, the coefficient of x^n in $K(x)$ must be non-zero. We may normalize $K(x)$ so that the coefficient of x^n is unity. From earlier arguments we know that $||H^1_{n,k}(x)||<||K(x)||$. However, $H^1_{n,k}(x)-K(x)$ would be of the form (3.2.5). A similar argument to that following (3.2.5) implies that $H^1_{n,k}(x)\equiv K(x)$. Thus $W^1_{n,k}(x)$ satisfies (1), (2) and (3) where $E^1_{n,k}=\{t_i\}_{i=1}^m$.

In order to see that the coefficients are non-zero, we note that $W^1_{n,k}(x)$ for $x > 0$ has $\frac{n}{2} + [\frac{k}{2}] + 1$ distinct zeros in $(0,1]$. By Descartes' Rule of Signs, there can be at most $\frac{n}{2} + [\frac{k}{2}] + 1$ zeros in $(0,1]$. This implies that all of the coefficients β_i used in $W^1_{n,k}$ are non-zero.

The proof for n odd follows this identical argument after consideration of the following lemma.

<u>Lemma 3.2.2</u>. Let V be the linear space spanned by the functions

$$\{x^{2j-1}\}_{j=1}^{\frac{n-1}{2}} \cup \{x^{n-1-2j}\}_{j=0}^{[\frac{k}{2}]} \quad \text{on } [0,1]$$

where n is odd and $n-2 \geq k \geq 0$. There exists a unique polynomial $s^*(x)$ of functions in V satisfying

$$\max_{0 \leq x \leq 1} |x^n - s^*(x)| < \max_{0 \leq x \leq 1} |x^n - v(x)|$$

for all $v(x)$ in V. $s^*(x)$ is further characterized by the property that there exists $m = (\frac{n+3}{2} + [\frac{k}{2}])$ points $\{t_i\}_{i=1}^m$ where $(0 < t_1 < t_2 < \ldots < t_m = 1)$ such that

$$(-1)^{m-i}(t_i^n - s^*(t_i)) = \max_{0 \leq x \leq 1} |x^n - s^*(x)|; \quad i = 1,2,\ldots,m.$$

Proof: There exists a best approximation of x^n by a function $b^0(x) \varepsilon V$ in the sense that

$$\max_{0\le x\le 1} |x^n - b^0(x)| \le \max_{0\le x\le 1} |x^n - v(x)|$$

for all $v(x) \varepsilon V$. Meinardus (1967, page 1).

Given the compact set $[\varepsilon,1]$, we denote by $b^{\varepsilon}(x)$, $0<\varepsilon<1$, the unique best approximation of x^n by functions in V defined on $[\varepsilon,1]$. Karlin and Studden (1966, page 280). If $0\le\varepsilon_1<\varepsilon_2<1$, then

$$\max_{x\varepsilon[\varepsilon_2,1]} |x^n - b^{\varepsilon_2}(x)| \le \max_{x\varepsilon[\varepsilon_2,1]} |x^n - b^{\varepsilon_1}(x)| \le \max_{x\varepsilon[\varepsilon_1,1]} |x^n - b^{\varepsilon_1}(x)|.$$

Thus we have for all ε in $(0,1)$

3.2.6. $$\max_{x\varepsilon[\varepsilon,1]} |x^n - b^{\varepsilon}(x)| \le \max_{x\varepsilon[0,1]} |x^n - b^0(x)| = A_1.$$

Choose $\eta>0$ such that $0<\eta<1$ and pick m points $\{x_i\}_{i=1}^m$, $\eta\le x_1<x_2<\ldots<x_m\le 1$. We write the system of equations

$$x_i^n - b^{\varepsilon}(x_i) = c_i \qquad i = 1,\ldots,m \quad 0<\varepsilon\le\eta$$

in the form

3.2.7. $$\overline{M}\,\overline{b}_{\varepsilon} = \overline{c}_{\varepsilon}$$

where $\overline{b}_{\varepsilon}$ is the vector of coefficients of the polynomial $x^n - b^{\varepsilon}(x)$ and $\overline{c}_{\varepsilon}$ is the vector of values c_i in (3.2.7). $\overline{M}$ is an mxm non-singular matrix determined by the $\{x_i\}_{i=1}^m$ and the functions in $V \cup \{x^n\}$. We may write (3.2.7) in the form

$$\overline{b}_{\varepsilon} = \overline{M}^{-1}\,\overline{c}_{\varepsilon}.$$

Since $\overline{M}^{-1}$ is a continuous linear transformation, we have that $||\overline{b}_\varepsilon|| \leq K||\overline{c}_\varepsilon||$ where K is some positive constant and "$||\ ||$" denotes the euclidean norm. Since $|c_i| \leq A_1$ by (3.2.6) for each i, i = 1,...,m, independent of ε, we have that

$$||\overline{b}_\varepsilon|| < K\, A_1,$$

where the positive constant $K\, A_1$ is independent of ε. Thus implies that the individual coefficients in $b^\varepsilon(x)$ are uniformly bounded by $K\, A_1$ for $\varepsilon\varepsilon[0,\eta]$. Therefore there exists an $\varepsilon_3>0$, $0<\varepsilon_3<\eta$, such that

$$\max_{x\varepsilon[0,\varepsilon_3]} |x^n-b^{\varepsilon_3}(x)| \leq A_1.$$

After consideration of (3.2.6) for ε_3 and the minimizing properties of $b^0(x)$, we have that

$$\max_{x\varepsilon[0,1]} |x^n-b^{\varepsilon_3}(x)| = A_1.$$

However, since $b^{\varepsilon_3}(x)$ is the unique best approximation of x^n on $[\varepsilon_3,1]$, it must agree completely with $b^0(x)$. So $b^{\varepsilon_3}(x)\equiv b^0(x)$ for $x\varepsilon[0,1]$. We define

$$b^{\varepsilon_3}\equiv s^*(x)$$

and note that by Karlin and Studden (1966, page 280) and the fact that there must be only (m-1) zeros in the derivative of $x^n-s^*(x)$, the lemma is proven.

We have seen that $W^1_{n,k}(x)$ when properly normalized becomes the minimizing polynomial for x^n. The next lemma shows that it has this minimizing property for all of the functions in (3.2.2) that appear in $W^1_{n,k}(x)$.

Lemma 3.2.3. Among all polynomials f(x) in the functions (3.2.2),

(1) $W^1_{n,k}(x)/\beta_{n-2j}$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where f(x) is any polynomial in (3.2.2) with the coefficient of x^{n-2j} unity for $j = 0,\ldots,[\frac{n}{2}]$, and

(2) $W^1_{n,k}/\beta_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where f(x) is any polynomial in (3.2.2) with the coefficient of $(x^{n-2j-1} - 2x_+^{n-2j-1})$ unity for $j = 0,\ldots,[\frac{k}{2}]$.

Proof: The case for which the coefficient of x^n is unity was done in the two preceding lemmas. Consider next the cases where the coefficient of x^{n-2j} is unity for $j = 0,\ldots, [\frac{n}{2}]$. As before, there is a best approximation of x^{n-2j} by the remaining functions in (3.2.2) which is odd or even as x^{n-2j} is odd or even. Meinardus (1967, pages 26 and 27). Upon dividing this best approximation by its norm, we see that it satisfies the same properties that $W^1_{n,k}(x)$ satisfied in lemma (3.2.1). (i.e. The construction would follow a similar development as that in lemmas (3.2.1) and (3.2.2).) $W^1_{n,k}(x)$ is unique in this respect so (1) is proven. (2) follows the same line of reasoning as (1).

We are now in a position to begin classification of the $\overline{c}_p$-optimal design for the parameters of the functions in (3.2.2) that have the same parity as x^n (i.e. The functions that are even or odd as n is even or odd).

Theorem 3.2.1. The optimal designs for estimating the following parameters in (3.2.1)

3.2.8. $$\begin{cases} \theta_{n-2j} & \text{for } j = 0,\dots,[\frac{n}{2}]; \quad n-2j\neq 0 \\ \beta_{n-2j-1} & \text{for } j = 0,\dots,[\frac{k}{2}]; \\ \text{i.e., parameters for functions of the same parity} \\ \text{as } x^n, \end{cases}$$

have their supports contained in the set $E^1_{n,k}$ and satisfy the following:

(i) When k is odd the optimal design for any parameter in (3.2.8) is unique and is supported by the full set $E^1_{n,k}$;

(ii) When k is even the optimal designs for the parameters in (3.2.8) are not unique and satisfy the following:

(a) The optimal designs are a convex combination of two probability measures $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ (or $\mu^0_{\beta_h}$ and $\mu^1_{\beta_h}$) where $\mu^0_{\theta_\ell}\{+1\}$ $=0(=\mu^0_{\beta_h}\{+1\})$ and $\mu^1_{\theta_\ell}\{-1\}=0$ $(=\mu^1_{\beta_h}\{-1\})$;

(c) In the convex combination described in (a) all the designs other than $\mu^0_{\theta_h}$ and $\mu^1_{\theta_\ell}$ ($\mu^0_{\beta_n}$ and $\mu^1_{\beta_h}$) are supported by the full set $E^1_{n,k}$;

(d) The vectors of weights associated with the optimal designs of (ii) lie on parallel lines;

(iii) The support for the optimal design for θ_0 is $\{0\}$.

Proof: Let $\bar{c}_p=(0,\ldots,0,1,0,\ldots,0)$ where the 1 is in the component of $\theta = (\theta_0,\ldots,\theta_n,\beta_{n-1-k},\ldots,\beta_{n-1})$ corresponding to some one of the parameters in (3.2.8). By (3.1.3) we have that

$$\inf_{\mu} V(\bar{c}_p,\mu) = \min_{\bar{a}} \sup_{-1\leq x\leq 1} [f_p(x)-(\bar{a},\bar{f}_\ell(x))]^2.$$

If we denote by β_p the coefficient of $b_p(x)$ in $W^1_{n,k}(x)$, we have that

$$\inf_{\mu} V(\bar{c}_p,\mu) = (\beta_p)^2$$

by lemma (3.2.3). Suppose that μ^*_p is $\bar{c}_p$-optimal. Then

$$\begin{aligned} V(\bar{c}_p,\mu^*_p) &= \sup_{a} (\bar{c}_p,a)^2[\int(a,f(x))^2\mu^*_p(dx)]^{-1} \\ &\geq (\beta_p)^2[\int(W^1_{n,k}(x))^2\mu^*_p(dx)]^{-1} \\ &\geq (\beta_p)^2. \end{aligned}$$

Since $|W^1_{n,k}(x)|=1$ only for $x\varepsilon\ E^1_{n,k}$, strict inequality holds at the last step unless μ^*_p has its support contained in the set $E^1_{n,k}$.

Assume first that k is odd. To find the $\bar{c}_p$-optimal design, Elfving's theorem (theorem 3.1.1) tells us there is a solution to the system

3.2.9. $$\beta\bar{c}_p = \sum_{\nu=1}^{n+k+2} \varepsilon_\nu P_\nu \bar{b}(x_\nu)$$

for $\beta=|1/\beta_p|>0$ where the $x_\nu\varepsilon\ E^1_{n,k}$, $\sum_{\nu=1}^{n+k+2} P_\nu=1$, $P_\nu\geq 0$ and $\varepsilon_\nu=\pm 1$. The system (3.2.9) describes n+k+2 equations in n+k+2 unknowns. The rank of the system is n+k+2. To see this, we show that if M_1 is the matrix of coefficients of the $\{p_\nu\}$, then M_1 has n+k+2 independent rows. If not, then a nontrivial linear combination of the rows of M_2 would be equal to the zero vector. This would give rise to a polynomial in the functions (3.2.2) which would have zeros at the $\{x_\nu\}$ and be normalized so that the coefficient of x^n is 1. The coefficient of x^n is non-zero since then the only linear combination of the remaining rows would be the trivial one due to the spacing of the $\{x_\nu\}$ and lemma (2.1.5). Let this polynomial be denoted by P(x). Now

$$W^1_{n,k}(x)/\beta_n - P(x) = \sum_{i=0}^{n-1} \alpha_i x^i + \sum_{i=n-1-k}^{n-1} \alpha^1_i (x^i - 2x^i_+)$$

and has at least $[\frac{n+k+2}{2}]-1$ distinct zeros in both [-1,0) and (0,1] with at least n+k+1 distinct zeros in [-1,1]. Lemma (2.1.5) implies that $P(x) \equiv W^1_{n,k}(x)/\beta_n$. This is the desired contradiction. M_1 is invertable so there is a unique solution. Lemma (3.2.4) shows that the support of the $\bar{c}_p$-optimal design in this case is supported by the full set $E^1_{n,k}$.

Assume now that k is even. To find the $\bar{c}_p$-optimal designs, Elfving's theorem (theorem 3.1.1) tells us there is a solution to the system

3.2.10. $$\beta\bar{c}_p = \sum_{\nu=1}^{n+k+3} \varepsilon_\nu p_\nu \bar{b}(x_\nu)$$

for $\beta = |1/\beta_p| > 0$ where the $x_\nu p \varepsilon E^1_{n,k}$ and $\sum_{\nu=1}^{n+k+3} p_\nu = 1$, $p_\nu \geq 0$ and $\varepsilon_\nu = \pm 1$. The system (3.2.10) describes n+k+2 equations in n+k+3 unknowns. The rank of the system is n+k+2. To see this, we show that if M_2 is the matrix of coefficients of the $\{p_\nu\}$, then M_2 has independent rows. If we consider a $(n+k+3) \times (n+k+3)$ matrix M with the n+k+2 rows of M_2 and the row vector $(x_{1+}^n, x_{2+}^n, \ldots, x_{\nu+}^n, \ldots, x_{n+k+3+}^n)$, then the rows of M are independent by lemma (2.1.5). (M is obtainable by elementary row operations from a square matrix whose determinant is nonzero.) This implies that the rows of M_2 are independent. Since the coefficient matrix for each parameter in (3.2.8) is $(\pm 1)\ M_2$, (ii)-(d) is shown. (ii)-(a), (b) and (c) will follow after consideration of lemma (3.2.6). (i,i,i) follows after noting that when n is even, the only solution to (3.2.9) or (3.2.10) is when $p_{([\frac{n+k+2}{2}]+1)}=1$ or $p_{([\frac{n+k+3}{2}]+1)}=1$ respectively.

Lemma 3.2.4. When k is odd, the system of equations (3.2.9) considered in theorem (3.2.1) has as its unique solution a set of p_ν's such that $p_\nu>0$ for all $\nu = 1,\ldots,n+k+2$.

Proof: Assume that n is even. Since M_1 is nonsingular, we can solve for the $\{p_\nu\}$ by Cramer's method. If when estimating the parameters (3.2.8) we have $p_{\nu_0} = 0$ for some $\nu_0 = 1,\ldots,n+k+2$, we are led to a nontrivial linear combination of the functions in (3.2.2) in the form

$$P(x) = \sum_{i=0}^{n} a_i x^i + \sum_{i=n-1-k}^{n-1} b_i (x_+^i - 2x_+^i)$$

where either

$$a_{n-2j} = 0 \quad \text{for some } j = 0,\ldots,[\tfrac{n}{2}], n-2j \neq 0$$

or $$b_{n-2j-1} = 0 \quad \text{for some } j = 0,\ldots,[\tfrac{k}{2}].$$

Let $Q(x)=P(x)+P(-x)$. Then

$$Q(x) = \sum_{i=0}^{n/2} 2a_{2j}x^{2j} + \sum_{i=\frac{n-k-1}{2}}^{\frac{n-2}{2}} 2b_{2j+1}\,(x^{2j+1} - 2x_+^{2j+1})$$

where $a_{2j} = 0$ or $b_{2j+1} = 0$ for the appropriate parameter mentioned above. If $x_{\nu_0} \neq 0$, then $Q(x)$ has at least $(\frac{n+k-1}{2})$ distinct zeros in both $[-1,0)$ and $(0,1]$ and a zero at $x=0$. This implies that $a_0=0$. By Descartes' Rule of Signs, $Q(x)$ can have at most $(\frac{n+k-3}{2})$ distinct zeros in both $[-1,0)$ and $(0,1]$. Thus $Q(x)\equiv 0$ on $(-1,1]$. If $x_{\nu_0} = 0$, then $Q(x)$ would have at least $(\frac{n+k+1}{2})$ distinct zeros in both $[-1,0)$ and $(0,1]$. By Descartes' Rule of Signs, $Q(x)$ can have at most $(\frac{n+k-1}{2})$ distinct zeros in $[-1,0)$ or $(0,1]$. This again implies that $Q(x) \equiv 0$ on $[-1,1]$.

Since $Q(x)\equiv 0$, $P(x)$ can only be of the form

$$P(x) = \sum_{i=0}^{\frac{n-2}{2}} a_{2j+1}\,x^{2j+1} + \sum_{i=\frac{n-k-1}{2}}^{\frac{n-2}{2}} b_{2j}\,(x^{2j} - 2x_+^{2j}).$$

Now $P(x)$ can have at most $(\frac{n+k-1}{2})$ distinct zeros in either $[-1,0)$ or $(0,1]$. However, $P(x)$ must have at least $(\frac{n+k+1}{2})$ distinct zeros in either $[-1,0)$ or $(0,1]$, depending on whether $\nu_0\varepsilon[\frac{n+k+3}{2}, n+k+2]$ or

$\nu_0 \varepsilon [1,\frac{n+k+3}{2}]$, and at least $(\frac{n+k-1}{2})$ distinct zeros in both $[-1,0)$ and $(0,-1]$. This implies that $P(x)\equiv 0$ for $x\varepsilon[-1,1]$ and contradicts the fact that $p_{\nu_0} = 0$. Thus we have that $p_\nu > 0$ for all $\nu = 1,\ldots,n+k+2$.

If n were odd, similar arguments hold for $Q(x)\equiv P(x)-P(-x)$.

<u>Lemma 3.2.5</u>. When k is even there exists a linear combination of the functions in (3.2.2), $K^{\ell}_{n,k}(x)$, such that

3.2.11. $$K^{\ell}_{n,k}(x_\nu)=\begin{cases} W^1_{n,k}(x_\nu) & \text{for } x_\nu = x_\ell \text{ or } x_{n+k+4-\ell} \\ 0 & \text{for } x \neq x_\ell \text{ or } x_{n+k+4-\ell} \quad \ell=1,\ldots,[\frac{n+k+4}{2}] \end{cases}$$

and $K^{\ell}_{n,k}(x)$ is of the form

3.2.12. $$K^{\ell}_{n,k}(x) = \sum_{j=0}^{[\frac{n+1}{2}]} b_{n-2j}x^{n-2j} + \sum_{j=0}^{[\frac{k}{2}]} b_{n-2j-1}(x^{n-2j-1} - 2x_+^{n-2j-1})$$

where all the coefficients b_j are non-zero and have the same sign as the corresponding β_j in (3.2.3).

Proof: Let K be the linear space spanned by the functions $\{x^{n-2j}\}_{j=0}^{[\frac{n}{2}]} \cup \{x^{n-2j-1}\}_{j=0}^{[\frac{k}{2}]}$. Let E be the set of points $E^1_{n,k}$ [0,1]. In E there are exactly $[\frac{n}{2}]+[\frac{k}{2}]+2$ points and in K there are exactly $[\frac{n}{2}]+[\frac{k}{2}]+2$ functions. The functions in K form a T-system on a subinterval containing the points in E. Thus there is a unique linear combination of the functions in K, say $K^{\ell}_{n,k}$, that satisfies (3.2.11) on [0,1]. If n is even we define $K^{\ell}_{n,k}(-x)\equiv K^{\ell}_{n,k}(x)$, and if n is odd we define $K^{\ell}_{n,k}(-x) \equiv -K^{\ell}_{n,k}(x)$ for $x \varepsilon [0,1]$. Thus we have that $K^{\ell}_{n,k}(x)$ is of the form (3.2.12) where all the coefficients b_j are non-zero and have the same sign as the corresponding β_j in (3.2.3), since the coefficients b_j must alternate in sign in the same manner as the β_j in order to have the appropriate number of required zeros.

Lemma 3.2.6. When k is even, the system of equations (3.2.10) has as its solution experimental designs satisfying (ii)-(a), (b) and (c) of theorem (3.2.1).

Proof: Writing the system of equations (3.2.10) in matrix form, we have

$$\text{3.2.13.}\quad \begin{bmatrix} \varepsilon_1 b_0(x_1) & \cdots \varepsilon_{n+k+3}\, b_0(x_{n+k+3}) \\ \vdots & \vdots \\ \varepsilon_1 b_n(x_1) & \cdots \varepsilon_{n+k+3}\, b_n(x_{n+k+3}) \\ \vdots & \vdots \\ \varepsilon_1 b_{n+k+1}(x) & \cdots \varepsilon_{n+k+3}\, b_{n+k+1}(x_{n+k+3}) \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_n \\ \vdots \\ p_{n+k+3} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left|\frac{1}{\beta_p}\right| \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

According to lemma (3.1.1) $\varepsilon_\nu(\pm W^1_{n,k}(x_\nu)) = 1$, where the $\pm$ is determined by the sign of the coefficient of $b_p(x)$, (sgn b_p), in $W^1_{n,k}(x)$. Applying the linear combination of (sgn b_p) $K^1_{n,k}(x)$ from lemma (3.2.5) to the rows of (3.2.13), we have that

$$\text{3.2.14.}\qquad p_1 + p_{n+k+3} = c_1.$$

Now $1 \geq c_1 > 0$ since we know that there is a solution where

$0 \leq p_1 + p_{n+k+3} \leq 1$ and $c_1 \neq 0$, for $1/\beta_p$ would have a positive multiplier.

Consider next the reduced system

$$3.2.14. \quad \begin{bmatrix} \varepsilon_1 b_0(x_1) & \cdots \varepsilon_{n+k+2}\, b_0(x_{n+k+2}) \\ \vdots & \vdots \\ \varepsilon_1 b_{n+k+1}(x_1) & \cdots \varepsilon_{n+k+2}\, b_{n+k+1}(x_{n+k+3}) \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_{n+k+2} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left|\frac{1}{\beta_p}\right| \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Applying in like manner the polynomials $(\operatorname{sgn} b_p)\ K^1_{n,k}(x)$ and $(\operatorname{sgn} b_p)\ W^1_{n,k}(x)$ to the above system, we find that $p_1 = c_1$ and $\sum_{\nu=1}^{n+k+2} p_\nu = 1$. Thus any solution to (3.2.14) is also a solution to (3.2.13). The coefficient matrix in (3.2.14) is nonsingular. To see this, take the linear combination of the rows suggested by $K^1_{n,k}(x)$ and place this in the (n+1)st row. This implies that the determinant of the coefficient matrix is

$$\pm 1 \det \begin{bmatrix} b_0(x_2) & \cdots b_0 & (x_{n+k+2}) \\ \vdots & & \vdots \\ b_{n-1}(x_2) & \cdots b_{n-1} & (x_{n+k+2}) \\ b_{n+1}(x_2) & \cdots b_{n+1} & (x_{n+k+2}) \\ \vdots & & \vdots \\ b_{n+k+1}(x_2) & \cdots & b_{n+k+1}(x_{n+k+2}) \end{bmatrix}.$$

This last determinant is non-zero by lemma (2.1.5). The symmetry in the above arguments would give $p_{n+k+3} = c_1$ and $\sum_{\nu=2}^{n+k+3} p_\nu = 1$ when considering the system of equations determined by the points $\{x_\nu\}_{\nu=2}^{n+k+3}$. It will be shown in theorems (4.2.1) and (4.2.2) that the unique solution of (3.2.14) is a probability measure as well as the unique solution to the symmetric system determined by the points $\{x_\nu\}_{\nu=2}^{n+k+3}$. We will denote these probability measures as $\mu^0_{\theta_p}$ and $\mu^1_{\theta_p}$ respectively, where $\theta_p = (\overline{\beta}, \overline{c}_p)$. Now any solution μ_p (a probability measure) must satisfy

$$\mu_p = q\, \mu^0_{\beta_p} + (1-q)\mu^1_{\theta_p}$$

for some $0 \leq q \leq 1$. To see this, we note that $\mu^0_{\theta_p}$ and $\mu^1_{\theta_p}$, when viewed as vectors of weights on the points $\{x_\nu\}_{\nu=1}^{n+k+3}$, are distinct vectors on a one dimensional set. Also, if q were outside the closed interval [0,1], then either q or (1-q) would be negative. This would imply that either $\mu_p\{-1\}$ or $\mu_p\{+1\}$ would be negative which would be contradictory. This proves (ii)-(a).

In order to prove (ii)-(b), we consider the system (3.2.14) rewritten in matrix form as

$$A\overline{p} = B$$

where A is an $(n+k+2)x(n+k+2)$ matrix and B is an $(n+k+2)x(1)$ matrix. (Note: det A is non-zero by earlier arguments.) Assume n is even.

$$\det A = (-1)^{\frac{n+k+2}{2}} \det \begin{bmatrix} b_0(x_1) & \cdots & b_0(x_{n+2}) \\ \vdots & & \vdots \\ b_{n+k+1}(x_1) & \cdots & b_{n+k+1}(x_{n+2}) \end{bmatrix} =$$

$$(-1)^{\frac{n+k+2}{2}}(-1)^{\frac{n+k}{2}}(-1)^{\frac{(n+k+1)(n+k+2)}{2}} \det \begin{bmatrix} b_0(x_2) & \cdots & b_0(x_{n+k+3}) \\ \vdots & & \vdots \\ b_{n+k+1}(x_2) & \cdots & b_{n+k+1}(x_{n+k+3}) \end{bmatrix} .$$

The last equality is obtained by first multiplying the rows corresponding to odd functions by (-1) and then symmetrically interchanging the columns. (Note: The x_ν are symmetric about zero.) The last equality above becomes

$$\det A = (-1)^{\frac{n+k}{2}} \det \begin{bmatrix} b_0(x_2) & \cdots & b_0(x_{n+k+3}) \\ \vdots & & \vdots \\ b_{n+k+1}(x_2) & \cdots & b_{n+k+1}(x_{n+k+3}) \end{bmatrix} .$$

When solving (3.2.14) by Cramer's method, we are led to the evaluation of

$$\det \begin{bmatrix} \varepsilon_1 b_0(x_1) & \cdots & 0 & \cdots & \varepsilon_{n+k+2}\, b_0(x_{n+k+2}) \\ \vdots & & \vdots & & \vdots \\ & & \left|\frac{1}{\beta_p}\right| & & \\ & & 0 & & \\ \vdots & & \vdots & & \vdots \\ \varepsilon_1 b_{n+k+1}(x_1) & \cdots & 0 & \cdots & \varepsilon_{n+k+2}\, b_{n+k+1}(x_{n+k+2}) \end{bmatrix} ,$$

where $(0,\ldots,0,\ |1/\beta_p|\ ,0,\ldots,0)'$ is the νth column.

This determinant equals

$$(-1)^{\frac{n+k+2}{2}} (-1)^{\nu+\ell}$$

$$\det \begin{bmatrix} b_0(x_1) & \cdots & b_0(x_{\nu-1}) & 0 & b_0(x_{\nu+1}) & \cdots & b_0(x_{n+k+2}) \\ \vdots & & \vdots & \vdots & & & \vdots \\ & & & 0 & & & \\ & & & \left|\frac{1}{\beta_p}\right| & & & \\ & & & 0 & & & \\ \vdots & & \vdots & \vdots & & & \vdots \\ b_{n+k+1}(x_1) & \cdots & b_{n+k+1}(x_{\nu-1}) & 0 & b_{n+k+1}(x_{\nu+1}) & \cdots & b_{n+k+2}(x_{n+k+2}) \end{bmatrix} =$$

3.2.16. $$(-1)^{\nu+\ell}(-1)^{\frac{n+k}{2}} \det$$

$$\begin{bmatrix} b_0(x_2) & \cdots & b_0(x_{n+k+4-(\nu+1)}) & 0 & b_0(x_{n+k+4-(\nu-1)}) & \cdots & b_0(x_{n+k+3}) \\ \vdots & & \vdots & \vdots & & & \vdots \\ & & & \left|\frac{1}{\beta_p}\right| & & & \\ \vdots & & \vdots & \vdots & & & \vdots \\ b_{n+k+1}(x_2) & \cdots & b_{n+k+1}(x_{n+k+4-(\nu-1)}) & 0 & b_{n+k+1}(x_{n+k+4-(\nu-1)}) & \cdots & b_{n+k+1}(x_{n+k+3}) \end{bmatrix}.$$

The equalities were obtained as before with

$$\ell = \begin{cases} 1 \text{ if } (\operatorname{sgn} b_p) \text{ is positive} \\ 0 \text{ if } (\operatorname{sgn} b_p) \text{ is negative.} \end{cases}$$

In solving (3.2.14) by Cramer's method, we would have

3.2.17. $$\mu^0_{\theta_p}\{x_\nu\} = (3.2.16)/(3.2.15) = \mu^1_{\theta_p}\{x_{n+k+4-\nu}\}.$$

The case for n odd is somewhat more involved but follows similar arguments. This proves (ii)-(b).

In order to show (ii)-(c), we assume that

$$\mu_q\{x_\nu\}=q\,\mu^0_{\theta_p}\{x_\nu\} + (1-q)\,\mu^1_{\theta_p}\{x_\nu\} = 0$$

for some q such that $0<q<1$. This implies that $\mu^0_{\theta_p}\{x_\nu\}=\mu^1_{\theta_p}\{x_\nu\}=0$, and therefore, by (3.2.17) that $\mu^0_{\theta_p}\{x_{n+k+4-\nu}\}=0$. This implies that $\mu_q\{x_{n+k+4-\nu}\}=0$. However, after applying the linear combination suggested by $K^\nu_{n,k}(x)$ of lemma (3.2.5) to (3.2.13), we have that $\mu_q\{x_\nu\} + \mu_q\{x_{n+k+4-\nu}\}=c_\nu>0$. This is the desired contradiction. When $\nu_0=(\frac{n+k+4}{2})$, $x_{\nu_0}=0$ and $\mu^0_{\theta_p}\{0\} = \mu^1_{\theta_p}\{0\}=c_{\nu_0}>0$.

<u>Example 3.2.1</u>. Consider a random variable Y(x) with mean $E\,Y(x) = \theta_0+\theta_1 x+\theta_2 x^2 + \beta_1(x-2x_+)$ where $x\varepsilon[-1,1]$. For this example $W^1_{2,0}(x)=1+8x+8x^2+8(x-2x_+)$ and $E^1_{2,0}= \{-1,-1/2,0,1/2,1\}$. The optimal designs for estimating θ_2, given as vectors of weights on the points $E^1_{2,0}$, are

$$q(1/4,1/2,1/4,0,0) + (1-q)(0,0,1/4,1/2,1/4) \quad \text{for } 0\le q\le 1.$$

The optimal designs for estimating β_1 are

$$q(1/8,3/8,3/8,1/8,0) + (1-q)(0,1/8,3/8,3/8,1/8) \text{ for } 0\le q\le 1.$$

The optimal design for estimating θ_0 is

$$(0,0,1,0,0).$$

Example 3.2.2. Consider a random variable Y(x) with mean $E\,Y(x) = \theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\beta_2(x^2-2x_+^2)$ where $x\varepsilon[-1,1]$. For this example $W_{3,0}^1(x)=3(2+\sqrt{3})x+3/2(12+7\sqrt{3})(x^2-2x_+^2)+\frac{(26+15\sqrt{3})}{2}x^3$ and $E_{3,0}^1=\{-1,-(\sqrt{3}-1),-(3\sqrt{3}-5),(3\sqrt{3}-5),(\sqrt{3}-1),1\}$. The optimal design $\mu_{\theta_1}^0$ for estimating θ_1, given as vectors of weights on the points $E_{3,0}^1$ to 5 decimal places, is

$$\mu_{\theta_1}^0 = (.05955,.14088,.41467,.35912,.02578,0).$$

$\mu_{\theta_1}^1$ is the symmetric image of the above, and all the optimal designs can be represented as a convex combination of these two. The optimal designs for estimating θ_3 are any convex combination of

$$\mu_{\theta_3}^0 = (.17863,.35566,.31101,.14434,.01036,0)$$

and its symmetric image $\mu_{\theta_3}^1$.

The optimal designs for estimating β_2 are any convex combination of

$$\mu_{\beta_2}^0 = (.11909,.27233,.33878,.22767,.04213,0)$$

and its symmetric image $\mu_{\beta_2}^1$.

Example 3.2.3. Consider a random variable Y(x) with mean $E\,Y(x) = \theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\theta_4x^4+\beta_3(x^3-2x_+^3)$ where $x\varepsilon[-1,1]$. For this example $W_{4,0}^1(x) = -1+8(3+2\sqrt{2})x^2+2(17+12\sqrt{2})x^4+8(7+5\sqrt{2})(x^3-2x_+^3)$ and

$E^1_{4,0}=\{-1,-2(\sqrt{2}-1),-(\sqrt{2}-1),0,(\sqrt{2}-1),2(\sqrt{2}-1),1\}$. The optimal designs for estimating θ_2, given as vector of weights on the points $E^1_{4,0}$ to 5 decimal places, are any convex combination of

$$\mu^0_{\theta_2} = (.07322,.16220,.26726,.31786,.15952,.01994,0)$$

and its symmetric image $\mu^1_{\theta_2}$.

The optimal designs for estimating θ_4 are any convex combination of

$$\mu^0_{\theta_4} = (.14645,.29315,.28452,.19822,.06903,.00863,0)$$

and its symmetric image $\mu^1_{\theta_4}$.

The optimal designs for estimating β_3 are any convex combination of

$$\mu^0_{\beta_3} = (.10984,.23548,.27589,.24242,.11428,.02209,0)$$

and its symmetric image $\mu^1_{\beta_3}$.

The optimal design for estimating θ_0 is

$$(0,0,0,1,0,0,0).$$

We now begin a development similar to the above to enable us to classify the $\bar{c}_p$-optimal designs for those functions in (3.2.2) that are of opposite parity of x^n. We will show in this case that the designs are unique for any particular choice of n and k.

Lemma 3.2.7. There exists a unique polynomial $W^2_{n,k}(x)$ which is a linear combination of the functions in (3.2.2), satisfying:

(1) $|W^2_{n,k}(x)| \leq 1$;

(2) The set $E^2_{n,k}=\{x: |W^2_{n,k}(x)|=1\}$ contains exactly $n+2[\frac{k+1}{2}]$ points and is symmetric about zero;

(3) $W^2_{n,k}(x)$ attains its supremum at each of the points of the set $E^2_{n,k}$ with alternating signs. $W^2_{n,k}(x)$ is of the form

3.2.18. $$W^2_{n,k}(x) = \sum_{j=0}^{[\frac{n-1}{2}]} \beta^1_{n-2j-1}\, x^{n-2j-1} + \sum_{j=1}^{[\frac{k+1}{2}]} \beta^1_{n-2j}(x^{n-2j} - 2x_+^{n-2j}),$$

and all the coefficients are non-zero. (Note that $W^2_{n,k}(x)$ is even or odd as n is odd or even).

Proof: Let V be the linear space spanned by the functions in (3.2.2), excluding x^{n-1}. If $g(x)\varepsilon$ V, then $g(-x)\varepsilon V$ since each function in (3.2.2) is either odd or even. There exists a best approximation of $b_{n-1}(x) = x^{n-1}$, say $Q_{n-1}(x)$, with respect to V, which is even (odd) as n is odd (even). Meinardus (1967, pages 26 and 27). We can construct this unique even or odd best approximation with an argument similar to that in lemma (3.2.1). We define

$$W^2_{n,k}(x) = \frac{x^{n-1}-Q_{n-1}(x)}{||x^{n-1}-Q_{n-1}(x)||}$$

and note that by a similar argument to that of lemma (3.2.1), $W^2_{n,k}(x)$ is of the form (3.2.18) with all non-zero coefficients.

Let $H(x)$ be a linear combination of the functions in (3.2.2) satisfying (1), (2) and (3). If n is odd, $\frac{H(x)+H(-x)}{2}$ satisfies (1), (2) and (3) and is a nontrivial even function. So $W^2_{n,k} = \frac{H(x)+H(-x)}{2}$. If we write $H(x)$ as the sum of its odd ($\mathcal{O}(x)$) and even ($\mathcal{E}(x)$) parts, we have $H(x)=\mathcal{E}(x)+\mathcal{O}(x)=W^2_{n,k}(x)+\mathcal{O}(x)$. Now $\mathcal{O}(x)$ must equal zero at the points in $E^2_{n,k}$ and must also have its derivative zero there. If the derivative were not zero at the points in $E^2_{n,k}$, then $\sup |H(x)|>1$. This implies that $\mathcal{O}(x)\equiv 0$. Thus, $H(x)\equiv W^2_{n,k}(x)$ and we have uniqueness. A similar argument holds for n even.

Lemma 3.2.8. Among all polynomials $f(x)$ in the functions (3.2.2),

(1) $W^2_{n,k}(x)/\beta^1_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where $f(x)$ is any polynomial in (3.2.2) with the coefficient of x^{n-2j-1} unity for $j = 0,\ldots,[\frac{n-1}{2}]$, and

(2) $W^2_{n,k}(x)/\beta^1_{n-2j}$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where $f(x)$ is any polynomial in (3.2.2) with the coefficient of $(x^{n-2j}-2x_+^{n-2j})$ unity for $j = 1,\ldots,[\frac{k+1}{2}]$.

Proof: The proof follows an argument similar to that in lemma (3.2.3) after consideration of lemma (3.2.7).

<u>Theorem 3.2.2</u>. The optimal designs for estimating the following parameters in (3.2.1):

3.2.19. $$\begin{cases} \theta_{n-2j-1} & \text{for } j = 0,\ldots,[\frac{n-1}{2}] \quad n-2j-1\neq 0 \\ \beta_{n-2j} & \text{for } j = 1,\ldots,[\frac{k+1}{2}] \\ \text{i.e., parameters for functions of the opposite parity of } x^n, \end{cases}$$

are unique and are supported by the full set $E^2_{n,k}$. The support for the optimal design for θ_0 is $\{0\}$.

Proof: Let $\bar{c}_p=(0,\ldots,0,1,0,\ldots,0)$ where the 1 is in the component of $\bar{\theta}=(\theta_0,\ldots,\theta_n,\beta_{n-1-k},\ldots,\beta_{n-1})$ corresponding to some one of the parameters in (3.2.19). By(3.1.3) and an argument similar to that in theorem (3.2.1), we have that

$$V(\bar{c}_p,\mu_p^*) \geq (\beta_p^1)^2[\int(W^2_{n,k}(x))^2\mu_p^*(dx)]^{-1}$$
$$\geq (\beta_p^1)^2,$$

where μ_p^* is $\bar{c}_p$-optimal and β_p^1 is the coefficient of $b_p(x)$ in $W^2_{n,k}(x)$. Since $|W^2_{n,k}(x)|=1$ only for $x\varepsilon E^2_{n,k}$, μ_p^* has its support contained in the set $E^2_{n,k}$, since

$$\inf_{\mu} V(\bar{c}_p,\mu) = (\beta_p^1)^2.$$

Elfving's theorem (theorem 3.1.1) tells us there is a solution to the system

3.2.20. $$\left|\frac{1}{\beta_p^1}\right|\bar{c}_p = \sum_{\nu=1}^{n+2[\frac{k+1}{2}]} \varepsilon_\nu p_\nu \bar{b}(x_\nu)$$

where the $x_\nu \in E^2_{n,k}$, $\sum_{\nu=1}^{n+2[\frac{k+1}{2}]} p_\nu=1$, $p_\nu \geq 0$ and $\varepsilon_\nu=\pm 1$. Any $\overline{c}_p$-optimal design must be a solution to (3.2.20).

Assume that k is odd. The system (3.2.20) describes (n+k+2) equations in (n+k+1) unknowns. The rank of the system is (n+k+1). To see this, we note that the (n+k+1) row vectors omitting the row vector corresponding to x^n, are independent by lemma (2.1.5). This implies that the solution is unique.

Assume that k is even. The system (3.2.20) describes (n+k+2) equations in (n+k) unknowns. The rank of the system is (n+k). Consider the (n+k) row vectors, omitting those corresponding to the functions x^n and $(x^{n-1}-2x_+^{n-1})$. If these rows were not independent, then we would have a nontrivial linear combination, say P(x), such that $P(x_\nu)=0$ for $x_\nu \in E^2_{n,k}$, and the coefficient of x^{n-1} is non-zero by lemma (2.1.5). After normalization, we may assume the coefficient of x^{n-1} in P(x) is unity. Since

$$\frac{W^2_{n,k}(x)}{\beta^1_{n-1}} - P(x)$$

has (n+k-1) distinct zeros falling between successive x_ν's in $E^2_{n,k}$, and is of the form $\sum_{i=0}^{n-2} \alpha_i x^i + \sum_{i=n-1-2}^{n-2} \alpha_i^1 (x^i-2x_+^i)$, it must be identically equal to zero. (See lemma (2.1.5).) This implies that $P(x)\equiv 0$. This implies the solution to (3.2.20) is unique.

Lemma (3.2.9) will verify that the optimal designs are supported by the full set. If n is odd, we see by inspection of the system (3.2.20) that the optimal design for estimating θ_0 is unique and is supported by {0}.

<u>Lemma 3.2.9</u>. The equations of the system (3.2.20) have as their respective unique solutions a set of p_ν's such that $p_\nu > 0$ for all $\nu = 1,\ldots,n+2[\frac{k+1}{2}]$.

Proof: Assume that k is even. Since the system (3.2.20) has a solution, we need only solve a reduced system of (n+k) independent equations. As in theorem (3.2.2), we eliminate the contribution of the functions x^n and $(x^{n-1}-2x_+^{n-1})$. The remaining coefficient matrix is square and nonsingular. We are thus led to the situation of lemma (3.2.4) with k odd. This verifies that the support mentioned in theorem (3.2.2) is on the full set $E^2_{n,k}$.

Assume that k is odd. As in the above, we eliminate the contribution of the function x^n. The remaining coefficient matrix is square and nonsingular. If $p_{\nu_0}=0$, for some $\nu_0=1,\ldots,n+k+1$, when solving by Cramer's method we would be led to a nontrivial linear combination of the form

$$P(x) = \sum_{i=1}^{n-1} a_i x^i + \sum_{i=n-1-k}^{n-1} b_i (x^i - 2x_+^i)$$

where either $a_{n-2i-1}=0$ for some $i = 1,\ldots,[\frac{n-1}{2}]$, $n-2i-1\neq 0$ or $b_i=0$ for some $i = 1,\ldots,[\frac{k+1}{2}]$. Also $P(x_i)=0$ for all $x_\nu \neq \pm x_{\nu_0}$. If n were odd, then P(x)+P(-x)=Q(x) would have the form

$$Q(x) = \sum_{i=1}^{\frac{n-1}{2}} 2a_{2i}x^{2i} + \sum_{i=\frac{n-k-2}{2}}^{\frac{n-3}{2}} 2b_{2i+1}(x^{2i+1} - 2x_+^{2i+1})$$

where $a_{2i}= 0$ or $b_{2i} = 0$ for some i. $Q(x_\nu)=0$ for all $x_\nu \neq \pm x_{\nu_0}$.

On (0,1) Q(x) must have $(\frac{n+k-2}{2})$ zeros, but by Descartes' Rule of Signs Q(x) may only have $(\frac{n+k-4}{2})$ zeros. Appropriate adjustments can be made for the case where $x_{\nu_0} = 0$. This implies that $Q(x)\equiv 0$, which in turn implies that $P(x)\equiv 0$. This contradicts the fact that $p_{\nu_0}=0$. A similar argument will hold for n even.

Example 3.2.4. Consider a random variable Y(x) with mean value $E\,Y(x) = \theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\theta_4x^4+\beta_3(x^3-2x_+^3)$ where $x\varepsilon[-1,1]$. For this example $W_{4,0}^2(x) = -3x+4x^3$ and $E_{4,0}^2 = \{-1,-1/2,1/2,1\}$. The optimal design for estimating θ_1, given as a vector of weights on the points $E_{4,0}^2$, is

$$(1/18,8/18,8/18,1/18).$$

The optimal design for estimating θ_3, given as a vector of weights on the points $E_{4,0}^2$, is

$$(1/6,2/6,2/6,1/6).$$

3.3. Optimal Designs for Basis II.

In this section we consider a random variable $Y(x)$ with mean

3.3.1. $$\sum_{i=0}^{n} \theta_i x^i + \sum_{i=n-1-k}^{n-1} \beta_i x_+^i$$

where $x\varepsilon[-1,1]$ and $n-1-k\geq 1$. As regression functions, we have the $(n+k+2)$ linearly independent and continuous functions

3.3.2. $$\bar{b}(x) = (1,x,\ldots,x^n,x_+^{n-k-1},\ldots,x_+^{n-1}).$$

In order to classify the $\bar{c}_p$-optimal designs, we will follow the pattern of section 3.2. When finding the best approximation of $b_p(x)$ by the remaining functions in (3.3.2), we will again use the polynomials of lemmas (3.2.1) and (3.2.7). It is clear that any linear combination of the functions in (3.3.2) can be formed by a linear combination of the functions in (3.2.2) and conversely. However, we will find that for identical $b_p(x)$ in both (3.3.2) and (3.2.2), the optimal designs are not the same in all cases. See examples (3.2.2) and (3.3.2). The following lemma concerning the zeros of a polynomial in the functions (3.3.2) will be needed when finding the best approximation of $b_p(x)$ in (3.3.2).

Lemma 3.3.1. Let $B(x) = \sum_{i=0}^{n} \alpha_i x^i + \sum_{i=n-1-k}^{n-1} \beta_i x_+^i$ for $x\varepsilon[-1,1]$ and $n-1-k\geq 1$.

If at least one of the $\alpha_i=0$ for $n-1\geq i\geq n-1-k$ or at least one of the $\beta_{n-2j-1}=0$ for $j = 0,\ldots,[\frac{k}{2}]$, then $B(x)$ cannot satisfy both

(1) B(x) has at least $(n+2[\frac{k}{2}]+2)$ distinct zeros in [-1,1] with at least $[\frac{n+2[\frac{k}{2}]+2}{2}]$ distinct zeros in both [-1,0) and (0,1], and

(2) B(x) does not vanish identically in any interval containing two of these distinct zeros.

Proof: Assume $k=n-2$ and n is even. B(x) must have n distinct zeros in both [-1,0) and (0,1]. By Descartes' Rule of Signs, B(x) must have at least n variations of sign presented by its coefficients in (0,1]. The same must be true for B(-x) for $x\varepsilon(0,1]$. This is clearly impossible with the missing coefficients. A similar result is true with n odd.

Assume k is less than n-2. Then the derivative of B(x) exists throughout [-1,1] ,and by Rolle's Theorem, the zeros of B(x) are separated by the zeros of $B^{(1)}(x)$. ($B^{(j)}(x)$ denotes the j-derivative of b(x).) $B^{(1)}$ cannot vanish identically in any interval between two of its zeros. Clearly, $B^{(n-2-k)}(x)$ must satisfy the above case with $k=n_1-2$, where $n_1=k+2$, by successive application of Rolle's Theorem. Again we arrive at a contradiction.

Lemma 3.3.2. If B(x) satisfies the hypothesis and condition (1) of lemma (3.3.1), then $B(x)\equiv 0$ for $x\varepsilon(-1,1]$.

Proof: Lemma (3.3.1) implies that $B(x)\equiv 0$ for some interval. This implies that $B(x)\equiv 0$ on at least one of [-1,0] or [0,1]. On either of the above intervals, a nontrivial B(x) could thus have at most k distinct zeros by Descartes' Rule of Signs if it vanished identically on the other. Condition (1) implies that we must have $\left[\frac{n+2[\frac{k}{2}]+2}{2}\right] > k$ distinct zeros. Thus $B(x$ for $x \in [-1,1]$.

We are now able to go directly to the minimizing polynomials.

Lemma 3.3.3. Among all polynomials $f(x)$ in the functions (3.3.2),

(1) $W^1_{n,k}(x)/\beta_{n-2j}$ minimizes $\sup_{-1 \le x \le 1} |f(x)|$ where $f(x)$ is any polynomial in (3.3.2) with the coefficient of x^{n-2j} unity for $j = 0,\ldots,[\frac{n+1}{2}]$;

(2) $W^1_{n,k}(x)/\beta_{n-2j-1}$ minimizes $\sup_{-1 \le x \le 1} |f(x)|$ where $f(x)$ is any polynomial in (3.3.2) with the coefficient of x^{n-2j-1} unity for $j = 0,\ldots,[\frac{k}{2}]$; and

(3) $W^1_{n,k}(x)/-2\beta_{n-2j-1}$ minimizes $\sup_{-1 \le x \le 1} |f(x)|$ where $f(x)$ is any polynomial in (3.3.2) with the coefficient of x_+^{n-2j-1} unity for $j = 0,\ldots,[\frac{k}{2}]$.

$W^1_{n,k}(x)$ is the polynomial described in lemma (3.2.1).

Proof: The cases for which the coefficient of x^{n-2i} is unity for $i=[\frac{n-k-3}{2}],\ldots,[\frac{n}{2}]$, and $i=0$, follow the same arguments as in lemma (3.2.3). This happens because any best approximation can be expressed in a unique manner since the remaining functions span the same space.

Assume now that we are minimizing some one of the remaining functions listed in (1), (2) or (3), and let θ_ℓ represent its coefficient in $W^1_{n,k}(x)$. Assume $Q(x)$ is a better minimizing polynomial than $W^1_{n,k}(x)/\theta_\ell$.

Now

$$W^1_{n,k}(x)/\theta_\ell - Q(x)$$

satisfies the hypothesis of lemma (3.3.2). This implies that $Q(x) \equiv W^1_{n,k}(x)$ and proves the lemma.

In the following theorem, the parameters listed in (3.3.3) correspond to those of the same parity as x^n when $W^1_{n,k}(x)$ is written in the basis (3.2.2) rather than in that of (3.3.2).

Theorem 3.3.1. The optimal designs for estimating the following parameters in (3.3.1):

$$\text{3.3.3.} \qquad \begin{cases} \theta_{n-2j} & \text{for } j = 0,\ldots,[\frac{n}{2}] \quad n-2j \neq 0 \\ \theta_{n-2j-1} & \text{for } j = 0,\ldots,[\frac{k}{2}] \\ \beta_{n-2j-1} & \text{for } j = 0,\ldots,[\frac{k}{2}], \end{cases}$$

have their supports contained in the set $E^1_{n,k}$ (see lemma 3.2.1) and satisfy the following:

(i) When k is odd the optimal design for each parameter listed in (3.3.3) is unique and is supported by the full set $E^1_{n,k}$;

(ii) When k is even the optimal designs for the parameters in (3.3.3) are not unique and satisfy the following:

(a) The optimal designs are a convex combination of two probability measures $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ (or $\mu^0_{\beta_h}$ and $\mu^1_{\beta_h}$) where $\mu^0_{\theta_\ell}\{+1\}=0$ $(=\mu^0_{\beta_h}\{+1\})$ and $\mu^1_{\theta_\ell}\{-1\}=0$ $(=\mu^1_{\beta_h}\{-1\})$;

(b) In the convex combination described in (a), all the designs other than $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ ($\mu^0_{\beta_h}$ and $\mu^1_{\beta_h}$) are supported by the full set $E^1_{n,k}$;

(c) The vectors of weights associated with the optimal designs of (a) lie on parallel lines;

(iii) The support for the optimal design for θ_0 is $\{0\}$.

Proof: By an argument similar to that in theorem (3.2.1), we have that for any $\overline{c}_p$-optimal design μ^*_p, for one of the parameters in (3.3.3),

$$V(\overline{c}_p,\mu^*_p) = (\beta_p)^2[\int(W^1_{n,k}(x))^2\mu_p(dx)]^{-1} = (\beta_p)^2.$$

Any μ^*_p must have its support contained in the set $E^1_{n,k}$. Elfving's theorem (theorem 3.1.1) tells us there is a solution to the system

$$3.3.4. \qquad \left|\frac{1}{\beta_p}\right|\overline{c}_p = \sum_{\nu=1}^{n+2[\frac{k}{2}]+3} \varepsilon_\nu p_\nu \overline{b}(x_\nu)$$

where the $x_\nu \in E^1_{n,k}$, $\sum_{\nu=1}^{n+2[\frac{k}{2}]+3} p_\nu=1$, $p_\nu \geq 0$ and $\varepsilon_\nu = \pm 1$.

Assume that k is odd and n is even. As in theorem (3.2.1), we know that there is a unique solution. Assume $p_{\nu_0}=0$ for $\nu_0 \neq (\frac{n+k+3}{2})$ (i.e. $x_{\nu_0} \neq 0$) when estimating one of the coefficients in (3.3.3). When solving (3.3.4) by Cramer's method, we are led to a polynomial

$$P(x) = \sum_{i=1}^{n} a_i x^i + \sum_{i=n-1-k}^{n-1} b_i x^i_+ \quad \text{where} \quad \sum_{i=1}^{n} a_i^2 + \sum_{i=n-1-k}^{n-1} b_i^2 > 0.$$

Now $p(x_\nu)=0$ for all $x_\nu \neq x_{\nu_0}$. Let $Q(x)=P(x)+P(-x)$. Then

$$Q(x)= \sum_{i=1}^{\frac{n}{2}} 2a_{2i}x^{2i} + \sum_{i=\frac{n-1-k}{2}}^{\frac{n-2}{2}} b_{2i}x^{2i} + \sum_{i=\frac{n-1-k}{2}}^{\frac{n-2}{2}} b_{2i+1}(2x_+^{2i+1} - x^{2i-1}).$$

If $Q(x_{\nu_0})=0$, then $Q(x)\equiv 0$ by Descartes' Rule of Signs on $[-1,0)$ and $(0,1]$. If $Q(x_{\nu_0})\neq 0$, then by Descartes' Rule of Signs, $Q(x)$ has the maximum number of zeros possible, $(\frac{n+k-1}{2})$, in both $[-1,0)$ and $(0,1]$. This implies that a_n and b_{n-1} must be non-zero and of opposite signs. Multiplication of $Q(x)$ by a unique non-zero constant c will make $Q(x_{\nu_0})$ of opposite sign as $W^1_{n,k}(x_{\nu_0})$, with either the coefficient of x^n or x^{n-1} in the two polynomials the same. $W^1_{n,k}(x)-c\,Q(x)$ must satisfy the conditions of lemma (3.3.2). This contradiction implies that $Q(x)\equiv 0$. Thus $P(x)$ must be of the form

$$\sum_{i=1}^{\frac{n}{2}} a_{2i-1}x^{2i-1} + \sum_{i=\frac{n-1-k}{2}}^{\frac{n-2}{2}} a_{2i}x^{2i} + \sum_{i=\frac{n-1-k}{2}}^{\frac{n-2}{2}} b_{2i}x_+^{2i}.$$

Descartes' Rule of Signs implies that $P(x)\equiv 0$. This is the desired contradiction and proves that $p_{\nu_0}\neq 0$. A similar argument holds for $\nu_0=(\frac{n+k+3}{2})$. For n odd, we would let $Q(x)=P(x)-P(-x)$ and follow similar arguments.

Assume k is even. Following arguments similar to those in theorem (3.2.1) and lemma (3.2.6), we find that (ii)-(a) and (c) of this theorem hold . In order to prove (ii)-(b), assume that for some $q\varepsilon(0,1)$, where $\mu_q^* = q\mu_{\theta_p}^0 + (1-q)\mu_{\theta_p}^1$, we have $\mu_q^*\{x_{\nu_0}\}=0$. This implies that $\mu_{\theta_p}^0\{x_{\nu_0}\}=\mu_{\theta_p}^1\{x_{\nu_0}\}=0$ so that $\mu_q^*\{x_{\nu_0}\}=0$ for all $q\varepsilon[0,1]$. If this were true, then the reduced system from (3.3.4)

3.3.5. $$\left|\frac{1}{\beta_p}\right| \overline{c}_p = \sum_{\substack{\nu=1 \\ \nu\neq\nu_0}}^{n+k+3} \varepsilon_\nu p_\nu \, \overline{b}(x_\nu)$$

would have more than one solution and the system would be singular. Consider the function $K_{n,k}^{\nu_0}$ of lemma (3.2.5). When applied to the coefficient matrix of the system (3.3.5), we would have an equivalent system with the row corresponding to x^n having all zeros except a ±1 in the $n+k+4-\nu_0$ column. An application of lemma (2.1.5) shows that this system is nonsingular and implies that (3.3.5) has a unique solution. This is the desired contradiction, implying $\mu_q^*\{x_{\nu_0}\}>0$. If $x_{\nu_0}=0$, then $\mu_q^*\{x_{\nu_0}\}= c_{\nu_0}>0$ by application of $K_{n,k}^{\nu_0}(x)$ to (3.3.4).

<u>Remark 3.3.1.</u> When estimating the parameters in (3.3.3) other than θ_{n-2_j-1} for $j = 0,\ldots,[\frac{k}{2}]$, the system (3.3.4) can be transformed into the system (3.2.9) or (3.2.10) by elementary row operations on the coefficient matrix of the p_ν's. This implies that the optimal designs for estimating the parameters θ_{n-2j}; $j = 0,\ldots,[\frac{n}{2}]$, in (3.3.3) are the same as those for θ_{n-2j}; $j = 0,\ldots,[\frac{n}{2}]$, in (3.2.8). Also the optimal

designs for estimating the parameters β_{n-2j-1}; $j = 0,\ldots,[\frac{k}{2}]$, in (3.3.3) are the same as those for β_{n-2j-1}; $j = 0,\ldots,[\frac{k}{2}]$, in (3.2.8).

Example 3.3.1. Consider a random variable Y(x) with mean $E\,Y(x)=\theta_0+\theta_1x+\theta_2x^2+\beta_1x_+$ where $x\varepsilon[-1,1]$. For this example, $W^1_{2,0}(x)=1+16x+8x^2-16x_+$ and $E^1_{2,0}=\{-1,-1/2,0,1/2,1\}$. The optimal designs for estimating θ_1, given as vectors of weights on the points $E^1_{2,0}$, are

$$q(1/8,4/8,3/8,0,0)+(1-q)(0,2/8,3/8,2/8,1/8) \text{ for } 0\leq q\leq 1.$$

See example (3.2.1) for the coefficients θ_0, θ_2 and β_1.

Example 3.3.2. Consider a random variable Y(x) with mean $E\,Y(x)=\theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\beta_2x_+^2$ where $x\varepsilon[-1,1]$. $W^1_{3,0}(x)$, $E^1_{3,0}$ and the optimal designs for θ_3, θ_1 and β_2 can be found in example (3.2.2). The optimal designs for estimating θ_2, given as vectors of weights on the points $E^1_{3,0}$ to 5 decimal places, are

$$q(.11909,.30011,.36656,.19989,.01435,0)+$$
$$(1-q)(0,.06991,.25545,.31100,.24455,.11909) \text{ for } 0\leq q\leq 1.$$

Example 3.3.3. Consider a random variable Y(x) with mean $E\,Y(x)=\theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\theta_4x^4+\beta_3x_+^3$ where $x\varepsilon[-1,1]$. $W^1_{4,0}(x)$, $E^1_{4,0}$ and the optimal designs for θ_0, θ_2, θ_4 and β_3 can be found in example (3.2.3). The optimal designs for estimating θ_3, given as vectors of weights on the points $E^1_{4,0}$ to 5 decimal places, are

$$q(.10984,.24590,.29672,.24242,.09344,.01168,0)+$$

$$(1-q)(0,.03251,.13511,.24242,.25505,.22507,.10984) \text{ for } 0\leq q\leq 1.$$

Example 3.3.4. Motion with a constant acceleration that has an instantaneous change in velocity.

Consider the motion of a moving particle on the s-axis whose coordinate is s at time t. We assume that the particle is subjected to a constant acceleration $-a\text{ft/sec}^2.a>0$, starting at s=H feet with velocity $v_0$ft/sec. The equation of motion of this particle is

3.3.6. $$s(t)=-\frac{a}{2}t^2+v_0t+H.$$

Assume that at time t_1 we have an instantaneous change in velocity and let s=0 when $t=t_1$. (3.3.6) becomes

3.3.7. $$s(t)=-\frac{a}{2}t^2+v_0t+\frac{a}{2}t_1^2-v_0t_1.$$

Thus, when $t=t_1$, the velocity $v=v_1$, and (3.3.5) becomes

3.3.8. $$s(t)=\frac{a}{2}t_1^2-v_0t_1+v_0t-\frac{a}{2}t^2+(v_1+a\,t_1-v_0)(t-t_1)_+$$

which is a monospline with knot at t_1.

If we let a=32 (the acceleration due to gravity), (3.3.8) would describe the motion of a ball subjected to the velocity v_0 at height H with a bounce at time t_1. Writing (3.3.8) in a basis suggested by that of (3.2.2), we would have

3.3.9. $$s(t)=\left(\frac{v_1+v_0-32t}{2}\right)(t-t_1)-16(t-t_1)^2 + \left(\frac{v_0-32t_1-v_1}{2}\right)[(t-t_1)-2(t-t_1)_+].$$

The coefficient of $[(t-t_1)-2(t-t_1)_+]$ in $2s(t)$ is the difference of the velocities just prior to, and after, impact. The coefficient of $(t-t_1)$ in $s(t)$ is the average velocity just prior to, and after, impact and would be zero if there were a perfect bounce. The coefficient of $(t-t_1)^2$ in $2s(t)$ corresponds to the acceleration. If an experiment were to be designed in which one of the above coefficients was of prime interest, then one should consult example (3.2.1).

Writing (3.3.9) in a basis suggested by that of (3.3.2), we would have

$$s(t) = 16t_1^2-v_0t_1+v_0t-16t^2+(v_1-v_0+32t_1)\ (t-t_1)_+.$$

The coefficient of $(t-t_1)_+$ is the difference of the velocities just prior to, and after, impact. The coefficient of t^2 in $2s(t)$ corresponds to the acceleration. The coefficient of t is initial velocity and the coefficient of unity is the initial height. If one is interested in an estimate of one of these coefficients, then one should consult example (3.3.1).

Example 3.3.5. As in example (3.3.4) and section 3.3, let us consider a random variable with mean value

3.3.10. $$\theta_0+\theta_1t+\theta_2t^2+\beta_1(t-t_1)_+.$$

Assume we are in the time interval [0,1] and have the ability to adjust the experiment in (3.3.4) in such a manner that t_1 can be arbitrarily chosen in (0,1). If our main interest lies in estimating β_1, we would want to know what values of t_1 would minimize the variance of our estimate. For any particular value of t_1, the $\overline{c}_4$-optimal designs will give the minimum variance estimate. So we need only minimize these variances for $t_1 \varepsilon (0,1)$.

The best approximation of $(t-t_1)_+$ by 1, t and t^2 on [0,1] is unique and alternates at least three times. Karlin and Studden (1966, page 280). Actually, if $t_1=1/2$, we have four alternations, and if $t_1 \neq 1/2$, we have exactly three. If $t_1 \varepsilon [1/2,2/3]$, we are led to the normalized polynomial

3.3.11. $$W(t) = 1-\frac{8t}{t_1} + \frac{8t^2}{t_1^2} - \frac{16}{t_1}(t-t_1)_+.$$

The variance of the best estimate of β_1 in this case is

3.3.12. $$\left(\frac{16}{t_1}\right)^2$$

and is a minimum for $t_1=2/3$. If $t_1 \varepsilon [2/3,1)$, we are led to the normallized polynomial

3.3.13. $$W(t) = 1-\frac{8t}{t_1} + \frac{8t^2}{t_1^2} + \frac{2(t_1-2)^2}{t_1^2(t_1-1)}(t-t_1)_+.$$

The variance of the best estimate of β_1 in this case is

3.3.14. $$\left(\frac{2(t_1-2)^2}{t_1^2(t_1-1)}\right)^2$$

and is a minimum for $t_1^0 = [1-(\sqrt{5}-2)]$. The two variances are equal for $t_1=2/3$, so that t_1^0 gives the minimum variance for $t_1 \varepsilon [1/2,1)$. Due to the symmetry involved in this problem, we will find the same minimum variance at $1-t_1^0=(\sqrt{5}-2)$.

We now begin the development that will enable us to classify the $\overline{c}_p$-optimal designs for those parameters corresponding to the functions in (3.3.2) that when written in the basis (3.2.2) would be of opposite parity to that of x^n. These parameters are listed in (3.3.15).

<u>**Lemma 3.3.4.**</u> Among all polynomials f(x) in the functions (3.3.2),

(1) $W_{n,k}^2(x)/\beta_{n-2j-1}^1$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where f(x) is any polynomial in (3.3.2) with the coefficient of x^{n-2j-1} unity for $j = [\frac{k+2}{2}],\ldots,[\frac{n-1}{2}]$, and

(2) $W_{n,k}^2(x)/-2\beta_{n-2j}^1$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where f(x) is any polynomial in (3.3.2) with the coefficient of x_+^{n-2j} unity for $j = 1,\ldots,[\frac{k+1}{2}]$.

Proof: The proof for (1) is the same as that for (1) in lemma (3.2.8). The functions in (3.3.2), omitting the x^{n-2j-1} in (1) above, span the same space as the approximating functions in lemma (3.2.8).

Assume we are approximating x_+^{n-2j} in (2) by the remaining functions in (3.3.2). Any best approximation of x_+^{n-2j} would also lead to a best approximation of $(-1/2)(x^{n-2j}-2x_+^{n-2j})$ with the same maximum deviation. If n were odd (even), then a best approximation would exist that would be even (odd). Upon subtracting this best approximation from $(-1/2)(x^{n-2j} - 2x_+^{n-2j})$, we note that the difference normalized must satisfy (1), (2) and (3) of lemma (3.2.7). Thus, by the uniqueness of $W^2_{n,k}(x)$, (2) holds.

Theorem 3.3.2. The optimal designs for estimating the following parameters in (3.3.1):

3.3.15. $$\begin{cases} \theta_{n-2j-1} & \text{for } j = [\frac{k+2}{2}],\ldots,[\frac{n-1}{2}] \\ \beta_{n-2j} & \text{for } j = 1,\ldots,[\frac{k+1}{2}] \end{cases}.$$

are unique and are supported by the full set $E^2_{n,k}$.

Proof: The designs and proof for the parameters in (3.3.15) are the same as those in theorem (3.2.2). As before, elementary row operations and Elfving's theorem establishes the equivalence.

Example 3.3.4. Consider the set up of example (3.2.4). According to the above theorem, the optimum design for estimating θ_1 when $E\ Y(x)=\theta_0+\theta_1 x+\theta_2 x^2+\theta_3 x^3+\theta_4 x^4+\beta_3 x_+^3$, given as a vector of weights on the points $E^2_{4,0}$, is (1/18,8/18,8/18,1/18). The optimum designs for estimating θ_3 are given in example (3.3.3).

Example 3.3.5. Consider a random variable Y(x) with mean $E\ Y(x)=\theta_0+\theta_1 x+\theta_2 x^2+\theta_3 x^3+\beta_1 x_+ +\beta_2 x_+^2$ where $x\varepsilon[-1,1]$. For this example, $W^2_{3,1}(x) = 1+16x+8x^2-16x_+$ and $E^2_{3,1}=\{-1,-1/2,0,1/2,1\}$. The optimal design for estimating β_1, given as a vector of weights on the points $E^2_{3,1}$, is

$$(1/16,4/16,6/16,4/16,1/16).$$

3.4. Optimal Designs for Basis III.

In this section we consider a random variable Y(x), with mean

3.4.1. $$\sum_{i=0}^{n} \theta_i (x+1)^i + \sum_{i=n-1-k}^{n-1} \beta_i x_+^i,$$

where $x \in [-1,1]$ and $n-1-k \geq 1$. As regression functions, we are considering the $(n+k+2)$ linearly independent and continuous functions.

3.4.2. $$\bar{b}(x) = (1,(x+1),\ldots,(x+1)^n, x_+^{n-k-1},\ldots,x_+^{n-1}).$$

When k=0, it will be shown that the $\bar{c}_p$-optimal designs for all the parameters in (3.4.1) will have their supportscontained in the same set, $E^1_{n,0}$. The basic format of this section will follow that of the two preceding ones. The discussions here will depend on, and be similar to, earlier proofs as this next lemma illustrates.

Lemma 3.4.1. $W^1_{n,k}(x)$ of lemma (3.2.1),when expressed in the basis (3.4.2),is of the form

$$W^1_{n,k}(x) = \sum_{j=0}^{n} \lambda_j (x+1)^j + \sum_{j=0}^{[\frac{k}{2}]} \gamma_{n-2j-1} x_+^{n-2j-1}$$

where all the $\lambda_j \neq 0$, $j = 0,\ldots,n$; and $\gamma_{n-2j-1} \neq 0$, $j = 0,\ldots,[\frac{k}{2}]$.

Proof: Without loss of generality we may assume that k is even. It is also clear that $\gamma_{n-2j-1} = -2\beta_{n-2j-1}$, $j = 0,\ldots,\frac{k}{2}$, in lemma (3.2.1). Thus we need only show that $\lambda_j \neq 0$ for $j = 0,1,\ldots,n$.

Since $W^1_{n,k}(x)$ alternates $(n+k+3)$ times in $[-1,1]$ and reaches its maximum at the endpoints, it has its maximal number of zeros $(n+k+2)$ in $(-1,1)$. (See lemma (2.1.6).) For the remainder of this proof let $W(x) \equiv W^1_{n,k}(x)$ and $W^{(j)}(x)$ represent the jth derivative of $W(x)$. Let us expand $W(x)$ in its Taylor series about the point $\{-1\}$. $W(-1) \neq 0$. If $n-1-k \geq 2$, $W^{(1)}(x)$ exists for all $x \in [-1,1]$, and by applying Rolle's theorem, $W^{(1)}(x)$ has $(n+k+1)$ distinct zeros which are separated by those of $W(x)$. This implies that $W^{(1)}(-1) \neq 0$. We can repeat this argument so that $W^{(j)}(-1) \neq 0$ for $j = 0,1,\ldots,n-k-2$. Since $W^{(n-k-2)}(x)$ has $(2k+4)$ distinct zeros in $(-1,1)$ with $(k+2)$ in $(-1,0)$, it follows that $W^{(j)}(-1) \neq 0, j = 0,\ldots,n$. Since

$$\lambda_j = \frac{W^{(j)}(-1)}{j!} \quad j = 0,\ldots,n,$$

this completes the proof.

Lemma 3.4.2. Among all polynomials $f(x)$ in the functions (3.4.2),

(1) $W^1_{n,k}(x)/\lambda_j$ minimizes $\sup_{-1 \leq x \leq 1} |f(x)|$ where $f(x)$ is any polynomial in (3.4.2) with the coefficient of $(x+1)^j$ unity for $j = 0,\ldots,n$, and

(2) $W^1_{n,k}(x)/\gamma_{n-2j-1}$ minimizes $\sup_{-1 \leq x \leq 1} |f(x)|$ where $f(x)$ is any

polynomial in (3.4.2) with the coefficient of x_+^{n-2j-1} unity for $j = 0,\ldots,[\frac{k}{2}]$.

Proof: (2) follows immediately from lemma (3.3.3)-(3).

Assume for some $j = 0,\ldots,n$ that there is a better approximation of $(x+1)^j$, say $P_2(x)$, than that suggested by $W^1_{n,k}(x)/\lambda_j$. This implies that $P_1(x) \equiv W^1_{n,k}(x)/\lambda_j - ((x+1)^j - P_2(x))$ is a polynomial in the function (3.4.2) with $(n+2[\frac{k}{2}]+2)$ distinct zeros and the coefficient of $(x+1)^j$ zero. $P_1(-1)\neq 0$ since we assumed that $P_2(x)$ was a better approximation. $P_1^{(j)}(-1)\neq 0$ for $j = 0,\ldots,n$, by repeated application of Rolle's theorem as in lemma (3.4.1). When k is odd, any additional zeros must come in pairs. This is impossible by lemma (2.1.5). Thus, all zeros of $P_1(x)$ and its derivatives must fall in $(-1,1)$. This contradicts the existance of a better approximation.

We are now able to classify the $\bar{c}_p$-optimal designs for the parameters of the functions in (3.4.2).

Theorem 3.4.1. The optimal designs for estimating the following parameters in (3.4.1);

3.4.3.
$$\begin{cases} \theta_i & \text{for } i = 1,\ldots,n \\ \beta_{n-2j-1} & \text{for } j = 0,\ldots,[\frac{k}{2}], \end{cases}$$

have their supports contained in the set $E^1_{n,k}$ of lemma (3.2.1) and satisfy the following:

(i) When k is odd, the optimal design for each parameter listed in (3.4.3) is unique and is supported by the full set $E^1_{n,k}$;

(ii) When k is even, the optimal designs for the parameters in (3.4.3) are not unique and satisfy the following:

(a) The optimal designs are a convex combination of two probability measures $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ (or $\mu^0_{\beta_h}$ and $\mu^1_{\beta_h}$) where $\mu^0_{\theta_\ell}\{+1\}=0$ $(=\mu^0_{\beta_h}\{+1\})$;

(b) In the convex combination described in (a), all the designs other than $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ ($\mu^0_{\beta_h}$ and $\mu^1_{\beta_h}$) are supported by the full set $E^1_{n,k}$.

(c) The vectors of weights associated with the optimal designs of (ii) lie on parallel lines.

(iii) The support for the optimal design for θ_0 is $\{-1\}$.

Proof: By arguments similar to those in theorems (3.2.1) and (3.3.1), the $\overline{c}_p$-optimal design(s) must be a solution to the system

$$3.4.4. \qquad \left|\frac{1}{\beta_p}\right| \overline{c}_p = \sum_{\nu=1}^{n+2[\frac{k}{2}]+3} \varepsilon_\nu p_\nu \overline{b}(x_\nu),$$

where the $x_\nu \in E^1_{n,k}$, $\sum_{\nu=1}^{n+2[\frac{k}{2}]+3} p_\nu=1$, $p_\nu \geq 0$ and $\varepsilon_\nu = \pm 1$. If the coefficients of $W^1_{n,k}(x)$ in the basis (3.4.2) are written as

$\beta=(\lambda_0,\lambda_1,\ldots,\lambda_n,\gamma_{n-1-k},\ldots,\gamma_{n-1})$, then β_p is the (p+1)st coefficient.

By elementary row operations the system (3.4.4) can be made equivalent to the systems (3.2.9) or (3.2.10) for the parameters β_{n-2j-1}, for $j = 0,\dots,[\frac{k}{2}]$ in (3.4.3); and the parameters β_{n-2j-1}, for $j = 0,\dots,[\frac{k}{2}]$ in (3.2.8); as well as θ_n.

Assume k is odd. The uniqueness argument follows that of theorem (3.2.1) after some elementary row operations. When solving (3.4.4), in this case by Cramer's method where $p_{\nu_0}=0$ for some ν_0, we would be led to a polynomial P(x) in the functions (3.4.2) where $P(x) = \sum_{i=0}^{n} d_i(x+1)^i + \sum_{i=n-1-k}^{n-1} b_i x_+^i$ and $\sum_{i=1}^{n} d_i^2 + \sum_{i=n-1-k}^{n-1} b_i^2>0$. We can express P(x) in the basis (3.3.2) and follow the arguments of theorem (3.3.1) to show that $P(x)\equiv 0$. This implies that $p_{\nu_0}>0$.

Assume k is even. By theorem (4.2.5) we know that there is a solution of (3.4.4) with $p_{n+k+3}=0$, say $\mu^0_{\theta_\ell}$. Thus, we can put the augmented matrix of the system of equations (3.4.4) ((n+k+2) equations in (n+k+3) unknowns) in reduced row-echelon form with the first (n+k+2) columns independent. The first (n+k+3) out of (n+k+4) columns of this reduced row-echelon form augmented matrix would be the same as that of theorem (3.2.1). The (n+k+4)th column is the vector of weights of $\mu^0_{\theta_\ell}$ (say p*) on the points of $E^1_{n,k}$. The (n+k+3)rd column consists of the direction components of the parallel lines mentioned in theorems (3.2.1) and (3.3.1) as it does in this theorem. By the symmetry in theorem (3.2.1), we have that the direction components have the form $-1:-a_2:-a_3:\dots:a_3:a_2:1$. If (n+k+4) is even, the directional components are symmetric about $a_{\frac{n+k+4}{2}} = 0$. When $k \neq n-2$, we have that $p^*+k(-1,-a_2,\ldots,a_2,1)$, for k small and positive, corresponds to a probability measure and is also a solution to (3.4.4). This is easily seen after noting that $p_j^*>0$ for $j = 1,\ldots,n+k+2$, by lemma (3.4.3). When $k=n-2$ we, have that $a_i>0$ $i = 1,\ldots,\frac{n+k+2}{2}$ and $p_j^*>0$ for $j = 1,\ldots,\frac{n+k+4}{2}$, so that $p^*+k(-1,-a_2,\ldots,a_2,1)$, for k small and positive, corresponds to a probability measure and is a solution to (3.4.4). Define

$\mu^1_{\theta_\ell} \equiv p^*+k^*(-1,-a_1,\ldots,a_1,1)$ where

$$k^* = \min_j \left\{ k>0 \,\middle|\, \begin{array}{l} p_j^*-ka_j=0;\ j = 1,\ldots,[\frac{n+k+3}{2}] \\ \\ p_j^*+ka_j=0;\ j = [\frac{n+k+3}{2}]+1,\ldots,n+k+2 \end{array} \right\}.$$

The optimal designs for θ_ℓ are any convex combination of $\mu^0_{\theta_\ell}$ and $\mu^1_{\theta_\ell}$ since this would give the only solutions to (3.4.4) that would be probability measures.

If $\mu_q\{x_\nu\}=q\,\mu^0_{\theta_\ell}\{x_\nu\}+(1-q)\,\mu^1_{\theta_\ell}\{x_\nu\}=0$ for some q such that $0<q<1$, then $\mu^0_{\theta_\ell}\{x_\nu\}=\mu^1_{\theta_\ell}\{x_\nu\}=0$. This contradicts the fact that $p_i^*>0$ for $i = 1,\ldots,n+k+2$, when $k\neq n-2$ or the fact that $p_i^*>0$ for $i = 1,\ldots,\frac{n+k+4}{2}$. $a_i>0$ for $i = 1,\ldots,\frac{n+k+2}{2}$ when $k=n-2$. (iii) follows by inspection of the system of equations in (3.4.4).

Lemma 3.4.3. If k is even, k>n-2 , and $\mu^0_{\theta_\ell}$ is as described in theorem (3.4.1), then $\mu^0_{\theta_\ell}\{x_\nu\}>0$ for $\nu = 1,\ldots,n+k+2$; where the $\{x_\nu\}_{\nu=1}^{n+k+3}$ are the ordered points of $E^1_{n,k}$.

If k is even and k=n-2, then $\mu^0_{\theta_\ell}\{x_\nu\}>0$ for $\nu = 1,\ldots,\frac{n+k+4}{2}$, and $\mu^0_{\theta_\ell}\{x_\nu\} = 0$ for $\nu = \frac{n+k+6}{2},\ldots,n+k+3$.

Proof: Assume k is even and k>n-2. If $\mu^0_{\theta_\ell}\{x_{\nu_0}\}=0$ for some $\nu_0=1,\ldots,n+k+2$, we are led to a nontrivial polynomial P(x) in the functions (3.4.2) that has (n+k+1) distinct zeros in [-1,1]. $P(x_\nu)=0$ for $\nu\neq\nu_0$ or ν_{n+k+3}. Thus P(x), if nontrivial , must be nontrivial in both [-1,0] and [0,1]. The coefficient of $(x+1)^\ell$ in P(x) is zero. P(x) must have an additional zero in $(-\infty,1]$ since the ℓth derivative of P(x) must be zero at x=-1. For P(x) to be nontrivial throughout (-1,1] and have (n+k+2) zeros there, the coefficient of $(x+1)^n$ must be non-zero. Thus we may normalize P(x) so that its coefficient of $(x+1)^n$ is unity. Due to the spacing of the zeros of P(x), it must be true that $W^1_{n,k}(x)-P(x)$ must have at least (n+k+1) distinct zeros. Lemma (2.1.5) implies that $W^1_{n,k}(x)-P(x)\equiv 0$. This contradiction implies that $P(x)\equiv 0$. Thus it must be that $\mu^0_{\theta_\ell}\{x_\nu\}>0$.

If k is even and k=n-2, it is easily seen that a nontrivial P(x) exists, with the appropriate zeros, when solving for $\mu^0_{\theta_\ell}\{x_\nu\}$ when

$\nu=\frac{n+k+6}{2}, \ldots, n+k+2$. One has the k+1 functions, $\{x_+^i\}_{i=n-k-1}^{n-1}$, to form a P(x) with k zeros in (0,1). In this case, $\mu_{\theta_\ell}^0\{x_\nu\}>0$ for $\nu = 1,\ldots,\frac{n+k+4}{2}$, by the preceding argument.

Example 3.4.1. Consider a random variable Y(x) with mean $E\ Y(x)=\theta_0+\theta_1(x+1)+\theta_2(x+1)^2+\beta_1 x_+$ where $x\varepsilon[-1,1]$. For this example, $W_{2,0}^1(x)=1-8(x+1)+8(x+1)^2-16x_+$ and $E_{2,0}^1=\{-1,-1/2,0,1/2,1\}$. The optimal designs for estimating θ_1, given as vectors of weights on the points $E_{2,0}^1$, are

$$q(3/8,4/8,1/8,0,0)+(1-q)(1/8,0,1/8,4/8,2/8), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_2 are

$$q(1/4,1/2,1/4,0,0)+(1-q)(0,0,1/4,1/2,1/4), \quad \text{for } 0\leq q\leq 1.$$

The optimal designs for estimating β_1 are

$$q(1/8,3/8,3/8,1/8,0)+(1-q)(0,1/8,3/8,3/8,1/8), \text{ for } 0\leq q\leq 1.$$

The optimal design for θ_0 is

$$(1,0,0,0,0).$$

Example 3.4.2. Consider a random variable Y(x) with mean $E\ Y(x)=\theta_0+\theta_1(x+1)+\theta_2(x+1)^2+\theta_3(x+1)^3+\beta_2 x_+^2$ where $x\varepsilon[-1,1]$. For this example, $W_{3,0}^1(x)=-1+\frac{9}{2}(2+\sqrt{3})(y+1)-(21+12\sqrt{3})(y+1)^2+\frac{(26+15\sqrt{3})}{2}(y+1)^3-3(12+7\sqrt{3})y_+^2$ and $E_{3,0}^1 = \{-1,-(\sqrt{3}-1),-(3\sqrt{3}-5),(3\sqrt{3}-5),(\sqrt{3}-1),1\}$. The optimal designs for estimating θ_1, given as vectors of weights on the points $E^1_{3,0}$ to 5 decimal places, are

$$q(.35584,.45189,.14071,.04811,.00345,0)+$$

$$(1-q)(.20504,.16038,0,.18882,.29496,.15080), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_2 are

$$q(.23020,.403778,.26289,.09622,.00691,0)+$$

$$(1-q)(.02132,0,.06800,.29112,.41068,.20888), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_3 are

$$q(.17863,.35566,.31100,.14434,.01036,0)+$$

$$(1-q)(0,.01036,.14434,.31100,.35566,.17863), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating β_2 are

$$q(.11909,.27233,.33878,.22767,.04213,0)+$$

$$(1-q)(0,.04213,.22767,.33878,.27233,.11909), \text{ for } 0\leq q\leq 1.$$

The optimal design for θ_0 is

$$(1,0,0,0,0,0).$$

Example 3.4.3. Consider a random variable $Y(x)$ with mean $E\,Y(x)=\theta_0+\theta_1(x+1)+\theta_2(x+1)^2+\theta_3(x+1)^3+\theta_4(x+1)^4+\beta_3 x_+^3$ where $x\varepsilon[-1,1]$. For this example, $W^1_{4,0}(x)=1-8(2+\sqrt{2})(x+1)+5(12+8\sqrt{2})(x+1)^2-8(10+7\sqrt{2})(x+1)^3+2(17+12\sqrt{2})(x+1)^4-16(7+5\sqrt{2})x_+^3$ and $E^1_{4,0}=\{-1,2-2\sqrt{2},1-\sqrt{2},0,\sqrt{2}-1,2\sqrt{2}-2,1\}$. The optimal designs for estimating θ_1, given as vectors of weights on the points $E^1_{4,0}$ to 5 decimal places, are

$$q(.34911,.43756,.13363,.06028,.01726,.00216,0)+$$

$$(1-q)(.25829,.26112,0,.06028,.15089,.17860,.09082), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_2 are

$$q(.22322,.37764,.24226,.11805,.03452,.00431,0) +$$

$$(1-q)(.05858,.05777,0,.11805,.27678,.32419,.16464), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_3 are

$$q(.17234,.32655,.27589,.16697,.05178,.00647,0) +$$

$$(1-q)(.00425,0,.02857,.16697,.29910,.33303,.16808), \text{ for } 0\leq q\leq 1.$$

The optimal designs for estimating θ_4 and β_3 can be found in example (3.2.3). The optimal design for θ_0 is

$$(1,0,0,0,0,0,0).$$

CHAPTER IV

SPECIAL CASES OF MONOSPLINE REGRESSION WITH NONCENTERED KNOTS

4.1. Introduction with Background Theorems.

In section two of this chapter we will be concerned with the regression functions in (3.2), (3.3) and (3.4) on a closed subinterval of [-1,1] with k even. This subinterval does not include the point {+1} but depends on n and k in such a manner that it includes the (n+k+2) remaining points of $E^2_{n,k}$. In other words, the interval is nearly symmetric, $[-1,1-\varepsilon]$, where $\varepsilon>0$ and small. We also consider the interval $[-1+\varepsilon,1]$ for the regression functions in (3.2) and (3.3).

In section three we consider polynomial monospline regression of the form

4.1.1. $$\sum_{i=0}^{n} \theta_i (x+1)^i + \sum_{i=1}^{k} \beta_i (x-\eta_i)_+^{n-1}$$

where $n\geq 2$, $x\varepsilon[-1,1]$ and $-1<\eta_1<\eta_2<\ldots<\eta_k<1$. We will call (4.1.1) a monospline of class (n,k) and denote this by $M_{n,k}(x)$. Johnson (1960, page 459) discusses the existance of a unique monospline $M^*_{n,k}(x)$ of the form (4.1.1). In his work, the set of parameters and the knots in

(4.1.1) are allowed to vary. We state his principal theorem.

Theorem 4.1.1. For each (n,k) there exists a unique monospline $M^*_{n,k}(x)$ of class (n,k) which deviates least from zero on [-1,1]. For $n \geq 2$, $M^*_{n,k}(x)$ achieves its maximum absolute deviation, with alternating signs, at precisely (n+2k+1) points of [-1,1], including both end-points, and this condition determines $M^*_{n,k}(x)$ uniquely.

We will refer to $M^*_{n,k}(x)$ as the Johnson monospline.

4.2. Nonsymmetrical Special Cases of Chapter III.

Let us consider a random variable, Y(x), with mean value

4.2.1. $$\sum_{i=0}^{n} \theta_i x^i + \sum_{i=n-1-k}^{n-1} \beta_i (x^i - 2x_+^i)$$

where $n-1-k \geq 1$, k is even and $x \epsilon [-1, c_{n,k}]$. The constant ,$c_{n,k}$,is chosen such that if $x_0 = \max\{x | x \epsilon E^1_{n,k} \cap [-1,1)\}$, then $x_0 < c_{n,k} < 1$. $E^1_{n,k}$ is defined in lemma (3.2.1). We note that there are (n+k+2) points of $E^1_{n,k}$ less than $c_{n,k}$. We will first establish the best minimizing polynomials, as was done in lemma (3.2.3), for the nonsymmetric interval $[-1, c_{n,k}]$. The minimizing polynomials are the same as those in lemma (3.2.3).

Lemma 4.2.1. Among all polynomials in the functions (3.2.2) where k is even,

(1) $W^1_{n,k}(x)/\beta_{n-2j}$ minimizes $\sup_{-1 \leq x \leq c_{n,k}} |f(x)|$ where f(x) is any

polynomial in (3.2.2) with the coefficient of x^{n-2j} unity for $j = 0,\ldots,[\frac{n}{2}]$, and

(2) $W^1_{n,k}(x)/\beta_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where $f(x)$ is any polynomial in (3.2.2) with the coefficient of $(x^{n-2j-1} - 2x_+^{n-2j})$ unity for $j = 0,\ldots,[\frac{k}{2}]$.

Proof: Assume we have a $P_j(x)$ such that $\sup_{-1\leq x\leq c_{n,k}} |P_j(x)| < \sup_{-1\leq x\leq c_{n,k}} W^1_{n,k}(x)/\beta_{n-2j}$ where $P_j(x)$ is a polynomial in (3.2.2) with the coefficient of x^{n-2j} unity. Assume j is zero. The difference $P_0(x) - W^1_{n,k}(x)/\beta_n$ has at least $(n+k+1)$ distinct zeros with at least $[\frac{n+k+2}{2}]$ in $[-1,0)$ and $[\frac{n+k}{2}]$ in $(0,1]$. By lemma (2.1.5), the above is identically equal to zero. This implies that $P_0(x) \equiv W^1_{n,k}(x)/\beta_n$.

Assume n is even and let

$$F(x) = \frac{P_j(x) + P_j(-x)}{2}.$$

The coefficient of x^n in $F(x)$, say a_n, is either zero or $|a_n| < |\frac{\beta_n}{\beta_{n-2j}}|$.

Since $\sup_{-1\leq x\leq c_{n,k}} |P_j(x)| < \sup_{-1\leq x\leq c_{n,k}} \left| W^1_{n,k}(x)/\beta_{n-2j}\right|$ and

$\sup_{-1\leq x\leq c_{n,k}} \left|\frac{P_j(x)}{a_n}\right| \geq \sup_{-1\leq x\leq c_{n,k}} \left| W^1_{n,k}(x)/\beta_n\right|$, we have that $|a_n| < |\frac{\beta_n}{\beta_{n-2j}}|$.

The coefficient of x^n in $P_j(x)$ is also a_n since we assumed that n was even. This implies that the difference $F(x)-\frac{W^1_{n,k}(x)}{\beta_{n-2j}}$ has at least (n+k+2) distinct zeros with at least $(\frac{n+k+2}{2})$ in both $(-\infty,0)$ and $(0,+\infty)$. The difference can be written in the form

$$4.2.2. \qquad \sum_{j=0}^{\frac{n}{2}} a_{2j}x^{2j} + \sum_{j=0}^{\frac{k}{2}} b_{n-2j-1}(x^{n-2j-1}+2x_+^{n-2j-1}),$$

with $a_{2j}=0$ for some $j = 1,\ldots,\frac{n-2}{2}$. (4.2.2) is nontrivial on any interval and can have at most $(\frac{n+k}{2})$ distinct zeros in either $(-\infty,0)$ or $(0,\infty)$. This implies that (4.2.2) is identically equal to zero and $F(x)\equiv W^1_{n,k}(x)/\beta_{n-2j}$. This contradicts the fact that

$$\sup_{-c_{n,k}\le x\le c_{n,k}} |F(x)| < \sup_{-c_{n,k}\le x\le c_{n,k}} |W^1_{n,k}(x)/\beta_{n-2j}| .$$ Thus (1) is proven for n even. The case where n is odd follows a similar argument with $F(x)=P_j(x)-P_j(-x)$. (2) follows the identical argument.

<u>Lemma 4.2.2.</u> Lemma (4.2.1) holds where we consider the interval $[-c_{n,k},1]$.

Proof: The symmetry in the above arguments establishes the proof.

We can now obtain the optimal designs for the parameters in (4.2.1) that correspond to functions of the same parity as x^n.

Theorem 4.2.1. The optimal designs for estimating the following parameters in (4.2.1) with k even,

$$4.2.3. \quad \begin{cases} \theta_{n-2j} & \text{for } j = 0,\ldots,[\frac{n}{2}] \quad n-2j \neq 0 \\ \beta_{n-2j-1} & \text{for } j = 0,\ldots,\frac{k}{2}, \end{cases}$$

are unique and have their supports contained in the set $\{E^1_{n,k}-\{1\}\}$.

Proof: The proof of this theorem follows that of theorem (3.2.1) and lemma (3.2.6). We note here that the system of equations (3.2.14) is just Elfving's theorem (theorem (3.1.1)) applied to this situation. The unique optimal designs are the $\mu^0_{\theta_\ell}$ (or $\mu^0_{\beta_h}$) of theorem (3.2.1).

The symmetry in this last theorem allows us to state the following:

Theorem 4.2.2. Theorem (4.2.1) holds when we consider the interval $[-c_{n,k},1]$ and the points $\{E^1_{n,k}-\{-1\}\}$. The unique optimal designs are the $\mu^1_{\theta_\ell}$ (or $\mu^1_{\beta_h}$) of theorem (3.2.1).

A similar procedure can be followed for the functions in (3.3.2) after we have the following lemma. We now assume the functions in (3.3.2) and the mean value (3.3.1) are defined for the interval $[-1,c_{n,k}]$ with k even.

Lemma 4.2.3. Among all polynomials f(x) in the functions (3.3.2) with k even,

(1) $W^1_{n,k}(x)/\beta_{n-2j}$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where f(x) is any

polynomial in (3.3.2) with the coefficient of x^{n-2j} unity for $j = 0,\dots,[\frac{n+1}{2}]$,

(2) $W^1_{n,k}(x)/\beta_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where $f(x)$ is any polynomial in (3.3.2) with the coefficient of x^{n-2j-1} unity for $j = 0,\dots,\frac{k}{2}$, and

(3) $W^1_{n,k}(x)/-2\beta_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where $f(x)$ is any polynomial in (3.3.2) with the coefficient of x_+^{n-2j-1} unity for $j = 0,\dots,\frac{k}{2}$.

Proof: The cases for which the coefficient of x^{n-2i} is unity for $i = [\frac{n-k-3}{2}],\dots,[\frac{n}{2}]$ and $i=0$ follow the same arguments as in lemma (4.2.1), since the remaining functions span the same space.

Assume now that we are minimizing some one of the remaining functions listed in (1), (2) or (3), and let $P_1(x)$ be a polynomial in (3.3.2), with the appropriate coefficient unity, which corresponds to a better approximation. The coefficient of x^n in $P_1(x)$, say a_n, is either zero or $|a_n|<|\frac{\beta_n}{\lambda}|$, where λ is the coefficient of the appropriate function we are considering in $W^1_{n,k}(x)$. Since

$$\sup_{-1\leq x\leq c_{n,k}} |P_1(x)| < \sup_{-1\leq x\leq c_{n,k}} |W^1_{n,k}(x)/\lambda| \text{ and}$$
$$\sup_{-1\leq x\leq c_{n,k}} |P_1(x)/a_n| \geq \sup_{-1\leq x\leq c_{n,k}} |W^1_{n,k}(x)/\beta_n| \text{ , we have that}$$

$|a_n| \le |\beta_n/\lambda|$. This implies that the difference

$$W(x)/\lambda - P_1(x)$$

has at least (n+k+2) distinct zeros, with at least $[\frac{n+2k+2}{2}]$ distinct zeros in both [-1,0) and (0,1], and is nontrivial in any interval. Lemmas (3.3.1) and (3.3.2) imply that $P_1(x) \equiv W(x)/\lambda$, and this contradicts the assumption that $P_1(x)$ was better. This proves the lemma.

Lemma 4.2.4. Lemma (4.2.3) holds when we consider the interval $[-c_{n,k},1]$.

Proof: Again the symmetry of the above arguments establishes the proof.

We now state two theorems without proof since their proofs would be repetitious.

Theorem 4.2.3. The optimal designs for estimating the following parameters in (3.3.1), when k is even and $x \varepsilon [-1, c_{n,k}]$,

$$\text{4.2.4.} \quad \begin{cases} \theta_{n-2j} & \text{for } j = 0,\ldots,[\frac{n}{2}],\ n-2j \neq 0 \\ \theta_{n-2j-1} & \text{for } j = 0,\ldots,\frac{k}{2}, \\ \beta_{n-2j-1} & \text{for } j = 0,\ldots,\frac{k}{2}, \end{cases}$$

are unique and have their supports contained in the set $\{E^1_{n,k} - \{1\}\}$. The unique optimal designs are the $\mu^0_{\theta_\ell}$ (or $\mu^0_{\beta_h}$) of theorem (3.3.1).

Theorem 4.2.4. Theorem (4.2.3.) holds when we consider the interval $[-c_{n,k},1]$ and the points $\{E^1_{n,k}-\{-1\}\}$. The unique optimal designs are the $\mu^1_{\theta_\ell}$ (or $\mu^1_{\beta_h}$) of theorem (3.3.1).

The symmetry inherent in the functions considered in sections 3.2 and 3.3 does not carry over completely to those of 3.4 . The examples in 3.4 and the following show that we have similarities for $[-1,c_{n,k}]$ but not for $[-c_{n,k},1]$.

Lemma 4.2.5. Among all polynomials f(x) in the functions (3.4.2) with k even,

(1) $W^1_{n,k}(x)/\lambda_j$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where f(x) is any polynomial in (3.4.2) with the coefficient of $(x+1)^j$ unity for $j = 0,\ldots,n$, and

(2) $W^1_{n,k}(x)/\gamma_{n-2j-1}$ minimizes $\sup_{-1\leq x\leq c_{n,k}} |f(x)|$ where f(x) is any polynomial in (3.4.2) with the coefficient of x_+^{n-2j-1} unity for $j = 0,\ldots,\frac{k}{2}$.

Proof: For the case $(x+1)^n$, assume that $P_1(x)$ is a polynomial in (3.4.2) with the coefficient of $(x+1)^n$ unity whose norm is less than that of $W^1_{n,k}(x)/\lambda_n$. Thus, $W^1_{n,k}(x)/\lambda_n - P_1(x)$ has at least (n+k+1) distinct zeros, does not vanish in any interval and can be put in the form $\sum_{i=0}^{n-1} a_i x^i + \sum_{i=n-1-k}^{n-1} b_i x_+^i$. The maximum number of possible zeros is

$(n+k)$. This implies that $W^1_{n,k}(x)/\lambda_n \equiv P_1(x)$ and gives the desired contradiction.

For the cases x_+^{n-2j-1}, assume that $P_j(x)$ is a polynomial in (3.4.2) whose norm is less than that of $W^1_{n,k}(x)/-2\beta_{n-2j-1}$. Let $a_{n,j}$ be the coefficient of x^n in $P_j(x)$. By an argument similar to that in lemma (4.2.3), we have that $a_n=0$ or $|a_n|<\left|\frac{\beta_n}{-2\beta_{n-2j-1}}\right|$. In either case, $W^1_{n,k}(x)/-2\beta_{n-2j-1} - P_j(x)$ has at least $(n+k+2)$ distinct zeros in $[-1,\infty)$. The hypothesis of lemma (3.3.1) is satisfied, so lemma (3.3.2) leads to the desired contradiction.

For the cases $(x+1)^j$, $j = 1,\dots,n-1$, we follow reasoning as with x_+^{n-2j-1} up to the point where we note that the difference of the two approximations has $(n+k+2)$ distinct zeros in $[-1,\infty)$. The difference does not vanish identically in any interval, and the $(n+k+2)$ distinct zeros are the maximal number. Expanding the difference in a Taylor series about $x=-1$, we must have no zero derivatives of all orders up to n, so the coefficient of $(x+1)^j$, $j = 1,\dots,n$, must be non-zero. This contradicts the fact that the difference does not contain a term corresponding to $(x+1)^j$ and completes the proof.

Lemma 4.2.6. Lemma (4.2.5) holds for the functions x_+^{n-2j-1}, $j = 0,\dots,\frac{k}{2}$, and $(x+1)^n$ when we consider the interval $[-c_{n,k},1]$.

We now know that the following is true:

Theorem 4.2.5. The optimal designs for estimating the following parameters in (3.4.1), when k is even and $x\varepsilon[-1,c_{n,k}]$,

$$4.2.5. \quad \begin{cases} \theta_i & \text{for } i = 1,\ldots,n \\ \beta_{n-2j-1} & \text{for } j = 0,\ldots,\frac{k}{2}, \end{cases}$$

are unique and have their supports contained in the set $\{E^1_{n,k}-\{+1\}\}$. The unique optimal designs are the $\mu^0_{\theta_\ell}$ (or $\mu^0_{\beta_h}$) of theorem (3.4.1).

Due to the partial symmetry the following holds:

Theorem 4.2.6. The optimal designs for estimating the following parameters in (3.4.1), when k is even and $x\varepsilon[-c_{n,k},1]$,

$$\{\theta_n \text{ and } \beta_{n-2j-1} \quad \text{for } j = 0,\ldots,\tfrac{k}{2}\}$$

are unique and have their supports contained in the set $\{E^1_{n,k}-\{-1\}\}$. The unique optimal designs are the $\mu^1_{\theta_\ell}$ (or $\mu^1_{\beta_h}$) of theorem (3.4.1).

4.3. Optimal Designs for the Johnson Monosplines.

For a given (n,k) let $\eta_1,\ldots,\eta_k$, be the knots of $M^*_{n,k}$ in theorem (4.1.1) (the Johnson monospline). As regression functions for a given (n,k), let us consider the linearly independent and continuous functions

$$4.3.1. \qquad \bar{b}(x) = \begin{cases} 1,(x+1),\ldots,(x+1)^n \\ (x-\eta_i)^{n-1}_+ \quad i = 1,\ldots,k \\ x\varepsilon[-1,1]. \end{cases}$$

In this section we are considering a random variable, $Y(x)$, with mean

4.3.2. $$\sum_{i=0}^{n} \theta_i \, (x+1)^i + \sum_{i=1}^{k} \beta_i (x-\eta_i)_+^{n-1}$$

where $x \varepsilon [-1,1]$ and $n-1 \geq 1$. In order to classify the optimal designs for the parameters in (4.3.2), we will first establish the best minimizing polynomials as was done in chapter III.

Lemma 4.3.1. There exists a unique polynomial $W_{n,k}^3(x)$ (a linear combination of the functions in (4.3.1)) satisfying:

(1) $|W_{n,k}^3(x)| \leq 1$ for $x \varepsilon [-1,1]$;

(2) The set $E_{n,k}^3 = \{x: |W_{n,k}^3(x)| = 1\}$ contains exactly $(n+2k+1)$ points including both $\{-1\}$ and $\{1\}$;

(3) $W_{n,k}^3(x)$ attains its supremum at each of the points of the set $E_{n,k}^3$ with alternating signs;

(4) $W_{n,k}^3(x)$ is of the form

$$\sum_{j=0}^{n} \lambda_j (x+1)^j + \sum_{j=n+1}^{n+k} \lambda_j (x-\eta_{n-j})_+^{n-1}$$

where all the λ_j are non zero; and

(5) $W_{n,k}^3(x) = M_{n,k}^*(x)/||M_{n,k}^*(x)||$.

Proof: Define $W_{n,k}^3(x) = M_{n,k}^*(x)/||M_{n,k}^*(x)||$. (1), (2) and (3) follow from theorem (4.1.1). Since $W_{n,k}^3(x)$ alternates $(n+2k)$ times in $[-1,1]$ and achieves its maximum at $\{-1\}$ and $\{+1\}$, it has its maximal number of (n+2k) zeros in (-1,1). Let us expand $W^3_{n,k}(x)$ in its Taylor series about {-1}. $W^3_{n,k}(-1)\neq 0$. The jth derivative of $W^3_{n,k}(x)$, say $W^{(j)}(x)$, exists for j = 1,...,n. If $n-3\geq 0$, $W^{(1)}(x)$ exists for all $x\varepsilon[1,1]$. Applying Rolle's theorem, $W^{(1)}(x)$ has (n+2k-1) distinct zeros which are separated by those of $W^{(0)}(x)$. This implies that $W^{(1)}(-1)\neq 0$. We can repeat this argument for $W^{(\ell)}(-1)\neq 0$, $\ell = 1,\ldots,n-2$; $n-2\geq 0$. Since $W^{(n-2)}(x)$ has (2k+2) distinct zeros in (-1,1) and is nontrivial in every subinterval of (-1,1), it must have at most two distinct zeros in $(-1,\eta_1]$. Therefore, $W^{(n-1)}(-1)\neq 0$ and $W^{(n)}(-1)\neq 0$. If any one of the λ_j for j = n+1,...,n+k+1, is zero, then $W^3_{n,k}(x)$ could not have its stated property of (n+2k) distinct zeros since it is nontrivial on every subinterval.

Lemma 4.3.2. Among all polynomials f(x) in the functions (4.3.1) defined on [-1,1] with $n\geq 2$, $W^3_{n,k}(x)/\lambda_j$ minimizes $\sup_{-1\leq x\leq 1} |f(x)|$ where f(x) is any polynomial in (4.3.1) with the coefficient of $(x+1)^j$ unity for j = 1,...,n, or the coefficient of $(x-\eta_{n-j})^{n-1}_+$ unity for j = n+1,...,n+k.

Proof: Assume for some j = 0,...,n, that there is a better approximation of $(x+1)^j$ than the one suggested by $W^3_{n,k}(x)/\lambda_j$. Let $P_j(x)$ represent this better minimizing polynomial. Since

$$\sup_{-1\leq x\leq 1} |P_j(x)| < \sup_{-1\leq x\leq 1} |W^3_{n,k}(x)/\lambda_j| ,$$

$W^3_{n,k}(x)/\lambda_j - P_j(x)$ is a polynomial in (4.3.1) with (n+k+2) distinct zeros, not vanishing identically on any interval and having the coefficient of $(x+1)^j$ zero. By repeated application of Rolle's theorem as in lemma (4.3.1), we find that the coefficient of $(x+1)^j$ in the difference must be non-zero. This gives the desired contradiction.

If we consider a similar argument for $(x+\eta_{n-j})_+^{n-1}$ for some $j = n+1,\ldots,n+k$, we would find that

$$W^3_{n,k}(x)/\lambda_j - P_j(x)$$

could have at most (n+2(k-1)) zeros and not vanish identically on any interval of [-1,1]. However, the requirement that it have at least (n+2k) distinct zeros leads to the desired contradiction.

We are now able to classify the $\bar{c}_p$-optimal designs for the parameters of the functions in (4.3.2). We note that the optimal designs for all the parameters have their supports contained in the same set $E^3_{n,k}$.

Theorem 4.3.1. The optimal designs for estimating the following parameters in (4.3.2),

$$\text{4.3.3.} \qquad \begin{cases} \theta_i, \ i = 0,\ldots,n \\ \beta_i, \ i = 1,\ldots,k, \end{cases}$$

have their supports contained in the set $E^3_{n,k}$ of (n+2k+1) points. ($E^3_{n,k}$ is defined in lemma (4.3.1).) The optimal designs for each parameter, when expressed as vectors of weights on the points $E^3_{n,k}$, lie on k-dimensional planes. These planes are parallel.

Proof: By lemma (4.3.2) and the use of (3.1.3) as in theorem (3.2.1), we have that any optimal design for the parameters in (4.3.3) has its support contained in the set $E^3_{n,k}$. To find the $\bar{c}_p$-optimal designs, Elfving's theorem (theorem 3.1.1) tells us there is a solution to the system

4.3.4. $$\left|\frac{1}{\lambda_p}\right| \bar{c}_p = \sum_{\nu=1}^{n+2+k} \varepsilon_\nu p_\nu \bar{b}(x_\nu),$$

where the λ_p and $x_\nu \in E^3_{n,k}$ are defined in lemma (4.3.2), $\sum_{\nu=1}^{n+2k} p_\nu = 1$, $p_\nu \geq 0$ and $\varepsilon_\nu = \pm 1$. The system (4.3.4) describes (n+k+1) equations in (n+2k+1) unknowns. The rank of the system is (n+k+1). If this were not true, then a nontrivial linear combination of the rows of the coefficient matrix would yield a polynomial with (n+2k+1) distinct zeros. These zeros are the points of $E^3_{n,k}$. Lemma (2.1.5) implies that only a trivial linear combination can have these zeros, so the rank is (n+k+1). Thus we have a k-dimensional set of solutions. The coefficient matrix of the system (4.3.4), aside from a multiplicative constant (±1), is the same for each p. This implies that the k-dimensional sets are parallel.

Example 4.3.1. Consider a random variable, Y(x), with mean $E\,Y(x)=\theta_0+\theta_1(x+1)+\theta_2(x+1)^2+\beta_1(x+1/3)_+ +\beta_2(x-1/3)_+$ where $x\varepsilon[-1,1]$. For this example, $W^3_{2,2}(x)=1-12(x+1)+18(x+1)^2-24(x+1/3)_+ -24(x-1/3)_+$ and

$E^3_{2,2}=\{-1,-2/3,-1/3,0,1/3,2/3,1\}$. The optimal designs for estimating θ_1, given as vectors of weights on the points $E^3_{2,2}$, are convex combinations of (3/8,4/8,1/8,0,0,0,0), (1/8,0,1/8,4/8,2/8,0,0) and (1/8,0,0,2/8,2/8,2/8,1/8). The optimal designs for estimating θ_2 are convex combinations of (1/4,2/4,1/4,0,0,0,0), (0,0,1/4,2/4,1/4,0,0) and (0,0,0,0,1/4,2/4,1/4). The optimal designs for estimating β_1 are convex combinations of (1/8,3/8,3/8,1/8,0,0,0), (0,1/8,3/8,3/8,1/8,0,0) and (0,2/16,3/16,0,2/16,6/16,3/16). The optimal designs for estimating β_2 are convex combinations of (3/16,6/16,2/16,0,3/16,2/16,0), (0,0,1/8,3/8,3/8,1/8,0) and (0,0,0,1/8,3/8,3/8,1/8). The optimal design for estimating θ_0 is

(1,0,0,0,0,0,0).

CHAPTER V

SOME EXTRAPOLATION AND MINIMAX EXTRAPOLATION DESIGNS

5.1. Introduction

When we consider the problem of estimating the regression of the form(3.3.1)

$$\sum_{i=0}^{n} \theta_i x^i + \sum_{i=n-1-k}^{n-1} \beta_i x_+^i$$

at a point x_0 outside of [-1,1] by observations restricted to points of [-1,1], we have an extrapolation problem. For a given design or probability measure μ on [-1,1], the variance of the best linear unbiased estimate of

$$\sum_{i=0}^{n} \theta_i x_0^i + \sum_{i=n-1-k}^{n-1} \beta_i x_{0+}^i$$

is proportional to (see (1.1.5))

$$V(x_0,\mu) = \sup_{\overline{b}} \frac{(\overline{b},\overline{b}(x_0))^2}{\int(\overline{b},\overline{b}(x))^2 d\mu(x)} .$$

$\overline{b}(x)$ is defined in (3.3.2). A design μ^* is said to be optimal for extrapolating to x_0 if it minimizes $V(x_0,\mu)$. In section 5.2, we consider some extrapolation problems that include regression functions somewhat more general than those of $\overline{b}(x)$ but include $\overline{b}(x)$ as a special case.

In section 5.3, we consider minimax extrapolation designs for regression of the form (3.3.1). A design μ^* is a minimax extrapolation design for $t\varepsilon[1,e]$ if

$$\min_{\mu} \max_{t\varepsilon[1,e]} V(t,\mu) = \max_{t\varepsilon[1,e]} V(t,\mu^*).$$

A minimax extrapolation design for $[-e,-1]$ is defined in a similar manner.

5.2. Extrapolation Designs

In this section, we consider the linearly independent and continuous regression functions

5.2.1. $$\begin{cases} 1,x,\ldots,x^n \\ (x-\xi_i)_+^{n-1-k_i},\ldots,(x-\xi_i)_+^{n-1}, \quad \xi_i \text{ fixed } i = 1,\ldots,h \end{cases}$$

where $n-1-k_i \geq 1$, $a<\xi_1<\ldots<\xi_h<b$, and $x\varepsilon[a,b]$. Let $m(x)$ be any polynomial (linear combination) in the functions in (5.2.1) and define $||m(x)|| = \sup_{a\leq x\leq b} |m(x)|$. Let $W(x)$ be a polynomial in (5.2.1) such that $||W(x)||=1$ and $W(x)/\beta_n$ minimizes $\sup_{a\leq x\leq b} |f(x)|$ where $f(x)$ is any polynomial in (5.2.1) with the coefficient of x^n unity. The coefficient of x^n in $W(x)/\beta_n$ is unity. Such a $W(x)$ exists. Meinardus (1967, page 1).

Lemma 5.2.1. Among all polynomials $m(x)$ in (5.2.1) such that $||m(x)||=1$, $W(x)$ has the largest coefficient of x^n in absolute value.

Proof: Assume for some $m(x)$ satisfying $||m(x)||=1$, whose coefficient of x^n is θ_n, that we have $|\theta_n|>|\beta_n|$. This implies that

$$\left|\left|\frac{m(x)}{\theta_n}\right|\right|<\left|\left|\frac{W(x)}{\beta_n}\right|\right|$$

which contradicts the minimizing properties of $W(x)/\beta_n$.

Lemma 5.2.2. If

(i) $W(x)$ alternates at least $\sum_{j=1}^{i} (k_i+2)$ times in $[a,\xi_i]$;

(ii) $W(x)$ alternates at most $n + \sum_{j=1}^{i-1} (k_i+2)$ times in $[a,\xi_i]$;

(iii) $W(x)$ alternates $n + \sum_{i=1}^{h} (k_i+2)$ times in $[a,b]$; and

(iv) the alternating points of $W(x)$ include $\{a\}$ and $\{b\}$;

then

$W(x)$ has the property that $|W(x)|\geq|m(x)|$ for all $x\leq a$ or $x\geq b$.

Proof: Assume that there exists a point $x_0>b$ and an $m(x)$ such that $|m(x_0)|>|W(x_0)|$ where $||m(x)||=1$. Let $|m(x_0)|-|W(x_0)|=k>0$. There exists an $\varepsilon>0$, $1>\varepsilon>0$, such that $(1-\varepsilon)|m(x_0)|-|W(x_0)|>0$. That is, ε is so small that $k-\varepsilon|m(x_0)|>0$. Without loss of generality, we may assume that $m(x_0)>W(x_0)>0$.

By lemma (5.2.1), $\lim_{x\to+\infty} \frac{|m(x)|}{|W(x)|} \leq 1$ so that $\lim_{x\to+\infty} \frac{(1-\varepsilon)|m(x)|}{|W(x)|} < 1$.

This implies that there is an $s>x_0$ such that $(1-\varepsilon)m(s)<W(s)$. So

there must be a zero if $W(x)-(1-\varepsilon)m(x)$ in (x_0,s). Since $||(1-\varepsilon)m(x)||<1$, there is a zero of the difference in (b,x_0), as well as $(n + \sum_{j=1}^{n} (k_i+2))$ distinct zeros in (a,b). These $(n+2+\sum_{i=1}^{h}(k_i+2))$ distinct zeros are situated so that lemma (2.1.5) implies $W(x)-(1-\varepsilon)m(x)\equiv 0$, the desired contradiction. A similar argument holds if $x_0\leq a$.

Let us consider a random variable, $Y(x)$, with mean

5.2.2. $$\sum_{i=0}^{n} \theta_i x^i + \sum_{i=1}^{h} \sum_{j=n-k_i-1}^{n-1} \theta_{ij} (x-\xi_i)_+^j$$

where $n-1-k_i\geq 1$, $a<\xi_1<\ldots<\xi_h<b$, and $x\varepsilon[a,b]$.

Theorem 5.2.1. If the $W(x)$ of lemma (5.2.1) satisfies the conditions (1) thru (4) of lemma (5.2.2), then the optimal designs for extrapolating to x_0, where $x_0\geq b$ (or $x_0\leq a$), and the optimal designs for estimating θ_n in (5.2.2), have their supports contained in the same set E of the $(n+1+\sum_{i=1}^{h} (k_i+2))$ alternating points of $W(x)$.

Proof: An argument similar to that in theorem (3.3.1) shows that the optimal designs for θ_n have their supports contained in the set E.

Let $\overline{f}(x)=(1,x,\ldots,x^n,(x-\xi_1)_+^{n-1-k_1},\ldots,(x-\xi_1)_+^{n-1},\ldots,(x-\xi_h)_+^{n-1-k_n},\ldots,(x-\xi_h)_+^{n-1})$. By the discussion in sections 3.1 and 5.1, we have that

$$\inf_{\mu} V(x_0,\mu) = \sup_{\overline{b}}(\overline{f}(x_0),\overline{b})^2[\sup_{a\leq x\leq b} (\overline{b},\overline{f}(x))^2]^{-1}.$$

This implies

$$\inf_{\mu} V(x_0,\mu) = \sup_{\overline{b}}\{(\overline{f}(x_0),\overline{b})^2 \mid \sup_{a\leq x\leq b} (\overline{b},\overline{f}(x))^2=1\}$$

$$= (W(x_0))^2$$

by lemma (5.2.2).

Suppose μ_0^* is $\overline{f}(x_0)$ optimal. Then

$$V(x_0,\mu_0^*) = \sup_{\overline{b}}(\overline{f}(x_0),\overline{b})^2[\int(\overline{b},\overline{f}(x))^2\mu_0^*(dx)]^{-1}$$

$$\geq (W(x_0))^2[\int(W(x))^2\mu_0^*(dx)]^{-1}$$

$$\geq (W(x_0))^2.$$

Since $|W(x)|=1$ only for $x\varepsilon E$, strict inequality holds at the last step unless μ_0^* has its support contained in the set E.

The above theorem applies to the regression problems considered in sections 3.2 , 3.3 , 3.4 and 4.3 with only slight modifications. In the first three cases $W^1_{n,k}(x)$, and in the last $W^3_{n,k}(x)$, correspond to the function W(x) considered above.

<u>Corollary 5.2.1</u>. In theorem (5.2.1) ,let h=1, $\xi_1=0$, a=-1 and b=+1. If k_1 is odd, the optimal extrapolation design is unique and supported by the full set $E^1_{n,k}$. If k is even, the optimal extrapolation designs are any convex combination of two distinct probability measures. Any design not an endpoint of the convex combination is supported by the full set $E^1_{n,k}$.

Proof: According to Elfving's theorem (3.1.1), the optimal extrapolation designs are solutions of

$$5.2.3. \qquad \beta\, \overline{b}(x_0) = \sum_{\nu=1}^{n+2[\frac{k}{2}]+3} \varepsilon_\nu\, p_\nu\, \overline{b}(x_\nu)$$

where the β is an appropriate constant (the vector $\overline{b}(x)$ is defined in (3.3.2)), $x_\nu \in E^1_{n,k}$, $\varepsilon_\nu = \pm 1$ and $|x_0| \geq 1$. When k is odd and some p_ν in (5.2.3) is zero, Cramer's method of solution implies there is a polynomial in the functions (3.3.2) with (n+k+2) distinct zeros. x_0 is a zero, as are (n+k+1) points of $E^1_{n,k}$. This is clearly impossible and implies $p_\nu > 0$ for all $\nu = 1,\ldots,n+k+2$. The proof for k even follows a parallel argument to that of theorem (3.3.1)-(ii) and theorem (4.3.1).

Example 5.2.1. Consider a random variable, Y(x), with mean $E\,Y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \beta_1 x_+$ where x [-1,1]. For this example,

$W(x) = 1 + 16x + 8x^2 - 16x_+$ and $E = \{-1,-1/2,0,1/2,1\}$. The optimal designs for extrapolating to $x_0 = 2$, given as vectors of weights on the points E, are

$$q(4/17,8/17,3/17,0,2/17)+(1-q)(0,0,3/17,8/17,6/17) \text{ for } 0 \leq q \leq 1.$$

Theorem (5.2.1) holds in the above example, while in this next example the theorem does not apply.

Example 5.2.2. Consider a random variable, $Y(x)$, with mean $E\,Y(x)=\theta_0+\theta_1 x+\theta_2 x^2+\beta_1(x-1/4)_+$ where $x\varepsilon[1,1]$. For this example, there is a $W_2(x)$ corresponding to a best approximation of θ_2 such that $W_2(x) = -7/25+96/25x+128/25x^2-64/5\,(x-1/4)_+$, and the set of alternating points $E_2=\{-1,-3/8,1/4,7/8\}$. The optimal design for estimating θ_2, given as a vector of weights on E_2, is $(1/4,1/2,1/4,0)$. When extrapolating to $x_0=2$, the optimal extrapolation design is $(168/747,336/747,68/747,175/744)$ on the points $\{-1,-3/8,1/4,1\}$.

5.3. Minimax Extrapolation Designs

In this section we are concerned with the regression situation as defined in section 3.3 . However, there is a parallel minimax discussion for sections 3.2 and 3.4 .

Lemma 5.3.1. $|W^1_{n,k}(x)|$, as defined in lemma (3.2.1), is strictly increasing in x for $x\geq 1$ and strictly decreasing in x for $x\leq -1$.

Proof: Assume that n is even. Now $\frac{dW^1_{n,k}(x)}{dx}$ is non-zero for all $x\geq 1$. If zero, then Rolle's theorem would imply that there are at least $(n-1+2[\frac{k}{2}]+1)$ distinct zeros in $(-1,1)$, and since $W^1_{n,k}(x)$ is even, we would have at least $(n+2+2[\frac{k}{2}])$ distinct zeros of the derivative in $(-\infty,\infty)$. This is clearly impossible. Since the coefficient of x^n is positive in $W^1_{n,k}(x)$, its derivative is strictly positive in $[1,\infty)$ and strictly negative in $(-\infty,-1]$ by the symmetry. A similar argument would hold for n odd, except the derivative of $W^1_{n,k}(x)$ is strictly positive in both $(-\infty,-1]$ and $[1,\infty)$.

<u>Theorem 5.3.1</u>. The minimax extrapolation designs for [1,e], (or [-e,-1]) are the extrapolation designs of theorem (5.2.1) for the points e (or -e) in the setting of section 3.3 .

Proof:

$$\max_{1\le t\le e} \sup_{\bar{b}} \frac{(\bar{b},\bar{b}(t))^2}{\int(\bar{b},\bar{b}(x))d\mu(x)} \ge \max_{1\le t\le e} \frac{(W^1_{n,k}(t))^2}{\int(W^1_{n,k}(x))^2 d\mu(x)}$$

$$\ge \frac{(W^1_{n,k}(e))^2}{\int(W^1_{n,k}(x))^2 d\mu(x)} \ge (W^1_{n,k}(e))^2.$$

Equality is reached in all cases above after consideration of lemmas (5.3.1) and (5.2.2) by the extrapolation designs to e of theorem (5.2.1). A similar proof would follow for [-e,-1].

BIBLIOGRAPHY

Birkhoff, G. and DeBoor, C. (1965). Piecewise Polynomial Interpolation and Approximation, Approximation of Functions. (Edited by H. L. Garabedian) 164-190 Elsevier, New York.

Elfving, G. (1959). Optimum allocation in linear regression theory. Ann. Math. Statist. Vol. 23, 255-262.

Elfving, G. (159). Design of Linear Experiments, Probability and Statistics. (Edited by U. Grenander) 58-74 Wiley, New York.

Hoel, P. G. and Levine, A. (1964). Optimal spacing and weighing in polynomial prediction. Ann. Math. Statist. Vol. 35, 1553-1560.

Johnson, R. S. (1960). On monosplines of least deviation. Trans. Amer. Math. Soc. Vol. 96, 458-477.

Karlin, S. (1968). Total Positivity. Vol. I. Stanford University Press, California.

Karlin, S. and Schumaker, L. (1967). The fundamental theorem of Algebra for Tchebycheffian monosplines. Jour. d'Analyse Math. Vol. 20, 233-270.

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

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

Karlin, S. and Ziegler, Z. (1966). Tchebycheffian spline functions. Jour. SIAM Num. Anal. Vol. 3, 514-543.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser. B. Vol. 21, 273-319.

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

Meinardus, G. (1967). Approximation of Functions: Theory and Numerical Treatment. Vol. 13, Springer-Verlag, Heidleberg.

Murty, V. N. (1961). Optimal designs of individual regression coefficients with a Tchebycheffian spline regression function. Department of Statistics, Purdue University, Mimeograph Series No. 191, West Lafayette, Indiana.

Rice, J. R. (1969). The Approximation of Functions: Vol. II-Advanced Topics. Addison-Wesley, Massachusetts.

Schoenberg, I. J. (1964). On interpolation by spline functions and its minimal properties. International Series Numerical Math. Vol. 5, 109-129.

Studden, W. J. (1968). Optimal designs on Tchebycheff points. Ann. Math. Statist. Vol. 39, 1435-1447.

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

Synge, J. L. and Griffith, B. A. (1959). Principles of Mechanics. McGraw-Hill, New York.

Theilheimer, F. and Starkweather, W. (1961). The fairing of ship lines on a high-speed computer MTAC , Vol. 15, 338-355.

Van Arman, D. J. (1968). Classification of experimental designs relative to polynomial spline regression functions. Department of Statistics, Purdue University, Mimeograph Series No. 166, West Lafayette, Indiana.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE

Optimal and Admissible Designs for Polynomial Monospline Regression

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (Last name, first name, initial)

Bruvold, Norman T.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1971	119	21

8a. CONTRACT OR GRANT NO. N00014-67-A-0226-00014 and GP 20306	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series #253
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Also supported by the National Science Foundation, Washington, D. C.	Office of Naval Research Washington, D. C.

13. ABSTRACT

We consider regression of the form $\sum_{i=0}^{n} a_i x^i + \sum_{i=1}^{h} \sum_{j=\ell_i}^{k_i} b_{ij}(x-\xi_i)_+^{n-j}$ where $n-1 \geq k_i \geq \ell_i \geq 0$, $a<\xi_1<\ldots<\xi_h<b$ and $x\varepsilon[a,b]$. We define admissibility in terms of a positive semi-definite difference of information matrices. Some sufficient conditions for admissibility on the spectrum of a design are given.

When $\ell_1=1$, $h=1$ and ξ_1 lies in the center of the interval $[a,b]$, optimal experimental designs for the individual regression coefficients are given. Some of the optimal designs are not unique but are convex combinations of two probability measures. Three distinct bases are considered.

Extrapolation and minimax extrapolation designs are given for the centered knot situation along with some other special cases.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
experimental designs						
monospline regression						
admissibility						
extrapolation						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified [illegible] may obtain copies of this report from D[illegible]"

(2) "Foreign [illegible] and dissemination of this report by [illegible] authorized."

(3) "U. S. Gover[illegible] may obtain copies of this report [illegible] DDC. Other qualified DDC users shall [illegible] through

____."

(4) "U. S. milita[illegible] may obtain copies of this report direct[illegible]. Other qualified users shall request [illegible]

____."

(5) "All distribu[illegible] this report is controlled. Qualified DDC [illegible] request through

____."

If the report has [illegible] to the Office of Technical Services, Department [illegible], for sale to the public, indicate this fact and [illegible], if known.

11. SUPPLEMENTA[illegible]: Use for additional explanatory notes.

12. SPONSORING [illegible] ACTIVITY: Enter the name of the departmental [illegible] laboratory sponsoring *(paying for)* the research [illegible]. Include address.

13. ABSTRACT: [illegible] giving a brief and factual summary of the document [illegible] of the report, even though it may also appear [illegible] the body of the technical report. If additional [illegible] required, a continuation sheet shall be attached.

It is highly desir[illegible] the abstract of classified reports be unclassified. [illegible] of the abstract shall end with an indication of the [illegible] classification of the information in the para[illegible] represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limit[illegible] the length of the abstract. However, the suggested [illegible] from 150 to 225 words.

14. KEY WORDS: [illegible] are technically meaningful terms or short phrases that [illegible] a report and may be used as index entries for cat[illegible] the report. Key words must be selected so that no [illegible] classification is required. Identifiers, such as equipm[illegible] designation, trade name, military project code name, [illegible] location, may be used as key words but will be fo[illegible] indication of technical context. The assignment of [illegible], roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

D13551

Unclassified

Security Classification