Design Problems in Model Robust Regression and Exact D-Optimality

by

Mong-Na Lo Huang*

Technical Report #83-32

Department of Statistics
Purdue University

August 1983

*This research was partially supported by NSF Grant No. MCS-8200631.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1. MODEL ROBUST DESIGN IN REGRESSION

Let X be a subset of an Euclidean space. An observation at $x \in X$ is assumed to be of the form

$$y(x) = f(x) + \epsilon ,$$

where ϵ denotes a random variable with mean 0 and variance σ^2 for all x, and $f \in \Gamma$, a class of possible regressions over X.

A design problem, that is, placement of uncorrelated observations in X, for the estimation of regression parameters is discussed here.

The class Γ can be one of the standard classes, such as polynomials of fixed degree. Box and Draper(1959) have discussed some of the consequences of a strict formulation of Γ which ignores the possibility that the true f may only be approximated by an element of Γ. For example Γ may consist of the linear functions while the true f may be a quadratic, and in estimation this may result in a large bias term.

In the subsequent interesting papers, some deal with finite dimensional Γ, for example Karson, Manson and Hader(1969), Kiefer(1973,1980); some treat the problem with infinite dimensional Γ, such as Huber(1975), Marcus and Sacks(1976), Notz(1980), Li and Notz(1982), Pesotchinsky(1982), Li(1981), Spruill(1982), and Sacks and Ylvisaker(1982).

Huber(1975) addressed the curve fitting problem in the first part of his paper. Another approach to curve fitting is taken by Agarwal and Studden(1978).

In the last part of Huber's(1975) paper, he treats the extrapolation problem in a class of regression functions with bounded (h+1)th derivative. Two closely related papers are Sacks and Ylvisaker(1982) and Spruill(1982). In the former paper, the problem of estimating linear functionals is considered for Γ as in Huber's(1975). The latter treats the extrapolation problem on a closed interval [a,b], with f in the Sobolov space on interval [a,c], where c is the extrapolation point.

In the first part of this thesis, the extrapolation problem is discussed again for $X = [0,\infty)$ and $X = [-1,1]$, and the design problem for the estimation of $f^{(k)}(0)$ for f in Γ and X on $[0,\infty)$ is also discussed. Those discussions will be found in Chapter II and Chapter III.

For more detailed discussions and further references, see Huber(1975) and Sacks and Ylvisaker(1982).

1.2. EXACT D-OPTIMAL DESIGN FOR POLYNOMIAL REGRESSION

Consider the regression design problem where the regression function is a polynomial of degree (k−1) on [a,b]. then

$$Ey(x) = \theta' f(x)$$

where $\theta=(\theta_0, \ldots, \theta_{k-1})'$, $f(x)=(1,x, \ldots, x^{k-1})'$.

An exact design specifies a probability measure ξ on [a,b] which concentrates mass p_i at x_i, $i=1,\ldots r$, where $p_i n = m_i$, $i=1,\ldots,r$, are integers. The design problem now is to choose the design in some optimal way.

If the unknown parameter vector $\theta=(\theta_0,\theta_1,\ldots,\theta_{k-1})'$ 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 $\sigma^2 n^{-1} M^{-1}(\xi)$, where $M(\xi) = \int f(x)f'(x)d\xi(x)$, $M(\xi)$ is commonly called the "information matrix" of the design ξ. If the matrix $M^{-1}(\xi)$ is "small" or $M(\xi)$ is "large", then roughly speaking $\hat{\theta}$ is close to θ. A simple measure of the magnitude of the information matrix $M(\xi)$ is its determinant. Thus an exact D-optimal design is defined as follows:

DEFINITION 1.2.1. An exact design ξ^* is said to be D-optimal if ξ^* maximizes $|M(\xi)|$ among all the exact designs ξ on [a,b].

The emphasis above is on restricting the m_i to be integers. An approach which is often taken in optimal design work is to extend consideration to the class of all "approximate designs", i.e. arbitrary probability measures ξ on X. This approach has the advantage of permitting a complete characterization of certain optimal designs. Hoel(1958) has obtained the result that an approximate design is D-optimal for polynomial regression of degree k-1 on $[a,b] = [-1,1]$ if and only if it concentrates equal mass at the roots of $(1-x^2)P_{k-1}'(x)$, where $P_{k-1}(x)$ is the (k-1)th Legendre polynomial. Karlin and Studden(1966) have extended the result to some special choices of $f(x)=(f_0,f_1,\ldots,f_{k-1})'$, where $f_i(x)$ is x^i multiplied by some weight functions. Its limitation is that, in practice, only an exact design may be implemented. It is usually the case that an optimal approximation design is not exact for many choices of n. For more discussions and further references about the approximate design, see Fedorov(1972), Karlin and Studden(1966). Therefore it is our purpose to find the exact D-optimal design in this thesis for polynomial regression.

Salaevskii(1966) conjectures that an exact D-optimal design ξ^* distributes observations as evenly as possible among the k support points of the approximate D-optimal design. Constantine and Studden(1981) have a simplified proof of Salaevskii's result that the conjecture holds for sufficiently large n. Gaffke and Krafft(1982) have proved Salaevskii's conjecture for quadratic regression for all n using a new and quite simple proof.

In the second part of this thesis, following the new approach, we are able to prove Salaevskii's result for large sample case quite simply. Also for polynomials of degree $\leqq$ 9, Salaevskii"s conjecture is proved except for a few cases.

CHAPTER II

MODEL ROBUST DESIGN IN REGRESSION FOR X ON $[0,\infty)$

2.1. INTRODUCTION

Let X be a subset of an Euclidean space. An observation at $x \in X$ is assumed to be of the form

$$y(x) = f(x) + \epsilon_x,$$

where $E\epsilon_x = 0$, $E\epsilon_x^2 = \sigma^2$ and $f \in \Gamma$, a class of possible regressions over X.

This chapter fixes Γ to be a special class as in Huber(1975) , which is described below, and deals with designs, that is, placement of uncorrelated observations in X, for the estimation of regression parameters.

Huber(1975) discussed the design problem of extrapolating a function f to a point $x_0 = -1$ outside X, where $X = [0,\infty)$. f is assumed to be in the class Γ_0, where Γ_0 is defined to be the collection of functions f that are h+1 times differentiable, $h \geqq 0$, and that the (h+1)th derivative is bounded by ϵ

$$|f^{(h+1)}(x)| \leqq \epsilon, \ x_0 \leqq x < \infty. \tag{2.1.1}$$

Also he confined himself to the use of estimates f, which are linear functions of the observations, and tried to find the design which minimizes the maximum risk

$$\sup_{f\in\Gamma_0} E(\hat{f} - f(-1))^2.$$

Unfortunately, Huber's proof for obtaining the optimal model robust extrapolation design under the criterion described above is incorrect. In Section 2.2 we shall give a counter example and show where the mistake was made. The result however still seems to be correct.

In Section 2.3 with the restriction that there are exactly h+1 design points, i.e. the minimal number of design points under certain constraints which are necessary for keeping the maximum risk bounded, we see that the optimal model robust extrapolation design on $[0,\infty)$ is as described in the Theorem 6.1 of Huber's paper. The design points are found for different values of n, ϵ, σ^2 and the point x_0 we wish to extrapolate to. The corresponding weight of each design point is determined by certain constraints.

In Section 2.4 the model robust design with exactly h+1 points on $[0,\infty)$ for estimating $f^{(k)}(0)$, which is the coefficient of x^k if f is a polynomial, is discussed. We also find the limiting design measure as h goes to ∞ for

$f^{(k)}(0)$ and $f^{(h+1-k)}(0)$ for k fixed and the limiting design for $f^{([nq])}(0)$ for $0 < q < 1$.

In the following, we reformulate the extrapolation problem into a mathematical setting as in Huber(1975).

Suppose that n uncorrelated observations on the response y(x) are to be obtained at levels $x_1,\ldots,x_n$, $x_i \in X$. Then $E(y(x_i)) = f(x_i)$, and $var(y(x_i)) = \sigma^2$, $i = 1,\ldots,n$.

In order to more conveniently formulate the design problem, let $x_1,\ldots,x_r$ now denote the distinct levels at which $n_1,\ldots,n_r$ observations are taken. Also let y_i be the average of n_i observations taken at x_i, $i=1,\ldots,r$, $\sum n_i = n$. Put $m_i = n_i/n$, thus $var(y_i) = \sigma^2/n_i$ and

$$\sum_{i=1}^{r} m_i = 1.$$

Recall that we are interested in extrapolating $f(x_0)$, $x_0 < 0$, from $X = [0,\infty)$, and $f \in \Gamma_0$. Also we only consider linear estimates based on the observations y_i, $i=1,\ldots,r$, to predict the value of $f(x_0)$.

Let

$$\hat{f} = \sum_{i=1}^{r} a_i y_i.$$

be a linear estimate of $f(x_0)$.

The design problem here is to choose r, m_i, x_i, a_i, $i=1,\ldots,r$, to minimize the maximum mean square error

$$\sup_{f\in\Gamma_0} E(\hat{f} - f(x_0))^2 = \sup_{f\in\Gamma_0} (var(\hat{f}) + (E(\hat{f}) - f(x_0))^2)$$

$$= \sup_{f\in\Gamma_0}\left[\frac{\sigma^2}{n}\sum_1^r \frac{a_i^2}{m_i} + (\sum_1^r a_i f(x_i) - f(x_0))^2\right]$$

$$= \frac{\sigma^2}{n}\sum_1^r \frac{a_i^2}{m_i} + \sup_{f\in\Gamma_0} (\sum_1^r a_i f(x_i) - f(x_0))^2 \qquad (2.1.2)$$

The minimization of (2.1.2) is called the exact problem. If (2.1.2) is minimized w.r.t. $\{m_i\}$ for fixed $\{a_i\}$ and $\{x_i\}$ and without regard to the integer nature of the n_i's, the resulting minimum occurs when

$$m_i = \frac{|a_i|}{\sum|a_i|}, \quad i=1,\ldots,r ,$$

and has value

$$\sup_f E(\hat{f} - f(x_0))^2 = \frac{\sigma^2}{n}(\sum_1^r |a_i|)^2 + \sup_f(\sum_1^r a_i f(x_i) - f(x_0))^2$$

$$= \frac{\sigma^2}{n}(\sum_1^r |a_i|)^2 + \sup_f (\int f dA)^2, \qquad (2.1.3)$$

where A is a pure jump function, with jumps of size a_i at x_i, and a jump size -1 at x_0, such that $A(x) = 0$, for $x < x_0$.

Let R(A) denote the maximum mean square error w.r.t. A in (2.1.3). The minimization of R(A) is called the approximate problem.

Note that Γ_0 contains all polynomials of degree $\leq h$. Hence $\int f\, dA$ cannot stay bounded for all f in Γ_0 unless it vanishes for all polynomials f of degree $\leqq$ h. Therefore,

$$\int x^j \, dA = 0, \quad 0 \leqq j \leqq h. \tag{2.1.4}$$

or

$$\sum_{i=1}^{r} a_i x_i^{\,j} = x_0^{\,j}.$$

By Taylor's theorem and (2.1.4), $\int f \, dA$ can be written as

$$\int_{x_0}^{\infty} f \, dA = \int_{x_0}^{\infty} \int_{x_0}^{\infty} (h!)^{-1} \, (x - t)_+^{\,h} \, f^{(h+1)}(t) \, dt dA$$

$$= \int_{x_0}^{\infty} \int_{x_0}^{\infty} \frac{(x - t)_+^{\,h}}{h!} \, dA(x) \, f^{(h+1)}(t) \, dt$$

$$= \int_{x_0}^{\infty} B(t) f^{(h+1)}(t) \, dt,$$

where

$$B(t) = \int_{x_0}^{\infty} \frac{(x - t)_+^{\,h}}{h!} \, dA(t)$$

$$= \frac{1}{h!} \sum_{1}^{r} a_i (x_i - t)_+^{\,h}$$

Hence,

$$\sup_{f \in \Gamma_0} \left| \int f \, dA \right| = \epsilon \int |B(t)| \, dt.$$

Thus

$$R(A) = \sigma^2/n \cdot \left(\sum_{1}^{r} |a_i| \right)^2 + \epsilon^2 \left(\int |B(t)| \, dt \right)^2. \tag{2.1.5}$$

At this point, we find that for any design which does not contain the left end point of the interval, i.e. the point 0 when $X = [0,\infty)$, it is always better to shift the design points to the left so that it contains the left end point. We prove this in the following lemma.

LEMMA 2.1.1. For any signed measure A_1, whose support points are on x_i, $i=0,\ldots,r$, $x_0 < 0 < x_1 <\ldots< x_r$, and with corresponding jump size $-1, a_1,\ldots,a_r$. There is a signed measure A_2 whose support points x_0, $y_1 < y_2 <\ldots< y_r$ are obtained by a linear transformation of $\{x_i, i=0,\ldots,r\}$, such that x_0 remains fixed and $y_1 = 0$, and whose corresponding jump sizes are also $-1, a_1,\ldots,a_r$. Then $R(A_2) \leqq R(A_1)$.

Proof: Let A_1 be a signed measure with support points x_i, $i=0,\ldots,r$, $x_0 < 0 < x_1 <\ldots< x_r$ and corresponding jump size $-1, a_1,\ldots,a_r$. We define a linear transformation i such that

$$x_0 = bx_0 + c,$$

$$0 = bx_1 + c.$$

Thus

$$0 < b = (x_1 - x_0)/(-x_0) < 1.$$

Let

$$y_i = bx_i + c, \ i=1,\ldots,r,$$

then using constraints (2.1.4), we obtain

$$\sum_{i=1}^{r} a_i y_i^{\,j} = \sum a_i (bx_i + c)^j$$

$$= (bx_0 + c)^j = x_0{}^j, \ j=0,1,\dots,h.$$

Therefore, the signed measure A_2 with support x_0, y_i, $i=1,\dots,r$, where $y_1 = 0$, and with corresponding jump sizes -1, $a_1,\dots,a_r$ satisfies the constraints (2.1.4), and

$$R(A_2) = \sigma^2/n\cdot(\sum|a_i|)^2 + \epsilon^2(\int_{x_0}^{y_r} |B_2(s)| \, ds)^2$$

$$= \sigma^2/n\cdot(\sum|a_i|)^2 + (h!)^{-2}\epsilon^2(\int_{x_0}^{y_r} |\sum a_i(y_i - s)_+{}^h| \, ds)^2$$

$$= \sigma^2/n\cdot(\sum|a_i|)^2$$

$$+ (h!)^{-2}\epsilon^2\cdot b^{2(h+1)}(\int_{x_0}^{x_r} |\sum a_i(x_i - t)_+{}^h| \, dt)^2$$

$$\leqq \sigma^2/n\cdot(\sum|a_i|)^2 + \epsilon^2(\int_{x_0}^{x_r} |B_1(t)| \, dt)^2$$

$$= R(A_1).$$

With the result of Lemma 2.1.1. in mind, evidently (2.1.5) can also be minimized by minimizing

$$\int |dA|$$

subject to the condition

$$\int |B(t)| \, dt = c.$$

Therefore, the approximate design problem of minimizing the maximum mean square error can be turned into the mathematical problem called Q1.

PROBLEM Q1: Minimize

$$\int |dA| \tag{2.1.6}$$

subject to the condition

$$\int |B(t)|\ dt = c, \tag{2.1.7}$$

and the side conditions (2.1.4)

$$\int x^j\ dA = 0,\ j=0,\dots,h.$$

Huber tried to solve Problem Q1 by solving Problem Q2, which is minimizing (2.1.6) subject to conditions (2.1.4) and

$$\int B(t)\ dt = c. \tag{2.1.8}$$

2.2. ON HUBER'S PROOF

In this section we shall start with an example which shows that the approach of Huber's proof is incorrect. In his proof he tried to solve problem Q2 and said that since the design is on exactly h+1 points, the bias function B(t) does not change sign, therefore the minimal solution of problem Q2 is the same as that of the problem Q1.

Now if we take his example for $h = 1$, the optimal model robust linear extrapolation design ξ_1 for $x_0 = -1$ takes observation at $r = 2$ points, $x_1 = 0$, $x_2 = 1/\delta$, and allocates a fraction $m_1 = (1 + \delta)/(1 + 2\delta)$ to x_1, $m_2 = \delta/(1 + 2\delta)$ to

x_2, where δ is the unique solution of

$$\frac{n\epsilon^2}{\sigma^2} = \frac{8(1 + 2\delta)}{1 + \delta}$$

in the interval $[0,\infty)$.

Let $n\epsilon^2/\sigma^2 = 12$, then $\delta = 1$ is the only solution in $[0,\infty)$. The resulted total risk for problem Q2 is

$$\epsilon^2((1 + 2)^2/12 + 1) = \epsilon^2(7/4).$$

If we consider another design ξ_2, where

$$x_1 = 0,\ x_2 = 5,\ x_3 = 10,$$

$$a_1 = 33/25,\ a_2 = -11/15,\ a_3 = 3/25,$$

with weight

$$m_i = |a_i| \,/\, \sum|a_i|,\ i=1,2,3;$$

then

$$\int B(t)\,dt = \sum a_i x_i^2 = 1,$$

and satisfies the condition

$$\sum_{i=1}^{3} a_i = 1$$

$$\sum_{i=1}^{3} a_i x_i = x_0 = -1.$$

We find that the total risk of design ξ_2 for problem Q2 is

$$\epsilon^2((1.88)^2/12 + 1),$$

which is much smaller than that of the design ξ_1.

This example shows that the optimal design of problem Q1 is not an optimal design in problem Q2, which contradicts with Huber's proof.

The reason for the mistake is that the solution of problem Q2 in $[0,\infty)$ does not exist. This can be shown after finding the solution of problem Q2 in a closed interval $[0,T]$.

In order to show the solution of problem Q2 in $[0,\infty)$ does not exist, we need to go back to the original set-up without substituting $m_i = |a_i|/\sum|a_i|$ into the variance part. the problem can be read as follows, called Q3:

Problem Q3: Minimize $\sum \frac{a_i^2}{m_i}$

where $m_i > 0$, $i=1,\ldots,r$, and $\sum m_i = 1$.

subject to

$$(h+1)! \int B(t)\,dt = c,$$

i.e.

$$\sum a_i x_i^{h+1} = x_0^{h+1} + c;$$

and

$$\sum a_i x_i^j = x_0^j, \quad j=0,1,\ldots,h;$$

where $\{x_i,\ i=1,\ldots,r\}$ are in $[0,T]$.

THEOREM 2.2.1. If T is large enough, the minimal solution of Problem Q3 is unique, where $\{x_i\}$ are the Tchebyscheff points of order h+1 in the interval [0,T]. $\{a_i\}$ are determined according to constraints (2.1.4), and the weight $m_i = |a_i|/\sum|a_i|$, for all i. i.e.

$$\underset{\sim}{a}' = V^{-1}(\underset{\sim}{x_0} + \underset{\sim}{1_c}),$$

where

$$\underset{\sim}{a}' = (a_1, \ldots, a_{h+2}),$$

$$\underset{\sim}{x_0}' = (1, x_0, \ldots, x_0^{h+1}),$$

$$\underset{\sim}{1_c}' = (0, \ldots, c)_{1\times(h+2)},$$

$$V = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ x_1 & \cdot & \cdot & \cdot & x_{h+2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1^{h+1} & \cdot & \cdot & \cdot & x_{h+2}^{h+1} \end{pmatrix}.$$

The proof of this theorem will be delayed till the end of this chapter.

From Theorem 2.2.1 , we know that for any constant c, if T is large enough,

$$\sum_{i=1}^{h+2} \frac{a_i^2}{m_i} = \left(\sum_{i=1}^{h+2} |a_i|\right)^2$$

$$= \left[\sum_{i=1}^{h+2} \left| \prod_{j\neq i} (x_0 - x_j) + c_0) / \prod_{j\neq i} (x_j - x_i) \right| \right]^2$$

where $x_j = T/2[1 + \cos((h+2-j)\pi/(h+1)]$, $1 \leqq j \leqq h+2$

Therefore,

$$\lim_{T\to\infty} \left(\sum_{i=1}^{h+2} |a_i| \right)$$

$$= \sum_{i=1}^{h+2} \left| \frac{\prod_{j\neq i}(1 - \cos((j-1)\pi/(h+1))}{\prod_{j\neq i} [\cos((i-1)\pi/(h+1)) - \cos((j-1)\pi/(h+1)]} \right|$$

$$= 1.$$

From the constraints of the problem, we know that

$$\sum_{i=1}^{r} |a_i| > \sum_{i=1}^{r} a_i = 1,$$

which means that the minimum value of $\sum|a_i|$ is not attainable for $\{x_i, \; i=1,...r\}$ in $[0,\infty)$ subject to constraints (2.1.4) for any $h \geqq 1$.

2.3. MODEL ROBUST EXTRAPOLATION DESIGN WITH MINIMAL NUMBER OF POINTS ON $[0,\infty)$

In some situations we are interested in the optimal design with as few observations as possible, which in our case is exactly h+1 points. In this case we see that the bias function B(t) is the B-spline function of order h+1 for the knot sequence $\{x_i, i=0,\ldots,h+1\}$ multiplied by a constant. So the bias function will not change sign. For more details about B-spline functions see Carl De Boor(1978). Therefore, Problem Q1 is equivalent to Problem Q2.

The corresponding a_i's are determined now by the constraints (2.1.4). In other words, $a_i = L_i(x_0)$ for $i=1,\ldots,h+1$, where $L_i(x)$ are the Lagrange interpolation polynomials on the h+1 points $x_1,\ldots,x_{h+1}$. Also the integral of the bias function can be expressed as

$$\left|\int_{x_0}^{x_{h+1}} B(t)\,dt\right| = \left|\frac{1}{(h+1)!}\left(x_0^{h+1} - \sum_{i=1}^{h+1} a_i x_i^{h+1}\right)\right|$$

$$= |((h+1)!)^{-1}(x_0 - x_1)\ldots(x_0 - x_{h+1})|,$$

where the second equality can be obtained either by the property that B(t) is a B-spline function, or using the constraints (2.1.4) to compute directly. We shall prove it using properties of the B-spline function.

Now we shall find where the optimal design points are in terms of different values of n, σ^2, ϵ and x_0. Huber's proof is used for part of the solution.

<u>THEOREM 2.3.1.</u> For any value of $x_0 < 0$, the optimal robust extrapolation design with exactly h+1 points on $[0,\infty)$ are on $0 = x_1 < x_2 < \ldots < x_{h+1}$, which after addition of another

point y constitute the set of Tchebyscheff points of order h+1 in the interval [0,y].

Proof: Huber took the constraints (2.1.4) and (2.1.7) into account. Then with the aid of Lagrange multipliers and a variational argument he obtained the necessary conditon that the optimal design points need to satisfy, namely that for some polynomial P_{h+1} of degree h+1,

$$P_{h+1}(x) = \pm 1,$$

for all the support points in the optimal design, and

$$P_{h+1}'(x_i) = 0 \text{ for all } x_i > 0.$$

This implies the stated result. For more details see Huber(1975).

In the following, the notation of z_k will be used extensively and thus is defined here.

Let

$$z_k = 1 + \cos((h+2-k)\pi/(h+1)),\ k=1,\ldots,h+1. \quad (2.3.1)$$

In order to find exactly where the optimal design points should be, we need to determine the end point of the interval described in Theorem 2.3.1. Note that $(h+1)! \cdot \int B(t)\,dt$ can be written as $x_0 \cdot (2^{-h} U_h(x_0,y))$ where $U_h(x,y)$ is defined to be the Tchebyscheff polynomial of the second kind on [0,y]. Also the notation $L_i(x_0,y)$ is used to indicate that the h+1 points $x_1,\ldots,x_{h+1}$, over which the Lagrange interpolation polynomial is defined, are determined by y. Thus we have the following theorem.

THEOREM 2.3.2. For any value of $x_0 < 0$, and a given ρ, where $\rho = \sigma^2((h+1)!)^2/(n\epsilon^2)$, the corresponding value of y_0 for the optimal robust extrapolation design on $[0,\infty)$ as described in Theorem 2.3.1. is the unique positive solution of

$$\rho = \frac{(-1)^{h+1}x_0(2^{-h}U_h(x_0,y))}{2^{2h+1}(\sum |L_i(x_0,y)|)}\cdot y^{h+1} \tag{2.3.2}$$

The corresponding weight of x_i, $i=1,\ldots,h+1$, is

$$|L_i(x_0,y_0)| \;/\; \sum|L_i(x_0,y_0)|.$$

Before proving the theorem, we give an example to see how to apply this theorem.

Example 2.3.1. Let the regression function be in Γ_0, where $h = 2$. We are interested in extrapolating $f(-1)$ from $[0,\infty)$. Then from the above theorem, given $n = 10$, $\sigma^2 = 1$, $\varepsilon = 0.05$, the optimal robust extrapolation design with exactly 3 design points is on 0, x_2, x_3, where

$$x_2 = (y_0/2)(1 + \cos 2\pi/3) = y_0/4$$

$$x_3 = (y_0/2)(1 + \cos\pi/3) = 3y_0/4,$$

where y_0 is the unique positive root of the following equation

$$\rho = \frac{(1 + y/4)(1 + 3y/4)}{2^5(1 + 12/y + 16/y^2)}\cdot y^3 .$$

where $\rho = \sigma^2((h+1)!)^2/(n\epsilon^2) = 1440$. The corresponding weight is $m_i = |a_i| / \sum |a_i|$, where $a_i = L_i(-1)$.

With the aid of a computer, we get $y_0 = 12.76$. Therefore,

the robust design is on

$$x_1 = 0,\ x_2 = 3.19,\ x_3 = 9.37,$$

with corresponding weight

$$m_1 = 0.7149,\ m_2 = 0.2506,\ m_3 = 0.0345.$$

Now we prove Theorem 2.3.2.

Proof: From Theorem 2.3.1. we know the optimal design is on $0 = x_1 < x_2 < \ldots < x_{h+1}$, where

$$x_k = (y/2)z_k,\ k=2,\ldots,h+1,$$

for some $y \in (0,\infty)$. Denote the corresponding signed measure by $A(y)$. The risk value is

$$R(A(y)) = (\sigma^2/n)\{\sum|L_i(x_0,y)|)^2 + (\epsilon x_0/((h+1)!))^2(2^{-h}U_h(x_0,y))^2\}$$

$$= \alpha\{\ \rho(\sum|L_i(x_0,y)|)^2 + x_0{}^2(2^{-h}U_h(x_0,y))^2\},$$

where $\alpha = \epsilon^2/((h+1)!)^2$, $\rho = \sigma^2((h+1)!)^2/(n\epsilon^2)$.

To find the y value which can minimize the risk, we differentiate $R(A(y))$ w.r.t. y and set it to 0 as follows:

$$0 = d/dyR(A(y))$$

$$= \alpha\{2\rho\ (\sum|L_i(x_0,y)|)\cdot d/dy(\sum L_i(x_0,y)|) +$$

$$2x_0{}^2(2^{-h}U_h(x_0,y))\cdot d/dy(2^{-h}U_h(x_0,y))\}.$$

How to simplify the derivative of $(\sum|L_i(x_0,y)|)$ and $(2^{-h}U_h(x_0,y))$ w.r.t. y is a major problem here. Therefore we use the following four lemmas to get the answer. The proofs of the lemmas will again be delayed till the end of this chapter.

<u>LEMMA 2.3.1.</u> $h!B(t) = \sum_{i=1}^{h+1} a_i(x_i - t)_+{}^h$ is a B-spline function under the constraints (2.1.4). The integral of the bias is then

given by

$$(h+1)!\int_{x_0}^{x_{h+1}} B(t)\ dt = x_0{}^{h+1} - \sum_{i=1}^{h+1} a_i x_i{}^{h+1}$$

$$= (x_0 - x_1)\dots(x_0 - x_{h+1}).$$

<u>LEMMA 2.3.2.</u> $\sum_{2\leqq i_1<\dots<i_r\leqq h+1} z_{i_1}\dots z_{i_r} = \frac{1}{2}\binom{2h-r+1}{r}$

where z_k, $k=1,\dots,h+1$ are as defined in (2.3.1).

<u>Lemma 2.3.3.</u> $d/dy(\sum|L_i(x_0,y)|)$

$$= (-1)^{h+1}\frac{x_0}{y^{h+1}}\sum_{k=1}^{h}\left\{(-1)^{h-k}2^{2k-1}(h-k+1)\binom{h+k}{h-k+1}x_0{}^{k-1}y^{h-k}\right\}.$$

<u>Lemma 2.3.4.</u> Let $2^{-h}U_h(x_0,y) = (x_0 - x_2)\ldots(x_0 - x_{h+1})$,

where $x_k = (y/2)z_k$, $k=2,\ldots,h+1$,

then

$$d/dy(2^{-h}U_h(x_0,y))$$

$$= \frac{(-1)^h}{2^{2h+1}} \cdot \frac{y^{h+1}}{x_0} \cdot \frac{d}{dy}(\sum |L_i(x_0,y)|).$$

Therefore, we get

$$0 = \rho(\sum |L_i(x_0,y)|) + \frac{(-1)^h \cdot x_0}{2^{2h+1}} \cdot y^{h+1}(\frac{1}{2^h}U_h(x_0,y)) \quad (2.3.3)$$

which can be written as (2.3.2).

Equation (2.3.3) multiplied by y^h is a polynomial of y of degree 3h+1 and has an unique solution in $(0,\infty)$. The first term on the right side of the equation multiplied by y^h is a polynomial of y of degree h and all its coefficients are of positive sign. The second term on the right side is a polynomial of y of degree 3h+1, whose first 2h coefficients are 0 and the rest are of positive sign. Therefore there is only one change of sign. By Descartes rule of signs(see Weisner(1938)), there is an unique positive root of (2.3.3), which completes the proof.

2.4. MODEL ROBUST DESIGN WITH MINIMAL NUMBER OF POINTS FOR ESTIMATING $f^{(k)}(0)/k!$ ON $[0,\infty)$

In this section we are interested in the design problem of estimating $f^{(k)}(0)$, a linear functional of f, for f in the class Γ_0, where Γ_0 is as defined in (2.1.1).

Again, suppose that we have n uncorrelated observations at $x_1,\ldots,x_n$, where r of them are distinct and each has n_i observations at it. The response $y(x_i)$ has mean $f(x_i)$ and variance σ^2. Let y_i be the average of the n_i observations at $x_i \in [0,\infty)$, and we only consider linear estimates $\hat{f} = \sum a_i y_i$ to estimate $f^{(k)}(0)$. Therefore, the design problem here is to choose r, m_i, x_i, a_i, i=1,...,r, to minimize the maximum mean square error

$$\sup_{f\in\Gamma_0} E(\hat{f} - f^{(k)}(0)/k!)^2. \tag{2.4.1}$$

Let A be a pure jump function, with jumps of size a_i at x_i, such that $A(x) = 0$ for $x < x_1$. Then (2.4.1) can be written as

$$\sup_{f\in\Gamma_0} E(\hat{f} - f^{(k)}(0)/k!)^2$$

$$= \sup_{f\in\Gamma_0} (\mathrm{var}(\hat{f}) + (E(\hat{f}) - f^{(k)}(0)/k!)^2)$$

$$= \frac{\sigma^2}{n}\sum_{1}^{r}\frac{a_i^2}{m_i} + \sup_{f\in\Gamma_0}\left(\sum_{i=1}^{r} a_i f(x_i) - f^{(k)}(0)/k!\right)^2$$

$$= \frac{\sigma^2}{n}\left(\sum_1^r |a_i|\right)^2 + \sup_{f\in\Gamma_0}\left(\int f\ dA - f^{(k)}(0)/k!\right)^2. \quad (2.4.2)$$

with $m_i = |a_i| / \sum |a_i|$.

Similarly, since Γ_0 contains all polynomials of degree $\leq$ h, we need the following constraints to have R(A) remain bounded. Let δ_{jk} denote the Kronecker delta where $\delta_{jk} = 0$ or 1 according as $j \neq k$ or $j=k$.

$$\int_0^\infty x^j\ dA = \delta_{jk},\ 0 \leq j \leq h,$$

or

$$\sum_1^r a_i x_i{}^j = \delta_{jk},\ 0 \leq j \leq h. \quad (2.4.3)$$

By Taylor's theorem and (2.4.3), $\int f\ dA$ can be written as

$$\int_0^\infty f\ dA = \int_0^\infty \int_0^\infty (h!)^{-1}(x - t)_+{}^h\ f^{(h+1)}(t)\ dtdA$$

$$= \int_0^\infty \int_0^\infty \frac{(x - t)_+{}^h}{h!}\ dA(x)\ f^{(h+1)}(t)\ dt$$

$$= \int_0^\infty B(t) f^{(h+1)}(t)\ dt,$$

where

$$B(t) = \int_0^\infty \frac{(x - t)_+{}^h}{h!}\ dA(x)$$

$$= \frac{1}{h!} \sum_{1}^{r} a_i (x_i - t)_+^h$$

Hence,

$$\sup_{f \in \Gamma_0} |\int f \, dA| = \epsilon \int |B(t)| \, dt.$$

Thus

$$R(A) = \sigma^2/n \cdot (\sum |a_i|)^2 + \epsilon^2 (\int |B(t)| \, dt)^2. \quad (2.4.4)$$

and this can be minimized by minimizing

$$\int |dA|$$

under the side condition

$$\int |B(t)| \, dt = c.$$

Similarly, if we confine ourselves to the design with the minimal number of points, namely h+1 points, it is easy to see that the bias function B(t) can have at most h−1 changes of sign, thus (2.4.2) and a repeated application of the mean value theorem shows that B(t) can not change signs, therefore

$$\int |B(t)| \, dt = | \int B(t) \, dt |$$

$$= \left| \int \frac{x^{h+1}}{(h+1)!} \, dA \right| .$$

The variance term for R(A) is determined by the constraints (2.4.3).

Let

$$\underset{\sim}{a}' = (a_1, \ldots, a_{h+1})$$

$$\underset{\sim}{X} = \begin{pmatrix} 1 & . & . & . & 1 \\ x_1 & . & . & . & x_{h+1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_1^h & . & . & . & x_{h+1}^h \end{pmatrix} .$$

$$\underset{\sim}{1}_k{}' = (0,\ldots,0,1,0,\ldots,0)_{1\times h+1}$$

where 1 is at the (k+1)th coordinate.

Then

$$\underset{\sim}{a} = \underset{\sim}{X}^{-1}\underset{\sim}{1}_k, \tag{2.4.5}$$

and a_i equals to α_{mi}, the coefficient of x^m in the ith Lagrange polynomial $L_i(x)$ which is defined as the polynomial of degree h which vanishes at the h points x_j, $j \neq i$, that is $L_i(x_j) = \delta_{ij}$.

An easy computation gives

$$\sum_{i=1}^{h+1} |a_i| = \sum_{i=1}^{h+1} \frac{\sum_{\neq i} x_{i_1} \cdots x_{i_{h-k}}}{\prod_{j\neq i} (x_i - x_j)} .$$

We also know that the bias term is $((h+1)!)^{-1}(\sum a_i x_i{}^{h+1})$. Then by substituting a_i in, we have

$$\sum_{i=1}^{h+1} a_i x_i{}^{h+1}$$

$$= (x_1^{h+1}, \ldots, x_{h+1}^{h+1}) \underset{\sim}{X}^{-1} \cdot \underset{\sim}{1}_k$$

$$= \frac{\det \begin{bmatrix} x^{h+1}, & 1, & x, \ldots, x^{k-1}, x^k & , \ldots, x^h \\ x_1 & , x_2, & \cdot \quad \cdot \quad \cdot & , x_{h+1} \end{bmatrix}}{\det \begin{bmatrix} 1, & x, & \cdot \quad \cdot \quad \cdot & , x^h \\ x_1, & x_2, & \cdot \quad \cdot \quad \cdot & , x_{h+1} \end{bmatrix}} ,$$

where $\det \begin{bmatrix} u_1, \ldots, u_m \\ x_1, \ldots, x_m \end{bmatrix}$ denotes the determinant

$$\begin{vmatrix} u_1(x_1) & u_2(x_2) & \cdot & \cdot & \cdot & u_1(x_m) \\ u_2(x_1) & u_2(x_2) & \cdot & \cdot & \cdot & u_1(x_m) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ u_m(x_1) & u_m(x_2) & \cdot & \cdot & \cdot & u_m(x_m) \end{vmatrix} .$$

Hence

$$\sum_{i=1}^{h+1} a_i x_i^{h+1} = \sum_{1 \leqq i_1 < \ldots < i_{(h+1-k)} \leqq h+1} x_{i_1} \ldots x_{i_{(h+1-k)}}$$

<u>THEOREM 2.4.1.</u> In estimating $f^{(k)}(0)/k!$ for some k, $1 \leqq k \leqq h$, the optimal robust design with exactly h+1 points on $[0,\infty)$ is on $0 = x_1 < x_2 < \ldots < x_{h+1}$, which after addition of another point y_k constitute the Tchebyscheff points of order h+1 in the interval $[0,\tilde{y}_k]$, where

$$\tilde{y}_k = 4(\rho(h+1-k)/(4k))^{1/(2h+2)}$$

for given ρ, where $\rho = \sigma^2((h+1)!)^2/(n\epsilon^2)$. The corresponding weight $m_i = |a_i|/\sum|a_i|$, where a_i are determined by the constraints (2.4.2)

Again, before giving the proof, we see an example of the application of this theorem.

Example 2.4.1. Let the regression function f is in Γ_0, where $h = 2$. We are interested in estimating $f^{(2)}(0)/2!$. Then the optimal robust design with exactly 3 points on $[0,\infty)$ is on 0, x_2, x_3, where

$$x_2 = (\tilde{y}_2/2)(1 + \cos 2\pi/3) = y_2/4,$$

$$x_3 = (\tilde{y}_2/2)(1 + \cos\pi/3) = 3y_2/4,$$

where

$$\tilde{y}_2 = 4(\rho/8)^{1/6}$$

for given ρ, where $\rho = \sigma^2(3!)^2/(n\epsilon^2)$. The corresponding weights m_i are $|a_i|/\sum|a_i|$, where

$$a_1 = (x_2 x_3)^{-1} = (16/3)\cdot\tilde{y}_2^{-2},$$

$$a_2 = [x_2(x_2 - x_3)]^{-1} = 8\cdot\tilde{y}_2^{-2},$$

$$a_3 = [x_3(x_3 - x_2)]^{-1} = (8/3)\cdot\tilde{y}_2^{-2}.$$

Therefore, if given $n = 10$, $\sigma^2 = 1$, $\epsilon = 0.05$, then $\rho = 1440$ and $\tilde{y}_2 = 9.5047$. We obtain that the optimal robust design for estimating $f^{(2)}(0)/2!$ has its observations on

$$x_1 = 0, \ x_2 = 2.3762, \ x_3 = 7.1285$$

with corresponding weight

$$m_1 = 1/3, \ m_2 = 1/2, \ m_3 = 1/6.$$

Now we prove the theorem.

Proof: The first part of the proof goes through as in Theorem 2.3.1 using the variational argument with the aid of Lagrange multipliers. The second part is the determination of the value of y_k.

Let $A(y)$ denote the signed measure with support points on $0 = x_1 < x_2 < \ldots < x_{h+1}$, where $x_i = (y/2)z_i$, $i=1,\ldots,h+1$. Then, by Lemma 2.3.2.

$$\sum_{i=1}^{h+1} |a_i| = \sum_{i=1}^{h+1} \frac{(y/2)^{h-k} \sum_{\neq i} z_{i_1} \cdots z_{i_{h-k}}}{(y/2)^h \prod_{j \neq i} (z_i - z_j)}$$

$$= \frac{2^{2k-1}}{y^k} \cdot \frac{h-k+1}{k} \binom{h+k}{h+1-k} ,$$

and

$$\sum_{i=1}^{h+1} a_i x_i^{h+1} = \left(\frac{y}{2}\right)^{h+1-k} \frac{1}{2^{h+1-k}} \binom{h+k}{h+1-k} ,$$

Therefore,

$$R(A(y)) = \frac{\epsilon^2}{((h+1)!)^2} \left\{ \rho \left[\frac{2^{2k-1}}{y^k} \cdot \frac{h-k+1}{k} \binom{h+k}{h+1-k} \right]^2 \right.$$

$$+ \left[\left(\frac{y}{2}\right)^{h+1-k} \cdot \frac{1}{2^{h+1-k}}\binom{h+k}{h+1-k}\right]^2 \Bigg\} ,$$

where $\rho = \sigma^2((h+1)!)^2/(n\epsilon^2)$.

Differentiate R(A(y)) w.r.t. y and equates to 0, we get the minimizing $\tilde{y}_k$,

$$\tilde{y}_k = 4(\ \rho(h+1-k)/(4k))^{1/(2h+2)} .$$

The corresponding weight of x_i is $m_i = |a_i|/\sum|a_i|$, for $i=1,\ldots,h+1$, where a_i are as in (2.4.3). Therefore the proof is completed.

For the use of the next two theorems, the weights given in Theorem 2.4.1. will be written out in a more explicit manner by a method similar to Studden(1978). Since the weight is invariant under scale transformations of x_i, we shall work with x_i in [0,2].

Let

$$R_{h+1}(x) = \sum_{i=1}^{h+1} \alpha_i x^i = \prod_{j=1}^{h+1} (x - x_j), \qquad (2.4.5)$$

where $x_j = z_j$, $j=1,\ldots,h+1$, z_j are defined as in (2.3.2). Therefore, by Lemma 2.3.2.

$$\alpha_i = \frac{(-1)^{h-i+1}}{2^{h-i+1}}\binom{h+i}{h-i+1} . \qquad (2.4.6)$$

THEOREM 2.4.2. The coefficients β_{mi} of x^m in the ith Lagrange polynomial are given by

$$\beta_{mi} = -\frac{1}{x_i^{m+1}}.\frac{R_{h+1,m}(x_i)}{R_{h+1}'(x_i)}, i\neq 1, \qquad (2.4.7)$$

where

$$R_{h+1,m}(x) = \sum_{i=1}^{m} \alpha_i x^i,$$

and

$$\beta_{m1} = \frac{\alpha_{m+1}}{R_{h+1}'(0)} \quad .$$

Proof: $R_{h+1}(x)$ vanishes at z_i, $i=1,\ldots,h+1$.

Hence the Lagrange polynomial is

$$L_i(x) = \frac{R_{h+1}(x)}{R_{h+1}'(z_\nu)(x - z_\nu)} \quad .$$

Then by equating coefficients and solving for β_{mi},

$$\beta_{mi} = \frac{\alpha_{k+1} + \alpha_{k+2}z_i + \ldots + \alpha_{h+1}z_i^{h-m}}{R_{h+1}'(z_i)} \quad ,$$

and

$$\frac{1}{z_i^{m+1}} \sum_{j=m+1}^{h+1} \alpha_j z_i^j = -\frac{R_{h+1,m}(z_i)}{z_i^{m+1}}$$

for $R_{h+1}(x)$ vanishes at z_i. Then the result follows.

Let $\xi(h+1,m)$ denote the optimal design for estimating $f^{(m)}(0)/m!$ with exactly h+1 points when $f \in \Gamma_0$. As $h \to \infty$, i.e. f gets smoother and smoother, we study the limiting behavior of the optimal design for estimating $f^{(m)}(0)/m!$ and $f^{(h+1-m)}(0)/(h+1-m)!$ for m fixed and the limiting design for $f^{[(h+1)q]}(0)/[(h+1)q]!$ for $0 < q < 1$ in the following theorem.

<u>THEOREM 2.4.3.</u> (i) The design $\xi(h+1,h+1-m)$ converges as $h \to \infty$ to the design ξ_1 with density proportional to

$$\frac{2 - x/2}{\sqrt{4x - x^2}}, \quad 0 < x < 4.$$

(ii) For fixed $q \in (0,1)$ the design $\xi(h+1,[(h+1)q])$ converges to ξ_q with density proportional to

$$\frac{2 - x/2}{(1 + (c/2)x)\sqrt{4x - x^2})}, \quad 0 < x < 4,$$

where $c = (1 - q^2)\cdot(2q^2)^{-1}$.

<u>Proof:</u> From (2.4.7) the weights on x_i have two parts, the parts $R_{h+1}'(x_i)$ in the denominator are proportional to $(2 - x_i)^{-1}$ for $i=1,\ldots,h+1$; the other parts $R_{h+1,m}(x_i)/x_i^{m+1}$ are essentially uniform on $\{x_i\}$ for case(i), where $k = h+1-m$, and are essentially proportional to the power series $(1 + cx_i)^{-1}$ for case (ii), where $k = [(h+1)q]$ and $c = -(2q^2)^{-1}(1 - q^2)$.

Note that

$$\frac{R_{h+1,k}(x_i)}{x_i^{k+1}} = \alpha_{k+1} + \alpha_{k+2}x_i + \ldots + \alpha_{h+1}x_i^{h-k}. \qquad (2.4.8)$$

Using the actual values of coefficients α_i from (2.4.7) we find

$$\left|\frac{\alpha_{k+2}}{\alpha_{k+1}}\right| = 2\cdot\frac{(1 + k/(h+1) + 1/(h+1))(1 - k/(h+1) - 1/(h+1))}{(2k/(h+1) + 3/(h+1))(2k/(h+1) + 1/(h+1))}$$

Then (2.4.8) may be seen to be essentially uniform for case (i) when $k = h+1-m$, and to be essentially proportional to the power series $(1 + cx_i)^{-1}$ for case (ii) when $k = [(h+1)q]$.

Moreover the weights are on $x_i = (\tilde{y}_k/2)z_i$. also we consider the transformation $\underset{\sim}{X} = (\tilde{y}_k/2)(1 - \cos\pi U)$ where U is uniform on (0,1). Since $\tilde{y}_k$ converges to the constant 4 for all k, we obtained the limiting densities for case(i) and (ii) as described in the theorem.

If $X_{h+1,m}$ denotes a random variable with corresponding probability measure $\xi(h+1,m)$, then the law of $Y_{h+1,m} = (h + 1)^2\cdot X_{h+1,m}$ converges weakly to a limiting measure or design which is discrete.

<u>THEOREM 2.4.4.</u> The variable $Y_{h+1,m} = (h + 1)^2\cdot X_{h+1,m}$ converges to a discrete variable Y. Y has weight w_i proportional to $\pi^m/(2m + 1)!$ on 0 and

$$\left|\sum_{j=0}^{m-1} \frac{(-1)^j}{(2j + 1)!} (i^2\pi^2)^j\right| \Big/ i^{2m}$$

on the points $i^2\pi^2$, $i=1,2,\ldots$.

Proof: The variable $Y_{h+1,m}$ has mass on

$x_k = (h+1)^2(\tilde{y}_m/2)z_k$, $k=1,\ldots,h+1$.

As $h \to \infty$, $x_k \to ((k-1)\pi)^2$ with $k=1,2,\ldots$. The terms $R_{h+1}'(x_i)$ may again be proportional to $(2-x_i)^{-1}$. If

$$w_{m1} = \alpha_{m+1}/R_{h+1}'(0),$$

$$w_{mk} = (R_{h+1}'(x_k)\cdot x_k^{m+1})^{-1}R_{h+1,m}(x_k),\ k \neq 1;$$

then

$$\sum_{k=1}^{h+1} w_{mk} \simeq ch^{2m+1}.$$

Using the exact expression for α_k in (2.4.7). it can be shown that $w_{mk}h^{-(2m+1)}$, $k=1,\ldots,h+1$, are proportional in the limit to the w_i, $i=0,1,\ldots$ given in the theorem.

2.5. APPENDIX

In order to prove Theorem 2.2.1 rewrite Problem Q3 in matrix form, then it reads as follows:

$$\text{minimize } \underset{\sim}{a}'W\underset{\sim}{a}$$

subject to

$$\underset{\sim}{X}\underset{\sim}{a} = \underset{\sim}{x}_0 + \underset{\sim}{1}_c\ ,$$

where $\underset{\sim}{X}$ denotes the matrix that its ith row is

$(x_1^{i-1},\ldots,x_r^{i-1})$, $i=1,\ldots,h+2$, and W denotes the $r\times r$ diagonal matrix with diagonal elements m_i^{-1}, $i=1,\ldots,r$.

$$\underset{\sim}{a}' = (a_1,a_2, \ldots ,a_r)_{1\times r},$$

$$\underset{\sim}{x}_0' = (1,x_0,x_0^2, \ldots , x_0^{h+1})_{1\times(h+2)},$$

$$\underset{\sim}{1}_c' = (0,0, \ldots ,0,c)_{1\times(h+2)} .$$

Let ξ be the design measure with support at points $x_1,\ldots,x_r$ and weight $m_1,\ldots,m_r$. Then

$$\underset{\sim}{X}W^{-1}\underset{\sim}{X}' = rM(\xi),$$

where

$$M(\xi) = \int f(x)f'(x)\, d\xi(x) ;$$

$$f'(x) = (1,x,\ldots,x^{h+1}).$$

Follow a similar proof as in Theorem 2.1 of Karlin and Studden(1966a),

$$\underset{\sim}{a}'W\underset{\sim}{a} = \sup_{d\in U^L,\, d\neq 0} [(\underset{\sim}{x}_0 + \underset{\sim}{1}_c,d)^2/(d'M(\xi)d)]$$

where $U = \{d|M(\xi)d = 0\}$.

Then Theorem 2.2. of Studden(1968) is used in our case where $f = (1,x,\ldots,x^{h+1})$, and $\underset{\sim}{C} = \underset{\sim}{x}_0 + \underset{\sim}{1}_c'$, $\{s_i, i=0,\ldots,h+1\}$ are the Tchebescheff points of order h+1 in the closed interval [0,T]. As long as T is large enough such that

$$\epsilon[(x_0 - s_0)\ldots(x_0 - s_h) + c] \geqq 0,$$

then

$$\epsilon D_i(c) = \epsilon \begin{vmatrix} 1 & 1 & \ldots & 1 & 1 & \ldots & 1 & 1 \\ s_0 & s_1 & & s_{i-1} & s_{i+1} & & s_{h+1} & x_0 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot \\ s_0^{h+1} & s_1^{h+1} & \ldots & s_{i-1}^{h+1} & s_{i+1}^{h+1} & \ldots & s_{h+1}^{h+1} & x_0^{h+1} + c \end{vmatrix} .$$

$$= \epsilon \prod_{j<k,\neq i} (s_k - s_j)\left(\prod_{j=0,\neq i}^{h+1} (x_0 - s_j) + c\right)$$

$$\geqq 0 \text{ for } i=0,1,\ldots,h+1.$$

Therefore by Theorem 2.2 of Studden(1968), the optimal design ξ_0 is supported on $s_0 <\ldots< s_{h+1}$ with mass

$$m_i = |D_i(C)| / \sum |D_i(C)| \ .$$

In other words, for $i=1,\ldots,h+2$,

$$x_i = s_{i-1}, \ m_i = |a_i| / \sum |a_i| ,$$

where $\underset{\sim}{a} = V^{-1}(\underset{\sim}{x}_0 + \underset{\sim}{1}_c)$.

<u>LEMMA 2.3.1.</u> $h!B(t) = \sum_{i=1}^{h+1} a_i (x_i - t)_+^h$

$$= (-1)^{h+1} \left[-(t - x_0)_+^h + \sum_{i=1}^{h+1} a_i (t - x_i)_+^h\right]$$

is a B-spline function under the constraints (2.1.4), and

$$(h+1)! \int B(t)\,dt = (x_0 - x_1)\dots(x_0 - x_{h+1}) \ .$$

Proof: From constraints (2.1.4) we know

$$a_i = L_i(x_0), \quad i=1,\dots,h+1;$$

where $L_i(x)$ are the Lagrange polynomials on $x_1,\dots,x_{h+1}$. So for $i=1,\dots,h+1$,

$$L_i(x_0) = \frac{g(x_0)}{(x_0 - x_i)g'(x_i)} \ ,$$

if $g(x) = \prod_{i=1}^{h+1} (x - x_i)$.

Then $h!B(t)$ can be written as

$$(-1)^{h+1}h!\cdot B(t) = [-(t - x_0)_+{}^h + \sum_{i=1}^{h+1} a_i(t - x_i)_+{}^h]$$

$$= g(x_0)\cdot\left\{-\frac{1}{g(x_0)}(t - x_0)_+{}^h + \sum_{i=1}^{h+1}\frac{1}{(x_0 - x_i)g'(x_i)}(t - x_i)_+{}^h\right\}$$

$$= g(x_0)\left\{-\sum_{i=0}^{h+1}\frac{(t - x_i)_+{}^h}{W'(x_i)}\right\} \ ,$$

where $w(x) = (x - x_0)\cdot g(x)$. Therefore by the definition of a B-spline function of order h+1 for the knot sequence

$\{x_i, i=0,\ldots,h+1\}$, $h!\cdot B(t)$ is a B-spline function multiplied by $(h+1)^{-1}g(x_0)$ and

$$(h+1)!\cdot\int B(t)\ dt = g(x_0) = (x_0 - x_1)\ldots(x_0 - x_{h+1}).$$

<u>LEMMA 2.3.2.</u> $$\sum_{2\leq i_1<\ldots<i_r\leq h+1} z_{i_1}\ldots z_{i_r} = \frac{1}{2}\binom{2h-r+1}{r}$$

where z_k, $k=1,\ldots,h+1$ are as defined in (2.3.1).

In the following we shall use some results about the ultraspherical polynomial which can be found in Szegö(1959).

<u>Proof:</u> Let $z_k' = \cos((h+2-k)\pi/(h+1))$, $k=2,\ldots,h+1$, which are the roots of the Tchebyscheff polynomial of the second kind $U_h(x)$ on $[-1,1]$. Then

$$2^{-h}U_h(x) = (x - z_2')\ldots(x - z_{h+1}')$$

$$= (x + z_2')\ldots(x + z_{h+1}').$$

The polynomial $U_h(x) = P_h^{(1)}(x)$, where $P_h^{(\alpha)}(x)$ is the ultraspherical polynomial with parameter λ. Moreover

$$P_h^{(\alpha)}(1) = \binom{h+2\alpha-1}{h} \quad .$$

We obtain the following results:

(i) When r = h

$$z_2 \ldots z_{h+1} = \prod_{k=1}^{h} \left(1 + \cos\frac{k}{h+1}\pi\right) = \frac{1}{2^h} U_h(1)$$

$$= \frac{1}{2^h}\binom{h+2-1}{h} = \frac{h+1}{2^h} \quad .$$

(ii) When r = h - 1, the property of the ultraspherical polynomial

$$d/dx P_h^{(\alpha)}(x) = 2\alpha \cdot P_{h-1}^{(\alpha+1)}(x)$$

together with

$$\frac{1}{2^h} \frac{d}{dx} U_h(x) = \sum_{k=2}^{h+1} \prod_{j \neq k} (x + z_j') \quad ,$$

it implies that

$$\frac{1}{2^h} U_h'(1) = \sum_{k=2}^{h+1} \prod_{j \neq k} \left(1 + \cos\frac{h+2-k}{h+1}\pi\right)$$

$$= \sum_{2 \leqq i_1 < \ldots < i_{h-1} \leqq h+1} z_{i_1} \ldots z_{i_{h-1}} \quad .$$

Also

$$U_h'(1) = 2P_{h-1}^{(2)}(1) = 2\binom{h+2}{h-1} \quad ,$$

implies that

$$\sum_{2 \leqq i_2 < \ldots < i_{h-1} \leqq h+1} z_{i_1} \ldots z_{i_{h-1}} = \frac{1}{2^{h-1}}\binom{h+2}{h-1}$$

$$= \frac{1}{2^{h-1}} \binom{2h+1 - (h-1)}{h-1}$$

(iii) In general, for $r = h - k$

$$\frac{d^k}{dx^k} P_h^{(1)}(x) = 2^k (k!) \cdot P_{h-k}^{(k+1)}(x) .$$

Hence,

$$\frac{1}{2^h} \cdot \frac{d^k}{dx^k} P_h^{(1)}(x) \Big|_{x=1} = 2^{-(h-k)} k! \binom{2h+1 - (h-k)}{h-k}.$$

Then comparing with

$$\frac{1}{2^h} \cdot \frac{d^k}{dx^k} U_h(x) \Big|_{x=1} = \frac{h \ldots (h-k+1)}{\binom{h}{h-k}} \sum z_{i_1} \ldots z_{i_{h-k}}$$

$$= k! \sum_{2 \leqq i_1 < \ldots \leqq h+1} z_{i_1} \ldots z_{i_{h-k}}$$

the lemma is proved.

LEMMA 2.3.3. $d/dy(\sum |L_i(x_0,y)|)$

$$= (-1)^{h+1} \frac{x_0}{y^{h+1}} \sum_{k=1}^{h} \left\{ (-1)^{h-k} 2^{2k-1} (h-k+1) \binom{h+k}{h-k+1} x_0^{k-1} y^{h-k} \right\} .$$

Proof: Note that $L_i(x,y)$ are the Lagrange interpolation polynomial on the h+1 points $x_1, \ldots, x_{h+1}$, where $x_i = y/2 \cdot z_i$. Therefore,

$$\sum |L_i(x_0,y)| = \sum (-1)^{i-1}L_i(x_0,y)$$

Let

$$\sum (-1)^{i-1}L_i(x,y) = b_0 + b_1x + \ldots + b_hx^h.$$

Then by the property that,

$$\sum (-1)^{j-1}L_i(x_j,y) = (-1)^{i-1},$$

for $i=1,\ldots,h+1$, we obtain $b_0 = 1$ and can solve for b_k, $k=1,\ldots,h$.

$$(1,b_1,\ldots,b_h) = (1,-1,\ldots,(-1)^h)V^{-1},$$

where $V = (v_{ij})$ is the Vandermonde matrix of $\{x_i, i=1,\ldots,h+1\}$. The inverse of V can be easily computed as

$$v_{ij}^{-1} = \left[\frac{(-1)^{i+j} \prod_{\substack{\mu<\nu \\ \mu,\nu\neq i}} (x_\nu - x_\mu)}{\prod_{1\leqq\mu<\nu\leqq h+1} (x_\nu - x_\mu)} \times \sum_{\substack{2\leqq i_1<\ldots<i_{h+1-j}\leqq h+1 \\ \neq i}} x_{i_1}\cdots x_{i_{h+1-j}}\right]$$

for $i,j=1,\ldots,h+1$.

Denote

$$\Pi_0 = \prod_{1\leqq\mu<\nu\leqq h+1} (x_\nu - x_\mu)$$

$$\Pi_i = \prod_{\substack{1\leqq\mu<\nu\leqq h+1 \\ \neq i}} (x_\nu - x_\mu)$$

for $i=1,\dots,h+1$;

then

$$b_k = \frac{(-1)^h h+1}{\Pi_0} \sum_{i=1} \Pi_i \sum_{\substack{1 \leqq i_1 < \dots < i_{h-k} \leqq h+1 \\ \neq i}} x_{i_1} \cdots x_{i_{h-k}} .$$

Now by using $\cos\alpha - \cos\beta = 2\sin(\alpha + \beta)/2 \sin((\beta - \alpha)/2)$ and arranging terms, we obtain

$$\Pi_1/\Pi_0 = 2^h (y/2)^{-h}(h + 1)$$

$$\frac{\Pi_i}{\Pi_0} = \frac{2 - z_i}{(y/2)^h (h + 1)/2^h}$$

for $i=2,\dots,h+1$; where z_i is as in Lemma 2.3.2.

Hence by (2.5.1) and Lemma 2.3.2. b_k can be written as

$$b_k = \frac{(-1)^k \cdot 2^h}{(y/2)^h (h+1)} \left(\frac{y}{2}\right)^{h-k} \left\{ \sum_{i \neq 1} z_{i_1} \cdots z_{i_{h-k}} + \sum_{i=2}^{h+1} \sum_{\neq i} z_{i_1} \cdots z_{i_{h-k}} \cdot (2 - z_i) \right\} .$$

$$= \frac{(-1)^k 2^h}{(y/2)^k (h+1)} \left\{ (2k + 1) \sum_{1 \leqq i_1 < \dots < i_{h-k} \leqq h+1} z_{i_1} \cdots z_{i_{h-k}} - (h-k+1) \sum_{1 \leqq i_1 < \dots < i_{h-k+1}} z_{i_1} \cdots z_{i_{h-k+1}} \right\}$$

$$= \frac{(-1)^k}{y^k} \cdot \frac{2^{2k-1}}{k} (h - k + 1) \binom{h+k}{h-k+1} , \quad \text{for } k=1,\dots,h.$$

We obtain the following result

$$d/dy(\sum|L_i(x_0,y)| = d/dy(1 + b_1x_0 + \ldots + b^hx_0{}^h)$$

$$= (d/dyb_1)x_0 + \ldots + (d/dyb^h)x_0{}^h,$$

where

$$\frac{d}{dy}b_k = (-1)^{k-1}2^{2k-1}\cdot(h-k+1)\binom{h+k}{h-k+1}y^{-(k+1)}.$$

Thus the lemma is established.

<u>LEMMA 2.3.5.</u> If $2^{-h}U_h(x_0,y) = (x_0 - x_2)\ldots(x_0 - x_{h+1})$ where $x_k = (y/2)z_k$, $k=2,\ldots,h+1$; then

$$d/dy\ 2^hU_h(x_0,y))$$

$$= -\frac{1}{2}\frac{1}{2^{2h}}\sum_{k=1}^{h}(-1)^{h-k}2^{2k-1}(h-k+1)\binom{h+k}{h-k+1}x_0{}^{k-1}y^{h-k}$$

$$= \frac{(-1)^h}{2^{2h+1}}\cdot\frac{y^{h+1}}{x_0}\cdot\frac{d}{dy}\sum_{i=1}^{h+1}|L_i(x_0,y)|\ .$$

Proof: $d/dy(2^{-h}U_h(x_0,y)) = d/dy\{(x_0 - x_2)\ldots(x_0 - x_{h+1})\}$

$$= \sum_{i=2}^{h+1} \prod_{\substack{j=2 \\ \neq i}}^{h+1} (x_0 - x_j)\left[-\frac{1}{2}(1 + \cos\frac{h+2-i}{h+1}\pi)\right]$$

$$= -\frac{1}{2}\left[\sum_{i=2}^{h+1} z_i \prod_{\substack{j=2 \\ \neq i}}^{h+1} (x_0 - x_j)\right]$$

$$= -\left[\frac{1}{2} hx_0^{h-1} - (\frac{2}{2} \sum_{2 \leqq i<j \leqq h+1} z_i z_j)x_0^{h-2}y + \right.$$

$$\left. \frac{(-1)^{h-1}}{2^{h-1}} \sum_{2 \leqq i_1<\ldots<i_h \leqq h+1}^{h} z_{i_1}\ldots z_{i_h} \; x_0^{k-1}y^{h-k}\right]$$

$$= -\frac{1}{2}\sum_{k=1}^{h} \frac{(-1)^{h-k}}{2^{h-k}} \frac{(h-k+1)}{2^{h-k+1}} \binom{h+k}{h-k+1} x_0^{k-1}y^{h-k}$$

$$= (-1)^h 2^{-(h+1)} x_0^{-1} y^{h+1} \{d/dy(\sum|L_i(x_0,y)|\}$$

CHAPTER III

MODEL ROBUST EXTRAPOLATION DESIGN FOR X ON [−1,1]

3.1. INTRODUCTION

In this chapter we shall discuss the model robust extrapolation design on [−1,1] under the same class Γ_0 of functions as in Chapter II

$$\Gamma_0 = \{ f \mid |f^{(h+1)}(x)| \leqq \epsilon, -\infty < x < \infty \}.$$

If we use a linear predictor $\hat{f} = \Sigma a_i y_i$ to extrapolate to $f(x_0), x_0 < -1$, we have the same set up as before. Therefore similar problems arise except that in this chapter the observations are only allowed to be taken in [−1,1].

Later in this section we shall discuss the similarities and differences between the case $X = [0,\infty)$ and the case $X = [-1,1]$. In Section 3.2, the design problem is connected to an approximation problem and a complete class of solutions is specified.

In Section 3.3, the minimizing design for different values of n, ϵ, σ^2 and x_0 is characterized and two examples are given to illustrate how to allocate the observations.

Naturally some properties of the optimal design for $X = [0,\infty)$ still hold in the case that $X = [-1,1]$. Firstly the optimal design has to include the left end point of the interval ,namely -1 for $X = [-1,1]$. Secondly under the restriction that there are exactly h+1 points in the design, the a_i's are determined by the constrains in (2.1.4). They are actually equal to $L_i(x_0)$, the value at x_0 of the ith Lagrange polynomial over the h+1 points. Thirdly, the bias function is a B-spline function multiplied by a constant. Fourthly a complete class of optimal design can be found by solving Problem Q2 in Chapter 2 for different values of c where

$$(-1)^{h+1}.(h+1)!\int B\,dt = c. \tag{3.1.1}$$

In this chapter, we shall work under the restriction that there are exactly h+1 points(the minimal number of points). Once the design points are specified , so are the estimator and the design . Therefore we shall only specify the design points for each design from now on.

One difference between the designs for $X = [0,\infty)$ and for $X = [-1,1]$ is that the values of c we need to consider to form the complete class are in different intervals. For case $X = [0,\infty)$, c belongs to an unbounded open interval interval $(|x_0^{h+1}|,\infty)$. But for case $X = [-1,1]$, it belongs to a bounded half open interval $(c_0,c_1]$, where

$$c_0 = |\ (x_0+1)^{h+1}\ | \tag{3.1.2}$$

$$c_1 = |(x_0+1)(x_0-\cos(\pi/h))\ldots(x_0-\cos((h-1)\pi/h))(x_0-1)|. \tag{3.1.3}$$

Another difference is that for $X = [0,\infty)$ there is only one "type" of design in the complete class of the optimal design, but there are two different "types" of designs in the case of $X = [-1,1]$. Here we use "type" to denote that the design points are from one special kind of polynomial, for example Tchebyscheff polynomials of order h+1 on different closed intervals.

The first lemma we shall prove is that the values of c we need to consider for case $X = [-1,1]$ are in $(c_0, c_1]$.

LEMMA 3.1.1. For any signed measure A with support on $\{-1, x_2, \ldots, x_{h+1}\}$, let

$$c = (-1)^{h+1}(h+1)!\int B\,dt$$

$$= |(x_0+1)(x_0-x_2)\ldots(x_0-x_{h+1})|. \tag{3.1.4}$$

(i) Thus $c > c_0$, c_0 is defined as in (3.1.2.),

(ii) If $c > c_1$, where c_1 is defined as in (3.1.3), then the signed measure A_1 with support on

$$\{-1, \cos((h-1)\pi/h), \ldots, \cos(\pi/h), 1\}$$

is better than A.

Proof: We know that the risk function we want to minimize with respect to A consists of two parts, the variance term and the bias term.

Note that the variance term here is the same as the variance for polynomial extrapolation problem of degree h. It is well known that the design with Tchebyscheff points of order h on [-1,1] is the best design to minimize the variance for the polynomial extrapolation problem. Therefore if A has a larger bias term than A_1, A_1 is better than A.

The lower bound of c is obtained from the fact that $x_i > -1$, for $i \geqq 2$ and

$$c = |(x_0+1)(x_0-x_2)\dots(x_0-x_{h+1})|.$$

Thus we shall consider Problem Q2 with X = [-1,1] and $c_0 < c \leqq c_1$.

Similarly as in Chapter II, by the Lagrangian method, we are to find a polynomial of degree h+1 and a set of points $\{-1,x_2,\dots,x_{h+1}\}$ which satisfies (3.1.4), such that

$$P_{h+1}(x_i) = (-1)^{i-1}, \text{ for } i=1,\dots,h+1,$$

and (3.1.5)

$$P_{h+1}'(x_i) = 0,$$

for all x_i in the interior of [-1,1].

For convenience we call this Problem Q4. If such a solution exists for each c, it is the solution of the design problem.

Let

$$c_2 = |(x_0+1)(x_0-y_2)\cdot\ \ \cdot\ \ \cdot(x_0-y_{h+1})|,$$

where

$$y_i = 2^{-1}(y_0+1)(1+\cos((h+2-i)\pi/(h+1))-1,\ i=1,\ldots,h+1,$$

and

$$y_0 = 4/(1+\cos(\pi/(h+1)) - 1. \qquad (3.1.6)$$

It is clear that for $c_0 < c \leqq c_2$, the Tchebyscheff polynomial of the first kind of order h+1 on $[-1,x_{h+2}]$, $x_{h+2} \leqq y_0$, and the set of Tchebyscheff points on $[-1,x_{h+2}]$ excluding x_{h+2} will serve the purpose just as in Chapter II and it is uniquely determined by c.

For $c_2 \leqq c \leqq c_1$, we shall prove that there exists an unique polynomial of degree h+1 and an unique set of points $x_i, i=1,\ldots,h+1, x_1 = -1$, satisfy the condition in (3.1.5) by turning the problem first into an approximation problem. The approximation problem is as follows:

<u>Problem Q5:</u> Among all polynomials of the (h+1)th degree with leading coefficient unity, which are equal to zero at point v, where $v \geqq v_0$, find that polynomial whose absolute maximum on [-1,1] is a minimum, i.e. determine the values of the coefficient $\alpha_0,\ldots,\alpha_h$, where $\alpha_h v^h+\ldots+\alpha_0=v^{h+1}$ for which

$$\max_{-1\leqq x\leqq 1} |x^{h+1} - (\alpha_h x^h +\ldots+ \alpha_0)| = \text{minimum}.$$

3.2. THE APPROXIMATION PROBLEM

Before going into Problem Q5, we introduce some notation.

E_{h+1} : the Euclidean space of (h+1) dimensions.

$\underset{\sim}{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_h)$: an element of E_{h+1}.

$$P_h(\underset{\sim}{\alpha}, x) = \sum_{i=0}^{h} \alpha_i x^i, \quad \alpha \in E_{h+1}.$$

$$E_{h+1}(v) = \{\underset{\sim}{\alpha} \in E_{h+1} \mid \alpha_h v^h + \alpha_{h-1} v^{h-1} + \ldots + \alpha_0 = v^{h+1}\}$$

where $v \geqq v_0$.

$$D_h(v) = \{P_h(\underset{\sim}{\alpha}, x), \underset{\sim}{\alpha} \in E_{h+1}(v)\}.$$

It is easy to see that $E_{h+1}(v)$ is a closed subset of E_{h+1} and $P_h(\underset{\sim}{\alpha}, x)$ is a polynomial of degree h.

By a similar argument as in Rice(1963) p.24, we have the following theorem about the existence of a solution. The proof is included for completeness.

THEOREM 3.2.1. There exists a polynomial of degree h+1 whose absolute maximum on [-1,1] is a minimum among all polynomials of the (h+1)th degree with leading coefficient unity, which are equal to zero at point v, where $v \geqq v_0$.

PROOF: Let $\{\alpha_k,\ k = 1, \ldots, \infty\}$ be a sequence in $E_{h+1}(v)$ such that

$$\lim_{k\to\infty} \max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha_k}, x)|$$

$$= \inf_{\alpha \in E_{h+1}(v)} \max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha}, x)| = \lambda(v).$$

For k sufficiently large, say $k > k_0$, then

$$\max_{-1 \leqq x \leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha}_k, x)| \leqq \lambda(v) + 1.$$

Hence with $\max_{-1 \leqq x \leqq 1} |x^{h+1}| = 1$,

$$\max_{-1 \leqq x \leqq 1} |P_h(\underset{\sim}{\alpha}_k, x)| \leqq \lambda(v) + 2. \qquad (3.2.1)$$

If we show that (3.2.1) implies that the parameters are bounded, then use the fact that a closed bounded subset of $E_{h+1}(v)$ is compact in $E_{h+1}(v)$ to assert the existence of a minimum in $E_{h+1}(v)$. That the parameters are bounded can be established by the following argument.

(i) For $P_h(\underset{\sim}{\alpha}, x) = \sum_{i=0}^{h} \alpha_i x^i$, there is a $\mu > 0$

such that $\max_i |\alpha_i| = 1$ implies

$$\max_{-1 \leqq x \leqq 1} |P_h(\underset{\sim}{\alpha}, x)| \geqq \mu > 0.$$

$$\max_i |\alpha_i| \geqq (\lambda(v) + 2)/\lambda(v)$$

implies that

$$\max_{-1 \leqq x \leqq 1} |P_h(\underset{\sim}{\alpha}, x)| \geqq \lambda(v) + 2.$$

Hence the sequence $\{\underset{\sim}{\alpha}_k\}$ satisfies

$$|\alpha_{i,k}| \leqq (\lambda(v) + 2)/\mu \text{ for } k \geqq k_0,$$

where $\alpha_{i,k}$ denotes the (i-1)th coordinate of the vector α_k. This guarantees the existence of a solution.

The characterization and uniqueness of the minimum polynomial is stated in Theorem 3.2.2 which can be proved easily by a proof similar to Karlin and Studden(1966) except we are doing it for polynomials in a subset. Again we include the proof for completeness.

THEOREM 3.2.2. The polynomial in $D_h(v)$ minimizing

$$\max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha}, x)| \tag{3.2.2}$$

for any $v \geqq v_0$, is uniquely determined, and the unique polynomial $P_h(\underset{\sim}{\alpha}(v), x)$ minimizing (3.2.2) is characterized by the property that there exist (h+1) points $\{x_i(v), i=1,...,h+1\}$, where

$$(-1=x_1(v)<x_2(v)<...<x_h(v)<x_{h+1}(v)=1)$$

such that

$$(-1)^i\delta\{(x_i(v))^{h+1} - P_h(\underset{\sim}{\alpha}(v), x_h(v))\}$$
$$= \max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha}(v), x)| = \lambda(v) > 0 \tag{3.2.3}$$

for i=1,...,h+1; where δ=+1 or −1.

Proof: The assertion of the uniqueness under (3.2.3) is easily proved with the aid of counting principle where double zeros are counted twice. Indeed, suppose $P_h(\underset{\sim}{\alpha}(v),x)$ fulfills condition (3.2.3). This means that $[-1,1]$ can be divided into at least h segments such that on each segment $P_h(\underset{\sim}{\alpha}(v),x)$ ranges from $x^{h+1} + \lambda(v)$ to $x^{h+1} - \lambda(v)$. In this event any polynomial $q_h(x)$ in $D_h(v)$ satisfying $|x^{h+1} - q_h(x)| \leqq \lambda(v)$, for $x \in [-1,1]$, obviously enjoys the property that $P_h(\underset{\sim}{\alpha}(v),x) - q_h(x)$ possesses at least h zeros (at least one in each segment) where double zeros are counted twice. Oalso by the fact that both $P_h(\underset{\sim}{\alpha}(v),x)$ and $q_h(x)$ are in $D_h(v)$ which means $P_h(\underset{\sim}{\alpha}(v),v) = q_h(v) = v^{h+1}$, $P_h(\underset{\sim}{\alpha}(v),x) - q_h(x)$ possesses another zero at v. This would make $P_h(\underset{\sim}{\alpha}(v),x) - q_h(x)$ possess h+1 zeros. Then by the property of polynomials that there can be at most h zeros, we conclude that $P_h(\underset{\sim}{\alpha}(v),x) = q(x)$.

The above argument proves that the polynomial $P_h(\underset{\sim}{\alpha}(v),x)$ deviates least from x^{h+1} in $D_h(v)$ and is uniquely determined by the property (3.2.3).

We now establish the property (3.2.3) by contradiction.

Suppose $P_h(\underset{\sim}{\alpha}(v),x) = \sum \alpha_i(v)x^i$ in $D_h(v)$ satisfies

$$\lambda(v) = \min_{\alpha \in E_{h+1}(v)} \max_{-1 \leqq x \leqq 1} |x^{h+1} - \sum_{i=0}^{h} \alpha_i x^i|$$

and $x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x)$ takes on the values $\pm\lambda(v)$ alternately at only $k \leqq h$ points. We suppose for definiteness that x^{h+1}

$- P_h(\underset{\sim}{\alpha}(v),x)$ assumes the value $+\lambda(v)$ before it takes the value $-\lambda(v)$. In this case there exist $k-1$ points $y_1,\ldots,y_{k-1}(-1<y_1<\ldots<y_{k+1}<1)$ such that $y_i^{h+1} - P_h(\underset{\sim}{\alpha}(v),y_i) = 0$, $i=1,\ldots,k-1$ and for some $\mu > 0$

$$\lambda(v) \geqq x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x) \geqq -\lambda(v) + \mu$$

for $x \in [-1,y_1]\cup[y_2,y_3] \ldots$

$$\lambda(v) - \mu \geqq x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x) \geqq -\lambda(v)$$

for $x \in [y_1,y_2]\cup[y_3,y_4] \ldots$

By Theorem I.5.2 in Karlin and Studden(1966a), there exists a polynomial $w(x)$ whose only zeros on the closed interval $[-1,v]$ are simple zeros at $y_1,y_2,\ldots,y_{k-1},v$ and in addition $w(x) < 0$, for $-1 \leqq x \leqq y_1$.

If ρ is chosen so that

$$|\rho w(x)| \leqq \lambda(v)/2$$

then

$$|x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x) + \rho w(x)| < \lambda(v), \quad x \in [-1,1].$$

Note that $w(v) = 0$, so

$$w(v) = 0, \ (P_h(\underset{\sim}{\alpha}(v),x) - \rho w(x)) \in D_h(v)$$

Therefore,

$$\min_{\alpha\in E_{h+1}(v)} \max_{-1\leqq x\leqq 1} |x^{h+1} - \sum_{i=0}^{h} \alpha_i x^i| < \lambda(v)$$

contradicting the fact that $\lambda(v)$ is the minimum deviation.

Thus any polynomial of least deviation necessarily obeys the required property.

Note that the two end points $\{-1,1\}$ are included in $\{x_h\}, i=1,\ldots,h+1$, simply because we can not have h of the x_i at the interior of $[-1,1]$. Otherwise $x^{h+1} - P_h(\alpha(v),x)$ would have derivative zero at those h points. Thus from the well known fact only the first kind Tchebyscheff polynomial on an interval $[-1,x_{h+2}]$, where $x_{h+2} < y_0$ and y_0 as in (3.1.6), may have such property. But for those Tchebyscheff polynomials their largest zero points would be less than v_0, i.e. they do not belong to $D_h(v)$, which would force $x_1=-1$ and $x_{h+1}=1$. Therefore the theorem is proved.

Note that

$$\begin{aligned}\lambda(v) &= \min_{\underset{\sim}{\alpha}\in E_{h+1}(v)} \max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha},x)| \\ &\geqq \min_{\underset{\sim}{\alpha}\in E_{h+1}} \max_{-1\leqq x\leqq 1} |x^{h+1} - P_h(\underset{\sim}{\alpha},x)| \\ &= 2^{-h} > 0,\end{aligned}$$

and $P_h(\underset{\sim}{\alpha}(v),x)$ is the unique polynomial which deviates least from x^{h+1} in $D_h(v)$. Thus we may divide $x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x)$ by $\lambda(v)$ and normalize it to be 1. Then let

$$W_h(v,x) = \epsilon(\lambda(v))^{-1}(x^{h+1} - P_h(\underset{\sim}{\alpha}(v),x) \text{ for all } v \geqq v_0,$$

where ϵ is determined so that $W_h(v,-1) = 1$.

Now we shall establish the one to one correspondence between Problem Q4 and Problem Q5 by showing the continuity property of $W_h(v,x)$ in terms of v.

LEMMA 3.2.2. If $v_0 \leqq v_1 < v_2$, then $x_i(v_1) < x_i(v_2)$, $i=1,...,h$.

Proof: (i) If $x_2(v_1) < x_2(v_2)$ and

$$x_i(v_1) < x_i(v_2) < x_{i+1}(v_2) < x_{i+1}(v_1)$$

for some i, $2 \leqq i \leqq h-1$, then $W_h(v_1,x)$ would intersect $W_h(v_2,x)$ three times between $(x_i(v_1),x_{i+1}(v_1))$, (i-2) times between $(x_1(v_1),x_i(v_1)$ and (h-i-1) times between $(x_{i+1}(v_1),x_h(v_1))$. We also know that $x_1(v_1) = x_1(v_2) = -1$, $x_{h+1}(v_1) = x_{h+1}(v_2) = 1$. Therefore in total $W_h(v_1,x)$ and $W_h(v_2,x)$ would intersect in at least h+2 points, which implies that $W_h(v_1,x) \equiv W_h(v_2,x)$. This leads to a contradiction. For the case that there are more than two points of $\{x_i(v_2)\}$ between any of $(x_i(v_1),x_{i+1}(v_1))$, $2 \leqq i \leqq h-1$ or there is none between them, a similar argument holds.

If $x_i(v_1) = x_i(v_2)$, for any $i=2,...,h$, $W_h(v_1,x)-W_h(v_2,x)$ would be counted as having two zeros at $x_i(v_1)$ which also would lead to a contradiction .

(ii)If $x_2(v_1) > x_2(v_2)$, a similar argument holds.

By (i) and (ii), either $x_i(v_1) < x_i(v_2)$ or $x_i(v_2) < x_i(v_1)$, for all $i=2,...,h$.

(iii)If $v_1 < v_2$, then $x_2(v_1) < x_2(v_2)$.

If not, then $x_i(v_2) < x_i(v_1)$,for $i=2,...,h$. They intersect

$h+1$ times in $[-1,1]$, and $DW_h(v_1,1) - DW_h(v_2,1) > 0$, where $DW_h(v_i,x) = d/dx \quad W_h(v_i,x)$, $i=1,2$. This and $W_h(v_1,v_1)=0$, $W_h(v_2,v_2)=0$, where $v_1 < v_2$, imply that they must intersect once again between $(1,v_1)$,which leads to the contradiction. Therefore $x_i(v_1)<x_i(v_2)$, for $i=2,\ldots,h$, if $v_0 \leqq v_1 < v_2$.

<u>LEMMA 3.2.3.</u> As $v_k \downarrow v$ ($\geqq v_0$), from the right, when $k\to\infty$,

$x_i(v_k) \downarrow x_i(v)$, for $i=1,\ldots,h+1$.

Proof: Choose arbitrary $h-1$ points η_i from $(-1,1)$, $i=2,\ldots,h$. For all $v\geqq v_0$, $W_h(v,\eta_i)=\epsilon_i(v)$ for some $\epsilon_i(v)$, $i=2,\ldots,h$, where $|\epsilon_i(v)|\leqq 1$. Together with $W_h(v,-1)=1$, $W_h(v,1)=(-1)^h$, it is easy to see that the coefficients of $W_h(v,x)$ are bounded for all $v\geqq v_0$ because $\epsilon_i(v)$ are bounded for all $v\geqq v_0$.

As $v_k\to v$ from the right, there exists a convergent subsequence $\{v_k'\}$ of $\{v_k\}$ such that $W_h(v_k',x)\to G(x)$ and $G(v)=0$. Also because $x_i(v_k')$ is monotonic and has a lower bound, there exists a subsequence of $\{v_k'\}$ such that

$x_i(v_k'') \downarrow \mu_i$, $i=2,\ldots,h$.

If $\mu_i > x_i(v_k'')$ for some i, then there are two polynomials equal to 0 at v both having the property characterized by (3.2.2), which contradicts the uniqueness. Thus $x_i(v_k'') \downarrow x_i(v)$ as $v_k'' \downarrow v$.

Again as $v_k\to\infty$ by the boundedness discussed in Lemma 2.3.3, we know there exists a convergent subsequence $\{v_k'\}$ of $\{v_k\}$ such that $W_h(v_k',x)\to G(x)$. In the next lemma we shall see that $G(x)$ is exactly the Tchebyscheff polynomial of the first kind of order h+1 in [-1,1].

LEMMA 3.2.4. Suppose $W_h(v_k,x) \to G(x)$ as $v_k \to \infty$, then $G(x)=T_h(x)$, where $T_h(x)$ is the Tchebyscheff polynomial of the first kind of order h in [-1,1].

Proof: By the characterization of $W_h(v_k,x)$ for each v_k, there are h zeros between (-1,1), say $\eta_{k,i}$, i=1,...,h and one zero at v_k. Then it can be written that

$$W_h(v_k,x) = \beta_k(x - v_k)\cdot\prod_i(x - \eta_{k,i})$$

the product Π is from i=1 to h.

Since $W_h(v_k,-1) = 1$ for all v_k,

$$\beta_k = 1 / ((-1)^{h+1}(1 + v_k)\cdot\prod_i(1 + \eta_{k,i})).$$

Therefore as $v_k \to \infty$,

$$W_h(v_k,x) = \beta_k\cdot\prod_i(x - \eta_{k,i})x - \beta_k v_k\cdot\prod_i(x - x_{k,i})$$

converges to a polynomial of degree h. By uniqueness and the characterization, we know it is the Tchebyscheff polynomial of the first kind of degree h in [-1,1].

The following theorem will connect the solution of the Approximation problem with the design problem.

THEOREM 3.2.3. There is a one to one correspondence between the minimal solution of Problem Q4(the design problem) for each c , where $c_2 \leqq c < c_1$, and the minimal solution of Problem Q5(the Approximation problem) for $v_2 \leqq v_k < \infty$). There c_2

corresponds to v_0 and c_1 corresponds to the limiting case as $v_k \to \infty$.

Proof: It follows quite easily from the continuity property of the polynomial $W_h(v_k, x)$ in terms of v_k and the two extremes of v_k correspond to the two bounds of c.

3.3. OPTIMAL ROBUST EXTRAPOLATION DESIGN ON [-1,1]

From the discussions of Section 3.1 and Section 3.2, we have obtained a complete class of optimal robust extrapolation design for $X=[-1,1]$. There are two 'types' of possible solution in the complete class, one is the first h+1 Tchebyscheff points on $[-1, x_{h+2}]$ for some $x_{h+2} \leqq y_0$, y_0 is defined as in (3.1.6); the other is $\{x_i(v)\}$, $i=1,\ldots,h+1$, as obtained from Section 3.2 for some $v \geqq v_0$. For the second type $\{-1,1\}$ are both in the design.

The following theorems will give us where the optimal design should allocate for different values of n, ϵ, σ^2 and x_0.

THEOREM 3.3.1. Let $x_\tau = x_0+1$ and $y_\tau = y_0+1$. For given $\rho=\sigma^2((h+1)!)^2/(n\epsilon^2)$, if $\rho \leqq \rho_0$, where

$$\rho_0 = \frac{(-1)^h x_\tau (2^{-h} U_h(x_\tau, y_\tau))}{2^{2h+1}(\Sigma |L_i(x_\tau, y_\tau)|)} y_\tau^{h+1} \qquad (3.3.1)$$

then the optimal design on exactly h+1 points is on the first (h+1)th Tchebyscheff points on [-1,y], where (y+1) is the unique positive solution of (2.3.1), with x_0 replaced by x_τ.

<u>Proof</u> Once the monotonicity with respect to ρ of the unique positive root of (2.3.1) with x_0 replaced by x_τ is established, the theorem is easily proved. This can be seen as follows. In view of Theorem 2.3.2 , the design with the first h+1 Tchebyscheff points of order h+1 on $[0,y(\rho)]$ is the optimal design for extrapolating x_τ. Therefore if $\rho \leqq \rho_0$, then the unique positive solution of (2.3.1) is $\leqq y_\tau$ and the design points would fall into [0,2]. Finally by the location invariant property of the problem, we may transform it to the interval [-1,1] and the theorem is established.

Now we prove the monotonicity as follows.

Let $0 < \rho_1 < \rho_2$. From Lemma 2.3.4, we see that $d(\Sigma|L_i(x_\tau,y)|)/dy$ is always negative for $x_\tau < 0$. Therefore

$$H(\rho_1,y(\rho_2)) < H(\rho_2,y(\rho_2)) < 0,$$

where

$$H(\rho_i,y) = d/dy[\rho_i(\Sigma|L_i(x_\tau,y)|)^2 + x_\tau^2(2^{-h}U_h(x_\tau,y))^2]. \qquad (3.3.2)$$

Again by the fact that there is only one positive root of (3.3.2) and the root is the minimum point of $R(A(y))$ with $\rho=\rho_1$, therefore, $0 < y(\rho_1) < y(\rho_2)$.

<u>THEOREM 3.3.2.</u> If $\rho > \rho_0$, then the optimal design is on $\{-1,x_2(v),\ldots,x_h(v),1\}$ for some $v \geqq v_0$, where $\{x_i(v)\}$, $i=1,\ldots,h+1$, are defined as in Section 3.2.

Proof: The proof follows easily from that of the previous theorem and is omitted.

Before giving examples of finding the optimal design for h=2 or h=3 through the procedure described above, we would like to know how the optimal design changes as x_0 changes.

By observing the two quantities $\Sigma|L_i(x_\tau,y)|$ and $(2^{-h}U_h(x_\tau,y))$ in $R(A(y))$, where $x_\tau=x_0+1$ and $R(A(y))$ are as defined in (2.3.2), we find that both of them are functions of $(-y/x_\tau)$. Therefore equation (2.3.1) can be rewritten as

$$\rho' = \frac{\rho}{(x_\tau)^{2(h+1)}} = \frac{(2^{-h}U_h(u))}{2^{2h+1}(\Sigma|L_i(u)|)}\, u^{h+1} . \qquad (3.3.3)$$

It implies that there is a one to one correspondence between ρ' and the positive root of (3.3.3). Thus if $\rho < \rho_0$, and ρ' is determined then the root is independent of x_τ. This property will be used to allocate the optimal design.

Now we shall give two examples concerning the cases h = 2 and h = 3.

EXAMPLE 3.3.1. For h = 2, if we want to extrapolate to x_0, $x_0 < -1$ from [-1,1]. Then the steps for finding the optimal design for a given value of ρ are as follows:

Step(i) Compare ρ with ρ_0 which can be found easily by (3.3.3) with $y=y_\tau$. We provide some of the critical values ρ_0 in Table 3.1 and Figure 3.1.

$\rho \leqq \rho_0$, goes to Step (ii) ;

if not, goes to Step (iii) .

Step(ii) Let $\rho' = \rho/(x_\tau)^6$, and find the corresponding u from Table 3.2, where Table 3.2 is made out of (3.3.3). In the case of h = 2, (3.3.3) can be written as

$$\rho' = \frac{1}{2^5} \frac{(1 + u/4)(1 + 3u/4)}{(1 + 12/u + 16/u^2)} . \quad u^3$$

where $u = -y/x_\tau$.

Then the optimal design is on $\{-1, -(x_\tau.u/4) - 1, -(3.x_\tau.u)/4 - 1\}$. with corresponding weight $m_i = |L_i(x_0)|/ \sum|L_i(x_0)|$, i=1,2,3.

Step(iii) According to Theorem 3.3.2, both end points are in the optimal design. Therefore for h = 2, there is only one point left to be determined which can be found explicitly through some calculation. In Table 3.3., some x_2 are given for different ρ and x_0.

For instance, if $x_0 = -3.00$, n = 20, $\sigma^2 = 1$, $\epsilon = .9987$, then $\rho = 1.865 > \rho_0 = .665$. By Table 3.4 x_2 is -1.65. Therefore, the optimal design is on -1, -1.65, 1, with weight .3892, .4712, .1395 respectively.

EXAMPLE 3.3.2. For h =3, all the steps are similar to Example 3.3.1. except in Step(ii) and Step(iii) we look up different tables and have different optimal designs. In Step(ii), we look up u from Table 3.4 or Figure 3.3 and

(3.3.3) in this case is

$$\rho' = \frac{1}{2^7}\,\frac{s(u)}{t(u)}\;.\;u^4\;,$$

where $s(u)=(1 + ((2-\sqrt{2})/4).u)(1 + u/2)(1 + ((2+\sqrt{2})/4).u)$

$t(u)=1 + 24/u + 80/u^2 + 64/u^3$.

Then the optimal design is on

$\{-1,(2-\sqrt{2})/4.(-x_\tau.u) - 1,(-x_\tau.u)/4 - 1,(2+\sqrt{2})/4.(-x_\tau.u) - 1\}$.

In Step(iii), we can not give an explicit formula for finding the other two points, but we do provide Table 3.5 of the optimal design points (x_2,x_3) for different x_0 and ρ.

Table 3.1.

Critical values ρ_0 for h = 2, h = 3.

x_0	ρ_0(h=2)	ρ_0(h=3)	x_0	ρ_0(h=2)	ρ_0(h=3)
−1.10	.065	.013	−6.00	1.474	.262
−1.50	.227	.040	−6.50	1.607	.286
−2.00	.382	.067	−7.00	1.740	.309
−2.50	.526	.092	−7.50	1.872	.333
−3.00	.665	.117	−8.00	2.005	.357
−3.50	.802	.142	−8.50	2.138	.381
−4.00	.938	.166	−9.00	2.270	.405
−4.50	1.073	.190	−9.50	2.402	.428
−5.00	1.207	.214	−10.00	2.535	.452
−5.50	1.340	.238	−10.50	2.667	.476

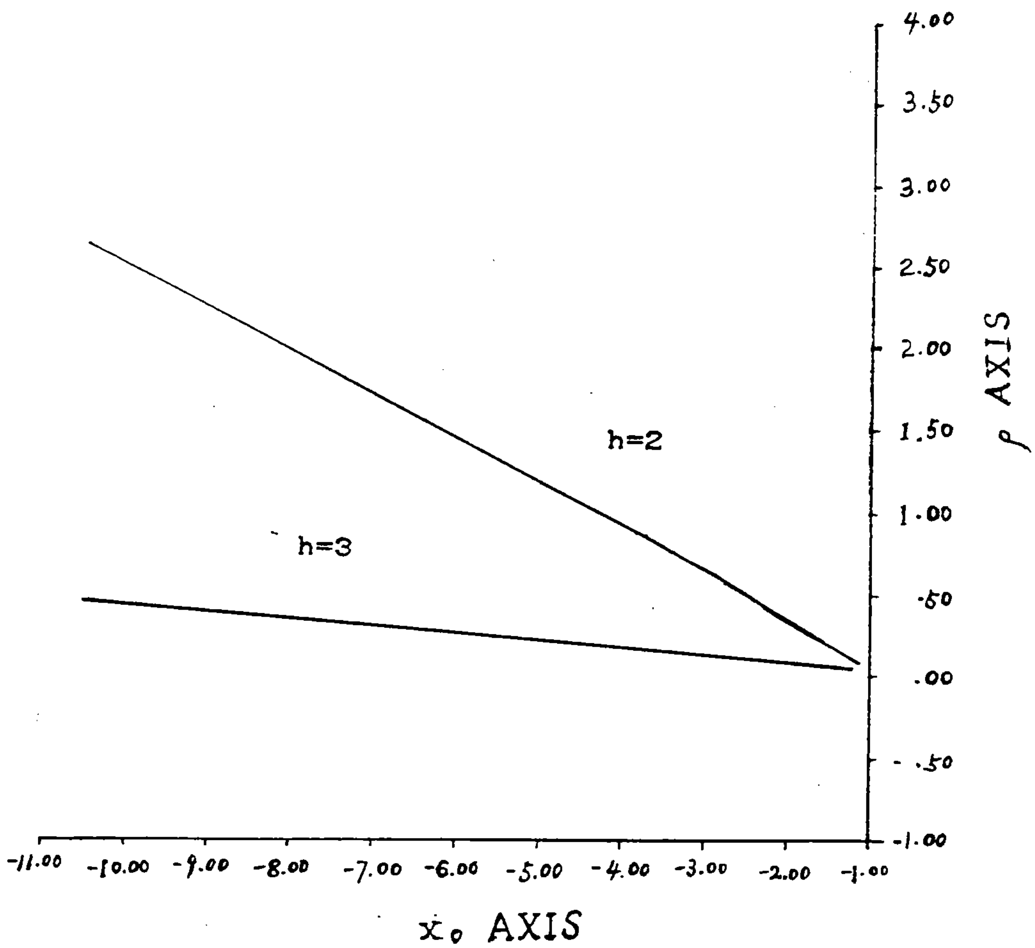

Figure 3.1. Critical values ρ v.s. x_0 for h=2, h=3.

Table 3.2.

Corresponding ρ and u for h = 2

ρ	u	ρ	u
1.000E-08	.0871	1.049E-02	1.3355
2.000E-08	.1000	2.097E-02	1.5269
4.000E-08	.1148	4.194E-02	1.7451
8.000E-08	.1317	8.389E-02	1.9939
1.600E-07	.1512	1.678E-01	2.2776
3.200E-07	.1735	3.355E-01	2.6009
6.400E-07	.1991	6.711E-01	2.9695
1.280E-06	.2284	1.342E+00	3.3895
2.560E-06	.2620	2.684E+00	3.8683
5.120E-06	.3005	5.369E+00	4.4142
1.024E-05	.3445	1.074E+01	5.0366
2.048E-05	.3950	2.147E+01	5.7465
4.096E-05	.4527	4.295E+01	6.5565
8.192E-05	.5188	8.590E+01	7.4809
1.638E-04	.5943	1.718E+02	8.5366
3.277E-04	.6807	3.436E+02	9.7425
6.554E-04	.7794	6.872E+02	11.1209
1.311E-03	.8921	1.374E+02	12.6970
2.621E-03	1.0209	2.749E+03	14.5002
5.243E-03	1.1678	5.498E+03	16.5642

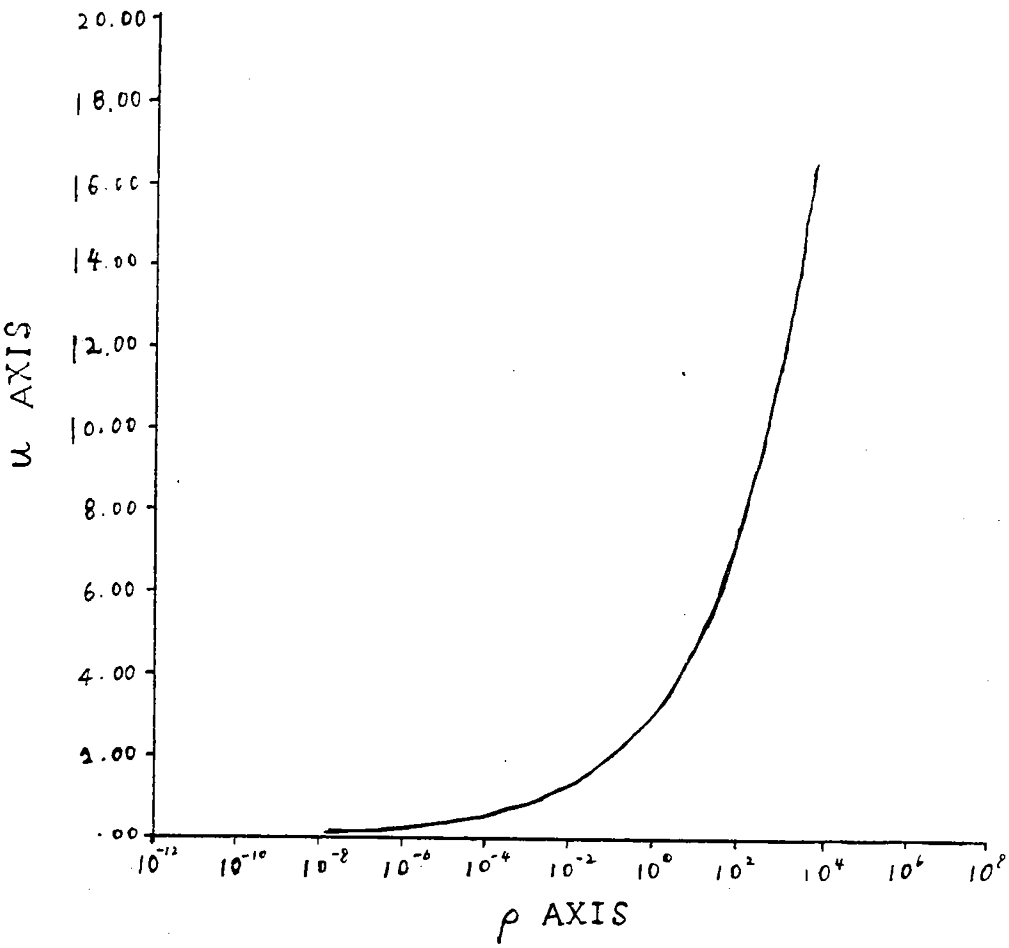

Figure 3.2. Corresponding ρ and u, h=2.

Table 3.3.

Optimal value of x_2 for different ρ, x_0 for h = 2

x_0	ρ_0	ρ_0+.2	+.40	+.60	+.80	+1.00	+1.20	+1.40	+1.60	+1.80
-1.10	-.333	-.130	-.080	-.058	-.045	-.037	-.031	-.027	-.024	-.021
-1.50	-.333	-.229	-.174	-.139	-.116	-.099	-.087	-.077	-.069	-.063
-2.00	-.333	-.261	-.214	-.181	-.156	-.137	-.122	-.110	-.100	-.091
-2.50	-.333	-.277	-.236	-.205	-.181	-.162	-.146	-.133	-.122	-.113
-3.00	-.333	-.287	-.251	-.222	-.200	-.181	-.165	-.152	-.140	-.130
-3.50	-.333	-.293	-.261	-.235	-.214	-.196	-.180	-.167	-.155	-.145
-4.00	-.333	-.298	-.270	-.246	-.225	-.208	-.193	-.179	-.168	-.158
-4.50	-.333	-.302	-.276	-.254	-.234	-.218	-.203	-.190	-.179	-.169
-5.00	-.333	-.305	-.281	-.260	-.242	-.226	-.212	-.200	-.188	-.178
-5.50	-.333	-.308	-.286	-.266	-.249	-.234	-.220	-.208	-.197	-.187

Table 3.4.

Corresponding ρ and u for h = 3

ρ	u	ρ	u
1.000E-10	.1319	1.074E-01	2.3159
4.000E-10	.1600	2.147E-01	2.5555
1.600E-09	.1940	4.295E-01	2.8207
6.400E-09	.2351	8.590E-01	3.1142
2.560E-08	.2846	1.718E+00	3.4387
1.024E-07	.3443	3.436E+00	3.7973
4.096E-07	.4161	6.872E+00	4.1934
1.638E-06	.5026	1.374E+01	4.6305
6.554E-06	.6069	2.749E+01	5.1128
2.621E-05	.7328	5.498E+01	5.6444
1.049E-04	.8849	1.100E+02	6.2303
2.097E-04	.9727	2.199E+02	6.8755
4.194E-04	1.0695	4.398E+02	7.5860
8.389E-04	1.1762	8.796E+02	8.3681
1.678E-03	1.2941	1.759E+03	9.2289
3.355E-03	1.4244	3.518E+03	10.1761
6.711E-03	1.5684	7.037E+03	11.2183
1.342E-02	1.7278	1.407E+04	12.3651
2.684E-02	1.9042	2.815E+04	13.6269
5.369E-02	2.0996	5.629E+04	15.0155

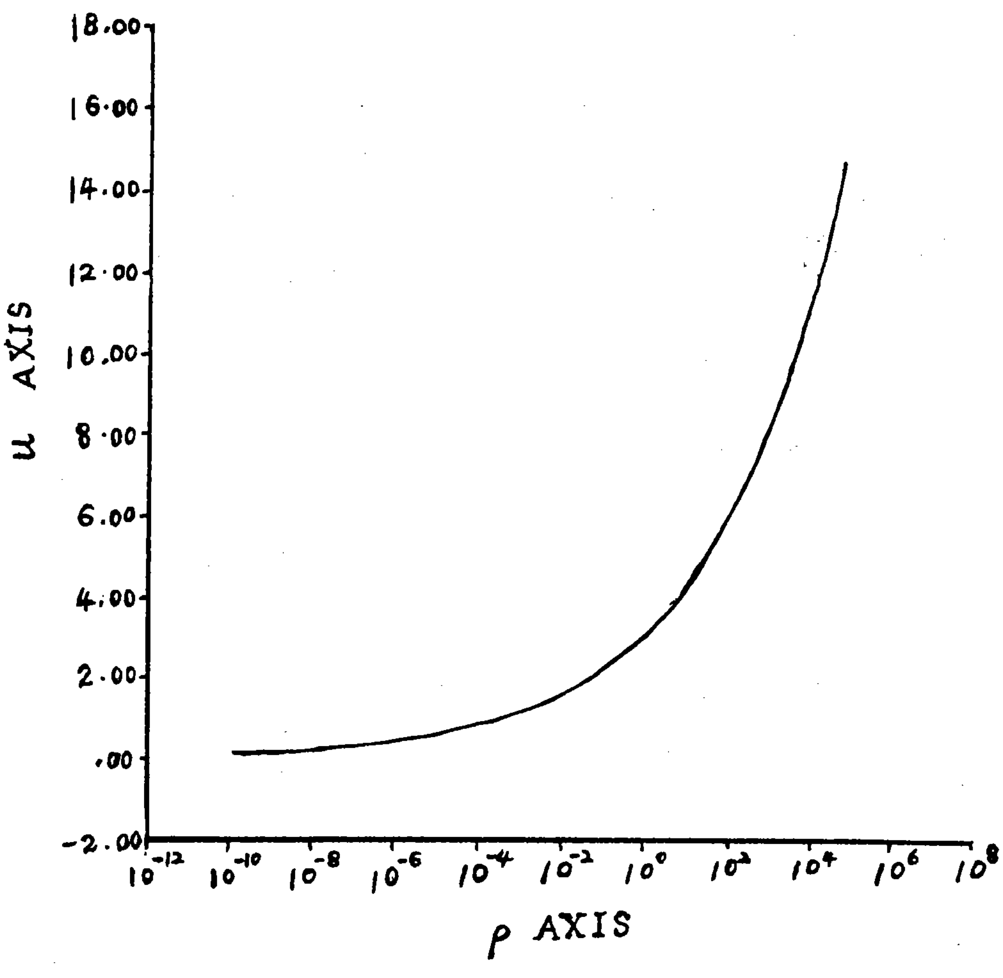

Figure 3.3. Corresponding ρ and u, h=3.

Table 3.5.

Optimal values of (x_2, x_3) for different ρ and x_0 for $h = 3$

x_0	ρ_0	ρ_0+.10	ρ_0+.20	ρ_0+.30	ρ_0+.40
-1.10	(-.657, .172)	(-.544, .446)	(-.526, .471)	(-.518, .480)	(-.514, .485)
-1.50	(-.657, .172)	(-.583, .377)	(-.557, .427)	(-.543, .448)	(-.535, .460)
-2.00	(-.657, .172)	(-.601, .337)	(-.574, .395)	(-.559, .424)	(-.548, .440)
-2.50	(-.657, .172)	(-.611, .311)	(-.586, .371)	(-.570, .404)	(-.559, .424)
-3.00	(-.657, .172)	(-.618, .292)	(-.595, .353)	(-.579, .387)	(-.567, .409)
-3.50	(-.657, .172)	(-.623, .279)	(-.601, .337)	(-.586, .373)	(-.574, .396)
-4.00	(-.657, .172)	(-.627, .268)	(-.606, .324)	(-.591, .360)	(-.580, .384)
-4.50	(-.657, .172)	(-.630, .312)	(-.611, .258)	(-.596, .348)	(-.585, .374)
-5.00	(-.657, .172)	(-.632, .251)	(-.614, .303)	(-.601, .338)	(-.590, .364)
-5.50	(-.657, .172)	(-.634, .245)	(-.618, .294)	(-.604, .329)	(-.594, .355)

CHAPTER IV

EXACT D-OPTIMAL DESIGNS FOR POLYNOMIAL REGRESSION

4.1. INTRODUCTION

As introduced in Chapter I, we are interested in finding the exact D-optimal designs for polynomial regression of degree k−1 on [−1,1] in this chapter.

Recall that for each $x \in [-1,1]$ an experiment can be performed. The outcome is a random variable $y(x)$, with mean value $\theta' f(x)$, where $\theta = (\theta_0, \ldots, \theta_{k-1})'$, $f(x) = (1, x, \ldots, x^{k-1})'$, and a common variance σ^2.

Suppose that n uncorrelated observations on the response $y(x)$ are to be obtained at levels $x_1, \ldots, x_n$. Let $Y = [y(x_1), \ldots, y(x_n)]'$, $X = (x_{ij})$, where $x_{ij} = (x_i)^j$, $1 \leqq i \leqq n$, $0 \leqq j \leqq k-1$. The unknown parameter vector $\theta = (\theta_0, \ldots, \theta_{k-1})'$ is estimated by the classical least squares estimator $\hat{\theta} = (X'X)^{-1}X'Y$. Then $E\hat{\theta} = \theta$ and $\mathrm{cov}(\hat{\theta}) = \sigma^2 (X'X)^{-1}$.

An exact design specifies a probability measure ξ on [−1,1] which concentrates mass p_i at x_i, $i=1,\ldots,r$, where $p_i n = m_i$, $i=1,\ldots r$, are integers. The information matrix in this case is $M(\xi) = X'X$. An exact design ξ^* is said to be D-optimal if ξ^* maximizes $|M(\xi)|$ among all the exact designs ξ on $[-1,1]$.

Salaevskii(1966) conjectures that an D-optimal design ξ^* distributes observations as evenly as possible among the k support points of the approximate D-optimal design. Constantine and Studden(1981) have provided a simpler proof of Salaevskii's result that the conjecture holds for sufficiently large n. Both of their proofs are based on the Taylor series expansion of the determinant of the information matrix with respect to the unknown exact D-optimal design points. Gaffke and Krafft(1982) have proved Salaevskii's conjecture for quadratic regression for all $n \geqq 3$ quite simply. Their proof is based on the geometric-arithmetic means inequality of the information matrix. Since we shall follow the approach of Gaffke and Krafft(1982) for general case, their clever idea is briefly described in the following.

If Salaevskii's conjecture holds true for quadratic regression for all $n \geqq 3$, then there are three different solutions ξ_1, ξ_2, ξ_3. If $n = 3p+1$, ξ_i puts p+1 points on one of x_i^*, where $x_1^* = -1$, $x_2^* = 0$, $x_3^* = 1$, and puts p points on the other two points. If $n = 3p+2$, ξ_i puts p points on one of x_i^*, where $x_1^* = -1$, $x_2^* = 0$, $x_3^* = 1$, and puts p+1 points on the other two points.

By the geometric-arithmetic means inequality, we have for any design ξ and for $1 \leqq \nu \leqq 3$,

$$\det M(\xi) \leqq \det M(\xi_\nu)(3^{-1}\operatorname{tr}M(\xi)M^{-1}(\xi_\nu))^3. \qquad (4.1.1)$$

Note that $\det M(\xi_\nu) = \Pi p_i |F|^2$ is independent of ν where F^2 will be defined later. So if we Let $c_n = \det M(\xi_\nu)$, then

$$\det M(\xi) \leqq c_n \min_{1\leqq\nu\leqq3} (3^{-1}\operatorname{tr}M(\xi)M^{-1}(\xi_\nu))^3.$$

To establish the theorem one has to show that for all ξ,

$$\min_{1\leqq\nu\leqq3} ((\operatorname{tr}M(\xi)M^{-1}(\xi_\nu)) \leqq 3. \qquad (4.1.2)$$

Let

$$d(x,\xi_\nu) = f'(x)M_f^{-1}(\xi)f(x). \qquad (4.1.3)$$

since $\operatorname{tr}M(\xi)M^{-1}(\xi_\nu) = \sum d(x_i,\xi_\nu)$, (4.1.2) was proved by showing that

$$\min_{1\leqq\nu\leqq3} \sum_{i=1}^{n} d(x_i,\xi_\nu) \leqq 3 ,$$

or

$$\min_{1\leqq\nu\leqq3} \sum_{i=1}^{n} p(p+1)d(x_i,\xi_\nu) \leqq 3p(p+1).$$

Since ξ_ν has support on -1,0,1(the approximate D-optimal design points for quadratic polynomial), we know that $p\cdot d(x_i,\xi_\nu) \leqq 1$ for all x_i.

Let r_ν be the number of x_i such that $(p+1)d(x_i,\xi_\nu) \leqq 1$.

Then

$$\sum_{i=1}^{n} p(p+1)d(x_i,\xi_\nu)$$

$$\leqq pr_\nu + (n-r_\nu)(p+1)$$

$$\leqq n(p+1) - r_\nu \ .$$

If $n = 3p+1$ then it turns out for each x_i, $(p+1)d(x_i,\xi_\nu) \leqq 1$ for some ν, which will imply $\max r_\nu \geqq p+1$.

Then

$$\min_{1\leqq\nu\leqq3} \{n(p+1) - r_\nu\} \leqq 3p(p+1) \ , \qquad (4.1.4)$$

and the result follows. If $n = 3p+2$, then for each x_i, $(p+1)d(x_i,\xi_\nu) \leqq 1$ for at least 2 curves and $\max r_\nu \geqq 2p+2$. Then (4.1.2) will hold.

For general k and $n = kp+t$, it will be shown that $\max r_\nu \geqq tp+t$. Then (4.1.2) will hold with 3 replaced by k on the right hand side of the inequality, and ν is from 1 to m, where m is the usual binomial coefficient with value $k!/(t!(k-t)!)$.

In Section 4.2, following the new approach Salaevskii's result for large sample case is proved. For polynomials of degree $\leqq$ 9, by the method we use to prove Salaevskii's conjecture for large sample case, we are able to give the value of N such that for $n \geqq N$, Salaevskii's conjecture is true. We give a list of the value of N for polynomial of degree $\leqq$ 9 in the following:

degree (k−1)	3	4	5	6	7	8	9
N	12	15	24	36	48	63	81

For cubic regression we already know that for $n = 4$, $n = 8$ the exact D-optimal design coincides with the approximate D-optimal design. Moreover for $n=9,10,11$, Salaevskii's conjecture is proved by using a modification of the new apporach, which is presented in Section 4.3.

4.2. ANOTHER LOOK AT SALAEVSKII'S RESULT FOR LARGE SAMPLE CASE

Again recall that the approximate D-optimal design for polynomial regression of degree k−1 on [−1,1] concentrates equal mass at the roots of $(1-x^2)P_{k-1}'(x)$, where $P_{k-1}(x)$ is the (k−1)th Legendre polynomial. At this point, we introduce the following notation:

(i)Let $\{x_1^*,\dots,x_k^*\}$ be the set of the approximate D-optimal design points for polynomial regression of degree k−1 on [−1,1].

(ii)Let $g_i(x), i=1,\dots,k$, be the fundamental Lagrange interpolation polynomials induced by the points $\{x_1^*,\dots,x_k^*\}$.

Then it is chear that

$$\sum_{i=1}^{k} g_i^2(x) \leqq 1, \quad x \in [-1,1]. \tag{4.2.1}$$

Suppose that we have n observations, where $n = kp+t$, for some $1 \leqq t \leqq k-1$.

(iii)Let $m = \binom{k}{t}$, $q = \binom{k-1}{t-1}$, where the brackets denote the usual binomial coefficient.

Then there are m designs which distribute the n observations as evenly as possible among $\{x_1^*, \ldots, x_k^*\}$. Thus let ξ_ν be one of such designs, say ξ_ν puts p+1 points on each of $x_{\nu_1}^*, x_{\nu_2}^*, \ldots, x_{\nu_t}^*$, where $x_{\nu_1}^* < \ldots < x_{\nu_t}^*$, and $\{\nu_i, i=1,\ldots,t\} \in \{1,\ldots,k\}$, and puts p points on each of the points in $S-\{x_{\nu_i}^*, 1 \leqq i \leqq t\}$.

(iv)Let $d(x,\xi_\nu)$ be the variance function of design ξ_ν, $1 \leqq \nu \leqq m$.

Then $d(x,\xi_\nu)$ can be written as

$$d(x,\xi_\nu) = f'(x)M_f^{-1}(\xi_\nu)f(x) = g'(x)M_g^{-1}(\xi_\nu)g(x)$$

$$= \sum_{i=1}^{k} \frac{g_i^2(x)}{p_{i,\nu}}, \qquad (4.2.2)$$

where

$$\begin{cases} p_{i,\nu} = p+1, \text{ if } i \in \{\nu_1, \ldots, \nu_t\}, \\ p_{i,\nu} = p, \quad \text{otherwise.} \end{cases}$$

Now by the geometric-arithmetic means inequality for the general case where the polynomial regression function is of degree k-1, we have for any design ξ and for all ξ_ν, $1 \leqq \nu \leqq m$,

$$\det M(\xi) \leqq \det M(\xi_\nu)(k^{-1}\mathrm{tr}M(\xi)M^{-1}(\xi_\nu))^k.$$

Let $c_n = \det M(\xi_\nu) = n^{-k}\cdot p^{k-t}(p+1)^t\cdot F^2$,

where

$$F^2 = \prod_{1\leqq i<j\leqq k} (x_j^* - x_i^*)^2 \qquad (4.2.3)$$

is the square of the Vandermonde determinant corresponding to the points $x_1^*,\ldots,x_k^*$. This in turn implies that for any ξ

$$\det M(\xi) \leqq c_n \min_{1\leqq\nu\leqq m} (k^{-1}\mathrm{tr}M(\xi)M^{-1}(\xi_\nu))^k. \qquad (4.2.4)$$

As in the quadratic case, we need to show that for all ξ

$$\min_{1\leqq\nu\leqq m} (\mathrm{tr}M(\xi)M^{-1}(\xi_\nu)) \leqq k. \qquad (4.2.5)$$

In the following, two lemmas which are useful for proving (4.2.5) are proved. The first lemma is a generalization of the inequality used in Gaffke and Krafft(1982). More notation are needed and introduced below.

It can easily be checked that $g_i^2(x)$ and $g_{i+1}^2(x)$ intersect only once in $[x_i^*, x_{i+1}^*]$, for $i=1,\ldots,k-1$. Therefore,

(v) let $\{x_0',\ldots,x_k'\}$ be the set of points where x_i' is the unique intersection point of $g_i^2(x)$ and $g_{i+1}^2(x)$ in $[x_i^*, x_{i+1}^*]$ for $i=1,\ldots,k-1$; and $x_0'=x_1^*$, $x_k'=x_k^*$.

From (i) it is clear that $x_0'=x_1^*= -1, x_k'=x_k^*=1$.

(vi)Let $R^2(x) = g_i^2(x)$, for all $x \in [x_{i-1}', x_i']$, $1 \leqq i \leqq k$.

Lemma 4.2.1. There exists p_0 such that

$$\sum_{i=1}^{k} g_i^2(x) \leqq \frac{p_0}{p_0+1} + \frac{1}{p_0+1} R^2(x) \leqq \frac{p}{p+1} + \frac{1}{p+1} R^2(x) \quad (4.2.6)$$

for $p \geqq p_0$, and for every $x \in [-1,1]$.

Proof: By the fact that S is the set of the approximate D-optimal design points, it is known that $x_2^*, \ldots, x_{k-1}^*$ are the local maxima of $\sum g_i^2(x)$ in $[-1,1]$. Also

$$\sum_{i=1}^{k} g_i^2(x_j^*) = 1, \quad j=1,\ldots,k.$$

This in turn implies that

$$\sum_{i=1}^{k} g_i^2(x_j') < 1 \quad , \quad j=1,\ldots,k-1; \quad (4.2.7)$$

and for $2 \leqq j \leqq k-1$

$$\frac{d}{dx} \sum_{i=1}^{k} g_i^2(x) \bigg|_{x=x_j^*} = 0,$$

$$\frac{d^2}{dx^2} \sum_{i=1}^{k} g_i^2(x) \bigg|_{x=x_j^*} < 0. \quad (4.2.8)$$

Let

$$R_{j,p}(x) = \sum_{i=1}^{k} g_i^2(x) - \frac{1}{p+1} g_j^2(x),$$

for $j=1,\ldots,k$; and for all positive integer p.

Then

$$R_{j,p}(x_j^*) = p/p+1 \ ,$$

and for i=2,...,k-1, j=1,...,k,

$$d/dx(R_{j,p}(x))\Big|_{x=x_j^*} = 0 \ .$$

From (4.2.8), for fixed i, j, there exists a $p_{i,j,1}$ such that for $p \geqq p_{i,j,1}$,

$$\frac{d^2}{dx^2}(R_{j,p}(x))\Big|_{x=x_i^*} < 0,$$

i.e. x_i^* is a local maximum of $R_{j,p}(x)$.

Let $p_{j,1} = \max_{2 \leqq i \leqq k-1} p_{i,j,1}$, then for $p \geqq p_{j,1}$, it is easy to see that $x_1^*,\ldots,x_{k-1}^*$ are the only local maxima for function $R_{j,p}(x)$. In order to find the absolute maxima of $R_{j,p}(x)$ for $p \geqq p_{j,1}$ at the interval $[x_j', x_{j+1}']$, we need to check the values of $R_{j,p}(x)$ at the boundary points x_j' and x_{j+1}'.

In view of (4.2.7), there exists a $p_{j,2}$ such that for $p \geqq p_{j,2}$ and for i = j or i = j+1,

$$R_{j,p}(x_i') \leqq p/(p+1),$$

Thus for every j, $1 \leqq j \leqq k$, we have

$$R_{j,p}(x) \leqq p/(p+1),$$

where $x \in [x_{j-1}', x_j']$ and $p \geqq \max(p_{j,1}, p_{j,2})$.

Let

$$p_0 = \max_{1 \leqq j \leqq k} (p_{j,1}, p_{j,2}), \qquad (4.2.9)$$

then for $p \geqq p_0$,

$$R_{j,p}(x) \leqq p/(p+1)$$

for every $x \in [x_{j-1}', x_j']$, and for all $j=1,\ldots,k$. Therefore the first inequality in (4.2.6) is proved. The second inequality follows from the fact that $R^2(x) \leqq 1$ for $x \in [-1,1]$.

<u>LEMMA 4.2.2.</u> For $p \geqq p_0$, p_0 is as defined in (4.2.9)

$$(p+1)d(x,\xi_\nu) \leqq 1, \text{ for } x \in [x_{\nu_j -1}', x_{\nu_j}'],$$

$1 \leqq \nu_j \leqq t$ and for all $1 \leqq \nu \leqq m$.

<u>Proof:</u> Divide the interval $[-1,1]$ into k subintervals by the points x_i', $i=0,\ldots,k$, such that

$$[-1,1] = \bigcup_{i=1}^{k} [x_{i-1}', x_i'].$$

In view of (4.2.2), for every ν, $1 \leqq \nu \leqq m$,

$$p(p+1)d(x,\xi_\nu) = (p+1)\sum_{i=1}^{k} g_i^2(x) - \sum_{j=1}^{t} g_{\nu_j}^2(x)$$

where $\{\nu_1,\ldots,\nu_t\}$ is a subset of $\{1,\ldots k\}$.

If $1 \in \{\nu_1, \ldots, \nu_t\}$, then by Lemma 4.2.1 it implies that for $x \in [-1, x_1']$,

$$p(p+1)d(x,\xi_\nu) = (p+1)\sum_{i=1}^{k} g_i^2(x) - g_1^2(x) - \sum_{j=2}^{t} g_{\nu_j}^2(x)$$

$$\leqq p + R^2(x) - g_1^2(x) - \sum_{j=2}^{t} g_{\nu_j}^2(x)$$

$$\leqq p \ .$$

Similarly, it can be proved that,

$(p+1)d(x,\xi_\nu) \leqq 1$, for $x \in [x_{\nu_j - 1}', x_{\nu_j}']$.

Thus the lemma is proved.

In view of Lemma 4.2.2, we see that there are at least q out of m functions of $\{d(x,\xi_\nu)\}$ such that

$$(p+1)d(x,\xi_\nu) \leqq 1 \ ,$$

for every $x \in [x_{i-1}', x_i']$ and $1 \leqq i \leqq k$. In other words, for every $x \in [-1,1]$, there are at least q indices ν such that $(p+1)d(x,\xi_\nu) \leqq 1$. The particular indices depend on the interval x is in.

Now we are ready to prove the main theorem.

THEOREM 4.2.1. For $n = kp+t \geqq N = kp_0$, where $1 \leqq t \leqq k-1$, and p_0 as defined in (4.2.9), there is an exact D-optimal design $\xi^* = \xi_\nu$, for some ν, $1 \leqq \nu \leqq m$.

Proof: As discussed previously, once (4.2.5) holds for all ξ, the theorem is true. From (4.2.2), we have

$$\operatorname{tr} M(\xi)M^{-1}(\xi_\nu) = \sum_{i=1}^{n} f'(x_i)M^{-1}(\xi_\nu)f(x_i)$$

$$= \sum_{i=1}^{n} d(x_i,\xi_\nu), \qquad (4.2.10)$$

where x_i, $i=1,\ldots,n$ are the support points of a design ξ.

Now let r_ν be the number of i's for which

$$(p+1)d(x_i,\xi_\nu) \leqq 1, \quad 1 \leqq \nu \leqq m.$$

In the following, we shall prove that for $p \geqq p_0$

$$\max_{1\leqq\nu\leqq m} r_\nu \geqq tp+t, \qquad (4.2.11)$$

then

$$p(p+1) \min_{1\leqq\nu\leqq m} \left(\sum_{i=1}^{n} d(x_i,\xi_\nu)\right) \leqq \min_{1\leqq\nu\leqq m} (pr_\nu + (n-r_\nu)(p+1))$$

$$= (kp+t)(p+1) - \max_{1\leqq\nu\leqq m} r_\nu$$

$$\leqq kp(p+1).$$

Together with (4.2.10), (4.2.5) is obtained for all ξ.

Now we prove (4.2.11) for $p \geqq p_0$. Let λ_i be the number of observations in the interval $[x_{i-1}', x_i']$, for $1 \leqq i \leqq k$, and let $\{\lambda_{[i]}\}$ be the ordered number of $\{\lambda_i\}$ such that

$$\lambda_{[1]} \geqq \lambda_{[2]} \geqq \dots \geqq \lambda_{[k]}.$$

therefore,

$$n = kp+t = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} \lambda_{[i]}.$$

From Lemma 4.2.2., we claim that for $p \geqq p_0$, there exists an s_ν, where $s_\nu = \sum_{i=1}^{t} \lambda_{[i]}$, such that $r_\nu \geqq s_\nu \geqq tp+t$.

For $p \geqq p_0$, from Lemma 4.2.2, we know $r_\nu \geqq s_\nu$.
The second inequality can be obtained as follows:

If $s_\nu < tp+t$, it means that

$$\sum_{i=1}^{t} \lambda_{[i]} < tp+t \Rightarrow \lambda_{[t]} < (p+1),$$

and

$$\sum_{i=t+1}^{k} \lambda_{[i]} = n - \sum_{i=1} \lambda_{[i]}$$

$$= (kp+t) - \sum_{i=1}^{t} \lambda_{[i]}$$

$$> (k - t)p$$

$$\Rightarrow \lambda_{[t+1]} > p+1,$$

which contradicts that $\lambda_{[t]} > \lambda_{[t+1]}$. Therefore the theorem is proved.

It is natural to ask that how many observations are sufficient for the theorem to be true . In the following, for polynomial of degree from 3 to 9, we give a list of the smallest p_0 and N values such that Lemma 4.2.1 holds. Therefore Theorem 4.2.1 is true for $n \geqq N = kp_0$.

degree (k −1)	3	4	5	6	7	8	9
p_0	3	3	4	5	6	7	8
N	12	15	24	36	48	63	81

4.3. CUBIC REGRESSION FOR N = 9, 10, 11

In the case of cubic regression, for p = 2, it can be checked that inequality (4.2.5) holds for $x \in [-1,-\epsilon] \cup [\epsilon,1]$, for a constant $\epsilon \leqq 0.03$. Also the left side of the inequality is greater than the right side by only a small amount for $x \in [-\epsilon,\epsilon]$. Therefore after making some modifications of the proof in the previous section, we are able to prove the conjecture for n = 9,10,11.

Again by the geometric-arithmetic means inequality, we have for any design ξ', ξ'' with 4 or more design points,

$$\det M(\xi') \leqq \det M(\xi'')(k^{-1}\mathrm{tr}M^{-1}(\xi'')M(\xi'))^k,$$

which can be written as

$$\det M(\xi') \leqq \det M(\xi_\nu)\left\{\frac{\det M(\xi'')}{\det M(\xi_\nu)}\left(\frac{1}{k}\mathrm{tr}M^{-1}(\xi'')M(\xi')\right)^k\right\}. \quad (4.3.1)$$

For convenience, we divide the interval $[-1,1]$ into 5 subintervals. Let

$$I_1 = [-1, x_1'],\ I_2 = [x_1', -0.03],$$
$$I_3 = [0.03, x_3'],\ I_4 = [x_3', 1],$$
$$I_5 = [-0.03, 0.03].$$

Case(i) $n = 9$. Then there are 4 designs ξ_ν, $1 \leqq \nu \leqq 4$, where ξ_ν puts 3 points on x_ν^* and 2 points on each of $S-\{x_\nu^*\}$.

From the proof of Lemma 4.2.2. it is easy to see that for $x \in I_\nu$,

$$3\cdot d(x,\xi_\nu) \leqq 1,\ 1 \leqq \nu \leqq 4.$$

Therefore for any design ξ with 3 points or more in any one of the intervals I_ν, $1 \leqq \nu \leqq 4$, from the proof of Theorem 4.2.1 , we have

$$\det M(\xi) \leqq \det M(\xi_\nu).$$

Since in the set $[-1,1] - I_5$ there can be at most 8 support points of design ξ, then there is at least one point in I_5. Now consider designs with one or more design points in I_5.

Let ξ'' be the exact design with 10 observations, which has support points x_1^*, x_2^*, x_3^*, x_4^* with corresponding weight 2/10, 3/10, 3/10, 2/10. Then

$$\mathrm{tr}M^{-1}(\xi'')M(\xi) = \frac{1}{9}\sum_{i=1}^{9}\sum_{\nu=1}^{4}\frac{g_\nu^2(x_i)}{p_\nu}$$

$$= \frac{1}{9}\sum_{i=1}^{9}\left\{\frac{10}{2}g_1^2(x_i) + \frac{10}{3}g_2^2(x_i) + \frac{10}{3}g_3^2(x_i) + \frac{10}{2}g_4^2(x_i)\right\}$$

$$= \frac{10}{9}\sum_{i=1}^{9}\left\{\frac{g_1^2(x_i)}{2} + \frac{g_2^2(x_i)}{3} + \frac{g_3^2(x_i)}{3} + \frac{g_4^2(x_i)}{2}\right\}$$

$$= \frac{10}{9}\sum_{i=1}^{9} d(x_i,\xi'')$$

where $d(x,\xi'')$ is the variance function of design ξ''.

It is clear that $3\cdot d(x,\xi'') \leqq 3/2$ for every $x \in [-1,1]$. Also it can be checked that

$$3\cdot d(x,\xi'') \leqq 1 \text{ , for } x \in I_2 \cup I_3,$$

and since

$$3\cdot d(x,\xi'') = (25/64)\{2(x^2 - 1)^2(5x^2 + 1) + 3(x^2 - 1/5)^2(x^2 + 1)\},$$

we have

$$3\cdot\cdot d(x,\xi'') \leqq 3\cdot d(.04,\xi'') < .840167 \text{ , for } x \in I_5.$$

Therefore,

$$\sum 3d(x_i,\xi'') \leqq (.840167 + 4 + 6) = (10.840167).$$

Together with the fact that for $1 \leqq \nu \leqq 4$

$$\det M(\xi_\nu) = 9^{-4} \cdot 2^3 \cdot 3 \cdot F^2,$$

$$\det M(\xi'') = 10^{-4} \cdot 3^2 \cdot 2^2 \cdot F^2$$

where F^2 is as defined in (4.2.3),
it implies that

$$\det M(\xi) \leqq \det M(\xi_\nu).$$

Therefore ξ_ν ,$1 \leqq \nu \leqq 4$, are the exact D-optimal designs.

For cases(ii) and (iii) with n = 11,10, the conjecture can be proved along the lines of that of case(i). We outline the proof as follows: Let $I = [-1,1]$.

Case(ii) n = 11. First eliminate those designs with more than 9 points in any of the following four intervals $I - I_1$, $I - I_4$, $I - I_2 - I_2 - I_5$, $I - I_3 - I_5$ by the property of the variance function of design ξ_ν with 11 design points. Then choose ξ'' to be the one with support on x_1^*, x_2^*, x_3^*, x_4^* and with corresponding weights 3/10, 2/10, 2/10, 3/10.

Case(iii) n = 10. Similarly, first eliminate those designs with more than 6 points in any one of the following four intervals, $I_1 \cup I_2$, $I_3 \cup I_4$, $I - I_1 - I_4$, $I - I_2 - I_3$. Then choose ξ'' to be the one with 11 design points on x_1^*, x_2^*, x_3^*, x_4^* and with corresponding weights either 3/11, 3/11, 3/11, 2/11 or 2/11, 3/11, 3/11, 3/11.

<u>REMARK.</u> For n = 5,6,7, we do not have a proof yet. It will be studied in the future work.

LIST OF REFERENCES

[1] Agarwal, G. and Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines, Ann. Statist., 8, 1307-1325.

[2] Box, G.E.P. and Draper, N.R. (1959). A basis for the selection of a response surface design, J. Am. Statist. Assoc., 54, 622-654.

[3] Constantine, K.B., and Studden, W.J. (1981). Optimal exact designs for polynomial regression. Purdue Mimeo Series #81-35.

[4] De Boor, Carl (1978). A practical guide to splines. Springer-Verlag, New York.

[5] Fedorov, V.V. (1972). Theorey of Optimal Experiments, Translated by Studden, W.J. and Klimko, E.M., Academic Press, New York.

[6] Hoel, P.G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Statist., 29, 294-299.

[7] Huber, P.J. (1975). Robustness and designs. A Survey of Statistical Design and Linear Models. North-holland, Amsterdam, 287-303.

[8] Karlin, S. and Studden, W.J. (1966a). Tchebycheff Systems: with Applications in Analysis and Statistics. Interscience, New York.

[9] Karlin, S. and Studden, W.J. (1966b). Optimal Experimental designs. Ann. Math. Statist., 37, 783-815.

[10] Karson, N.J., Manson, A.R., and Hader, R.j. (1969). Minimum bias estimation and experimental designs for response surfaces. Technometrics, 11, 461-476.

[11] Kiefer, J. (1973). Optimum designs for fitting biased multiresponse surfaces. Multivariate Analysis III(P.R. Krishnaiah, Ed.) 287-297. Academic Press, New York.

[12] Kiefer, J. (1980). Designs for extrapolation when bias is present. Multivariate Analysis V. North Holland, Amsterdam, 79-93.

[13] Li, K.-C. (1981). Robust regression designs when the design space consists of finitely many points. Purdue Mimeo Series 81-45.

[14] Li, K.-C., and Notz, W.I. (1982). Robust designs for nearly linear regression. J. Statist. Plann. Inference, 6, 135-152.

[15] Marcus, M.B., and Sacks, J. (1976). Robust designs for regression problems. Statistical Decision Theory and Related Topics II, Academic Press, New York, 245-268.

[16] Notz, W.I. (1979). Optimal robust designs for some regression problems. Purdue Mimeo Series 80-29.

[17] Pesotchinsky, L. (1982). Optimal robust designs: linear regression in R_k. Ann. Statist., 10, 511-525.

[18] Sacks, J., and Ylvisaker, D. (1982). Model robust design in regression. Northwestern University, Discussion paper number 60.

[19] Spruill, M.C. (1982). Optimal designs for minimax extrapolation. Preprint.

[20] Studden, W.J. (1968). Optimal designs on Tchebyscheff points. Ann. Statist., 39, 1435-1447.

[21] Studden, W.J. (1978). Designs for large polynomial regression. Commun. Statist.-Theor. Meth. A7(14), 1391-1397.

[22] Szegö, G. (1959). Orthogonal Polynomials(revised edition) Amer. Math. Soc. Colloq. Publ., 23, New York.

[23] Weisner, L. (1938). Theory of Equations. The Mcmillian Company, New York.