OPTIMAL RESTRICTED EXPERIMENTAL DESIGNS

by

Kenneth Constantine[1]
Purdue University

Mimeograph Series #81-36

Department of Statistics

August 1981

[1]Research supported by an NSF Graduate Fellowship and National Science Foundation Grant No. MCS-7901707.

CHAPTER I

INTRODUCTION

1.1 Restricted Experimental Design

Consider the linear model $y(x) = \theta' f(x) + \varepsilon$, which is assumed to hold for each "level" x belonging to a compact space $\mathcal{X}$. Here $f(x) = [f_0(x), f_1(x), \ldots, f_n(x)]'$ denotes a (column) vector of n + 1 known regression functions which are linearly independent and continuous on $\mathcal{X}$, $\theta = (\theta_0, \theta_1, \ldots, \theta_n)'$ denotes the unknown (column) vector parameter of interest, ε denotes the unobservable "error" random variable with expected value 0 and unknown variance σ^2 for all $x \in \mathcal{X}$, and $y(x)$ denotes the observable "response" random variable at level x.

Suppose that the θ_i's are to be estimated using N uncorrelated observations on the response random variable $y(x)$ at levels $x_1, \ldots, x_N$ of x. The linear model for these data is $Y = X\theta + e$, where $Y = [y(x_1), \ldots, y(x_N)]'$, where

$$X = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \vdots & & \vdots \\ f_0(x_N) & f_1(x_N) & \cdots & f_n(x_N) \end{bmatrix},$$

and where $e = (\varepsilon_1, \ldots, \varepsilon_N)'$. Here the vector e has expected value $E(e) = 0$ and covariance matrix $Cov(e) = \sigma^2 I_N$. It will be assumed throughout that estimation of the θ_i's is to be based on the classical estimator $\hat{\theta} = (X'X)^{-1}X'Y$. For this estimator, $E(\hat{\theta}) = \theta$ and

$Cov(\hat{\theta}) = \sigma^2(X'X)^{-1}$. Of course, if rank(X) < n + 1, then the inverse operation should be replaced by a generalized inverse.

Since the covariance matrix of $\hat{\theta}$ is a measure of the joint imprecision of the $\hat{\theta}_i$'s, a reasonable goal is to "minimize" this matrix. The sense in which it is minimized will define an optimality criterion which should be appropriate to the particular problem. Because σ^2 is unknown, the preceeding goal may be best realized by choosing those levels $x_1,\ldots,x_N$ which minimize $(X'X)^{-1}$ in the appropriate sense. The design problem is to determine the optimal choice of these levels.

This design problem may be better posed once it is reformulated. To that end, let the levels of x at which observations are taken be relabelled so that $x_1,\ldots,x_r$ denote distinct levels at which $n_1,\ldots,n_r$ observations are made (respectively). Here $n_1 +\ldots+ n_r = N$.

Now let an exact design ξ^N be a probability measure on $\mathcal{X}$ which concentrates mass n_i/N at x_i, for $i = 1,\ldots,r$. Thus the support of an exact design prescribes the levels of x at which observations are to be made and the corresponding masses prescribe the proportions of the N observations which are to be allocated to each of these levels.

Next let the information matrix (per observation) of an exact design be $M(\xi^N) = X'X/N$. Note that $Cov(\hat{\theta}) = \sigma^2 M^{-1}(\xi^N)/N$. Hence a reformulation of the design problem is to choose an exact design measure ξ^N on $\mathcal{X}$ which minimizes $M^{-1}(\xi^N)$ (in the appropriate sense). Note also that

$$M_{k\ell}(\xi^N) = \sum_{j=1}^{N} X_{jk}X_{j\ell}/N$$

$$= \sum_{i=1}^{r} f_k(x_i)f_\ell(x_i)n_i/N$$

$$= \int_{\mathcal{X}} f_k(x)f_\ell(x)d\xi^N(x), \quad \text{for } 0 \leq k,\ \ell \leq n.$$

Hence $M(\xi^N) = \int_{\mathcal{X}} f(x)f(x)'d\xi^N(x)$ (where the prime denotes transpose).

The determination of an optimal design is more tractable mathematically if the class of designs is extended to include all "approximate designs". Here an approximate design ξ is any probability measure on $\mathcal{X}$. Thus an approximate design prescribes where in $\mathcal{X}$ observations are to be taken and how they are to be allocated. It will be implicitly assumed that all design measures are defined on the Borel subsets of $\mathcal{X}$.

Now let the information matrix of any (approximate) design ξ be $M(\xi) = \int_{\mathcal{X}} f(x)f(x)'d\xi(x)$. Thus the approximate design problem is to determine a design ξ which minimizes $M^{-1}(\xi)$ in an appropriate sense.

The limitation of the approximate design approach should be noted. In practice, only an exact design can be implemented. It may happen that an optimal approximate design is not exact for certain values of N; it might even be exact for no value of N. This problem will be less troublesome for large N, but for any value of N an optimal approximate design should strongly suggest how to best allocate observations throughout $\mathcal{X}$. With this limitation in mind, designs will henceforth be assumed to be approximate.

The standard approach to optimal design problems has been to determine a best approximate design $\xi \in \Xi_0$, where Ξ_0 is the set of all probability measures on $\mathcal{X}$. That is, Ξ_0 is the closed convex hull of $\{\delta_x | x \in \mathcal{X}\}$, where δ_x denotes a measure concentrating mass 1 at the point x. Thus the standard assumption has been that any proportion of the observations may be allocated to any point of $\mathcal{X}$. Under this assumption, various optimality questions have been answered for certain choices of f(x) and $\mathcal{X}$ by Elfving ('52), de la Garza ('54), Guest ('58), Hoel ('58), Kiefer and Wolfowitz ('59), Kiefer and Wolfowitz ('60), Hoel and Levine ('64), Karlin and Studden ('66b), and many others.

The present interest is to address some of the same questions under a restriction on the class Ξ of allowable designs. This restriction will be assumed to be such that Ξ is still closed and convex. Several types of design restrictions will now be introduced.

1.2 The E-Problem

Consider the case that E is a compact subset of $\mathcal{X}$ and $\Xi = \{\xi | \xi(E) = 1\}$. Equivalently, $\Xi = \{\xi | \xi(\mathcal{X}-E) = 0\}$. Such a restricted class of designs would be appropriate to a situation where observations cannot be obtained for $x \in \mathcal{X} - E$ but can be obtained at will for $x \in E$.

It should be noted that if $\mathcal{X}$ is replaced by $\mathcal{X}' = E$ and Ξ is replaced by $\Xi_0' = \{\delta_x | x \in \mathcal{X}'\}$, then the E-problem fits the unrestricted design framework. The reason for considering the E-problem in a restricted design setting is to better compare the answers obtained for optimality questions on E to those answers either known for $\mathcal{X}$ or obtained for other design restrictions. It will typically be the case that $\mathcal{X}$ is connected but E is not.

As an example of the E-problem, consider the (univariate) setting with $\mathcal{X} = [a,b]$ and $\mathcal{X} - E = \bigcup_{j=1}^{J} G_j$, where the $G_j = (\alpha_j,\beta_j)$ are disjoint open intervals. Krein and Nudelman ('77) refer to this example as the "E_J-problem". The E_J-problem, with $J = 1$, might arise if one control instrument can only be used to set levels of $x \in [a,\alpha_1]$ and another can only be used to set levels of $x \in [\beta_1,b]$. Another example of the E-problem is $E = \{x_1,\ldots,x_k\}$. This example might arise if a digital instrument is to be used to control the level of x.

1.3 The (φ,ψ)-Problem

Consider the case that φ and ψ are measures on $\mathcal{X}$ with the properties that $\varphi(\mathcal{X}) < 1 < \psi(\mathcal{X})$ and $d\varphi \leq d\psi$. That is, $\varphi(A) \leq \psi(A)$ for any Borel set $A \subset \mathcal{X}$. Equivalently, φ is absolutely continuous with respect to ψ and $\frac{d\varphi}{d\psi}(x) \leq 1$ for all $x \in \text{Support}(\psi)$. Now let $\Xi = \{\xi | d\varphi \leq d\xi \leq d\psi\}$. Thus φ and ψ impose lower and upper "limits" on how observations may be allocated within $\mathcal{X}$.

Remark 1.3.1: An equivalent version of this problem is to set $\eta = \xi - \varphi$, set $\nu = \psi - \varphi$, and consider $\mathcal{H} = \{\eta | 0 \leq d\eta \leq d\nu\}$. Then the requirement that $\xi(\mathcal{X}) = 1$ is replaced by the requirement that $\eta(\mathcal{X}) = 1 - \varphi(\mathcal{X})$. In this sense the (φ,ψ)-problem is equivalent to the $(0,\nu)$-problem.

For an example of the (φ,ψ)-problem, let $\mathcal{X} = \{x_1,\ldots,x_k\}$, let $\varphi_j = \varphi(\{x_j\})$ and $\psi_j = \psi(\{x_j\})$ for $j = 1,\ldots,k$, let $0 \leq \varphi_j \leq \psi_j$ for each j, and let $\varphi_1 + \ldots + \varphi_k < 1 < \psi_1 + \ldots + \psi_k$. This yields the setting of regression on a discrete set with lower and upper limits imposed on the proportion of observations taken at each point. The discrete nature of $\mathcal{X}$ might be due to a digital control instrument, the lower limits φ_j might be due to observations already obtained, and the upper limits ψ_j might represent limitations on the instrument's capability to provide observations. For any point x_j with no lower (upper) limit, it would be appropriate to set $\varphi_j = 0 (\psi_j = 1)$.

For another example of the (φ,ψ)-problem, let $\mathcal{X} = [a,b]$, let $d\varphi = 0$, and let $d\psi(x) = \gamma dx$ for $\gamma \geq 1/(b-a)$. This example might arise from a setting where x denotes time and observations are obtained continuously in time. The parameter γ restricts the rate at which they may be obtained.

1.4. The p-Problem

Consider the case that $\{G_\omega | \omega \in \Omega\}$ is a collection of disjoint sets which are open in $\mathcal{X}$ and $p_\omega \in (0,1)$ for each $\omega \in \Omega$. Let $\Xi = \{\xi | \xi(G_\omega) \leq p_\omega;\ \omega \in \Omega\}$. Thus only a limited proportion of the observations may be obtained from each G_ω. However, within each G_ω observations may be allocated at will provided that the restriction $\xi(G_\omega) \leq p_\omega$ is satisfied. It may be noted that the G_ω need to be open in order for Ξ to be closed.

As an example of the p-problem, consider the (univariate) setting with $\mathcal{X} = [a,b]$, with $\Omega = \{1,\ldots,s\}$, with $G_j = (\alpha_j,\beta_j)$, and with $0 < p_j < 1$ for $j = 1,\ldots,s$. This example, with $s = 1$ and $G_1 = (\alpha_1,b]$ might arise in an industrial setting where x denotes time and the design restriction is due to limited manpower on a night shift.

1.5 The Marginal Restriction Problem

Consider the case that $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and a measure ξ_1^* on $\mathcal{X}_1$ is prescribed. Let $\Xi = \{\xi | \xi(A \times \mathcal{X}_2) = \xi_1^*(A)$ for all Borel sets $A \subset \mathcal{X}_1\}$. Thus designs are restricted to have first marginal ξ_1^*. Note that since $d\xi(x_1,x_2) = d\xi_1^*(x_1)d\xi(x_2|x_1)$, the restricted design problem is to determine the optimal allocation of observations with respect to x_2 for each $x_1 \in \text{Support}(\xi_1^*)$.

Cook and Thibodeau ('80) relate an example of the marginal restriction problem in which $x_1 \in \mathcal{X}_1 = \{0, 1, \ldots, 10\}$ denotes a pretest score and $x_2 \in \mathcal{X}_2 = \{0,1\}$ denotes a variable which indicates whether a subject belongs to the treatment or control group. For this example, the design problem is how best to assign subjects to one of the two groups given their pretest scores.

It should be noted at this point that a design restriction might also be a hybrid of any of these four basic types of restriction.

1.6 Comparison of Design Restrictions

The following general framework encompasses all of the preceding design restrictions. First consider Ξ_0 to be embedded in the linear space of all signed measures on $\mathcal{X}$. Suppose that for each $\gamma \in \Gamma$, L_γ is a continuous linear functional on this space and λ_γ is a real number. Then each of the preceeding design restrictions may be seen to satisfy

$$\Xi = \{\xi | L_\gamma(\xi) \leq \lambda_\gamma;\ \gamma \in \Gamma\} \cap \Xi_0 \tag{1.6.1}$$

by appropriate choice of Γ, the L_γ, and the λ_γ.

For the E-problem, the choices $\Gamma = \{1\}$, $L_1(\xi) = \int_{\mathcal{X}-E} d\xi$, and $\lambda_1 = 0$ confirm that $\Xi = \{\xi | \xi(\mathcal{X} - E) = 0\}$ satisfies (1.6.1).

Recall that, according to remark 1.3.1, the (φ,ψ)-problem is equivalent to the $(0,\nu)$-problem, where $\nu = \psi - \varphi$. That is, it suffices to consider $\mathcal{H} = \{\eta | d\eta \leq d\nu\}$. Now set $\mathcal{H}_0 = \{\eta | \eta(\mathcal{X}) = 1 - \varphi(\mathcal{X})\}$, set $\Gamma = \text{Support}(\nu)$, set each $L_x(\eta) = \frac{d\eta}{d\nu}(x)$, and set each $\lambda_x = 1$. These choices confirm that $\mathcal{H} = \{\eta | L_x(\eta) \leq \lambda_x;\ x \in \mathcal{X}\} \cap \mathcal{H}_0$.

For the p-problem, the choices $\Gamma = \Omega$, $L_\omega(\xi) = \int_{G_\omega} d\xi$, and $\lambda_\omega = p_\omega$ confirm that $\Xi = \{\xi | \xi(G_\omega) \leq p_\omega;\ \omega \in \Omega\}$ satisfies (1.6.1).

For the marginal restriction problem, the choices $\Gamma = \{1,2\} \times \mathcal{B}_1$ (where $\mathcal{B}_1$ is the class of all Borel subsets of $\mathcal{X}_1$), $L_{1A}(\xi) = -L_{2A}(\xi) = -\int_{A \times \mathcal{X}_2} d\xi$, and $\lambda_{1A} = -\lambda_{2A} = -\xi_1^*(A)$ confirm that $\Xi = \{\xi | \xi(A \times \mathcal{X}_2) = \xi_1^*(A)$ for all Borel sets $A \subset \mathcal{X}_1\}$ satisfies (1.6.1).

The linearity and continuity of the L_γ are immediate in each case. These two properties also imply that Ξ is convex and closed in general.

Some rather loose relationships between the different design restrictions may now be indicated.

Consider first the marginal restriction problem with $\text{Support}(\xi_1^*) = \{s_1,\ldots,s_k\} = \mathcal{X}_1 \subset \mathbb{R}$. This situation may be viewed as a special case of the p-problem with $\Omega = \{1,\ldots,k\}$, with $p_j = \xi_1^*(\{s_j\})$, and with $G_j = \{s_j\} \times \mathcal{X}_2$.

Consider next the E-problem where $E \subset \mathbb{R}$. This situation may be viewed as a limiting case of the p-problem with $\Omega = \{1\}$, with $G_1 = \mathcal{X} - E$, and with $p_1 = 0$. It may also be loosely viewed as a limiting case, as $\gamma \uparrow \infty$, of the (φ,ψ)-problem with $d\varphi = 0$ and

$$d\psi(x) = \begin{cases} 0 & x \in \mathcal{X} - E \\ \gamma dx & x \in E. \end{cases}$$

Consider finally the (φ,ψ)-problem where $\varphi = 0$ and $\text{Support}(\psi) = \{x_1,\ldots,x_k\} = \mathcal{X} \subset \mathbb{R}$. This situation may be viewed as a special case of the p-problem with $\Omega = \{1,\ldots,k\}$, with $p_j = \psi(\{x_j\})$, and with $G_j = \{x_j\}$. In the case that the support of ψ is not finite, the (φ,ψ)-problem might be loosely viewed as a limiting case of the p-problem as a partition of ψ's support becomes arbitrarily fine.

1.7 Program of Text

In chapter II, the problem of characterizing the admissible designs for polynomial regression on an interval is first embedded in a more general moment space problem which is of interest in its own right. For the E-problem and (φ,ψ)-problem restrictions, theorems 2.2.6 and 2.3.5 characterize the admissible designs in terms of their "index" $I(\xi)$. For the p-problem restriction, theorem 2.4.4 shows that an analogous condition on $I(\xi)$ is necessary for admissibility. Theorems 2.4.8 and 2.4.12 show that this condition is also sufficient in two special cases. For each of these three types of design restrictions, the index $I(\xi)$ "counts" the support points of a design in a way appropriate to the particular restriction. In sections 2.2 and 2.4, consideration is also given to the secondary problem of improving upon an inadmissible design.

In chapter III, the problem of determining a D-optimal restricted experimental design is considered. Theorem 3.1.1 provides a characterization of D-optimality which is a generalization of the equivalence result of Kiefer and Wolfowitz ('60). Subsequent sections of the chapter provide applications of this theorem to the various types of design restrictions already introduced.

Chapter IV considers the problem of c-optimality in the restricted design setting. A characterization of c-optimality is given by theorem 4.1.1 which generalizes a result of Kiefer and Wolfowitz ('59). The remainder of the chapter is devoted to applications of this theorem to the different types of design restrictions. In particular, section 4.2 includes the (implicit) solution to a problem of Hoel ('65).

CHAPTER II

ADMISSIBILITY FOR POLYNOMIAL REGRESSION

2.1 Preliminaries

Recall that the most general statement of the optimal design problem is to determine a design measure $\xi \in \Xi$ which "minimizes" $M^{-1}(\xi)$. The sense in which this minimization is undertaken will define an optimality criterion. It is often the case that a particular criterion corresponds to a real-valued function Φ defined on the set of non-negative definite matrices. Thus a "Φ-optimal" design ξ would be one which minimizes $\Phi(M^{-1}(\xi))$. Examples that will be treated in subsequent chapters are $\Phi(M^{-1}(\xi)) = |M^{-1}(\xi)|$ ("D-optimality") and $\Phi(M^{-1}(\xi)) = c'M^{-1}(\xi)c$ ("c-optimality"), where $c \in \mathbb{R}^{n+1}$.

Now different functions Φ will establish different orderings on the set of information matrices $\{M(\xi) | \xi \in \Xi\}$. Hence different criteria would be expected to yield different optimal designs. However, in many cases (including D and c-optimality) the function Φ is "monotone". That is, if $M^{-1}(\xi) \leq M^{-1}(\xi')$, then $\Phi(M^{-1}(\xi)) \leq \Phi(M^{-1}(\xi'))$. Here the inequality $A \leq B$ for non-negative definite matrices A and B should have the customary meaning that B - A is non-negative definite. Hence arises the "admissibility" problem: characterize those designs whose inverse information matrices are minimal with respect to the partial ordering "$\leq$". Now recall the elementary fact that positive definite matrices A and B satisfy $A \leq B$ if and only if $A^{-1} \geq B^{-1}$. Thus the equivalent statement of the admissibility problem is to characterize those designs whose information matrices are maximal with respect to the partial ordering "$\leq$".

Definition 2.1.1: A design ξ is admissible if and only if there does not exist another design ξ' such that $M(\xi') \geq M(\xi)$.

Once the admissible designs have been characterized, all others may be eliminated from consideration since they would be inferior for any monotone Φ.

Consider now the setting of univariate polynomial regression of degree n, with $\mathcal{X} = [a,b]$ and $f(x) = (1, x, \ldots, x^n)'$. Then the information matrix for a design ξ has elements $M_{ij}(\xi) = \int_a^b f_i(x) f_j(x) d\xi(x)$ $= \int_a^b x^{i+j} d\xi(x) = \mu_{i+j}$, where each $\mu_k = \int_a^b x^k d\xi(x)$ and $0 \leq i, j \leq n$. The fact that the information matrix for polynomial regression is such a Hankel matrix has enabled the following theorem to be established. The importance of this theorem is that it relates the admissibility problem for polynomial regression to a readily posed moment space problem.

Theorem 2.1.1: ξ is admissible for polynomial regression of degree n if and only if there exists no other design which shares the same values of $\mu_1, \ldots, \mu_{2n-1}$ but has a larger value of μ_{2n}.

Proof: A proof of this theorem may be found in Karlin and Studden ('66a).

The preceding theorem yields the conclusion that characterizing the admissible designs for polynomial regression of degree n is equivalent to characterizing those designs giving rise to moment points $(\mu_0, \mu_1,\ldots,\mu_{2n-1}, \mu_{2n})'$ which lie on the upper boundary of the moment space generated by all designs $\xi \in \Xi$.

The problem of characterizing the admissible unrestricted designs for polynomial regression has been treated by de la Garza ('54), Kiefer ('59), and Karlin and Studden ('66a). The characterization that they have developed is given by the following theorem.

Theorem 2.1.2: ξ is admissible for polynomial regression of degree n on [a,b] if and only if the support of ξ includes n-1 or fewer interior points.

This theorem will turn out to be a special case of theorem 2.2.6.

According to theorem 2.1.1, the restricted admissibility problem for polynomial regression of degree n is to characterize those designs $\xi \in \Xi$ whose value of μ_{2n} is at least as large as any other design which shares the same values of $\mu_0, \mu_1,\ldots,\mu_{2n-1}$. This problem is most readily solved in a more general setting. In order to lay out the setting, the following definition is necessary.

Definition 2.1.2: The continuous functions $u_0, u_1,\ldots,u_m$ are a Tchebycheff system on [a,b] if and only if the determinant

$$\begin{vmatrix} u_0(x_0) & u_0(x_1) & \cdots & u_0(x_m) \\ u_1(x_0) & u_1(x_1) & \cdots & u_1(x_m) \\ \vdots & \vdots & & \vdots \\ u_m(x_0) & u_m(x_1) & \cdots & u_m(x_m) \end{vmatrix} > 0 \qquad (2.1.1)$$

whenever $a \leq x_0 < x_1 < \ldots < x_m \leq b$.

At times it may be convenient to denote the determinant in (2.1.1) by either $\begin{vmatrix} u_0, & u_1, \ldots, u_m \\ x_0, & x_1, \ldots, x_m \end{vmatrix}$ or $|u(x_0), u(x_1), \ldots, u(x_m)|$, where the (column) vector $u(x) = [u_0(x), u_1(x), \ldots, u_m(x)]'$. The following lemma provides an important relationship between a Tchebycheff system $u_0, u_1, \ldots, u_m$ and "polynomials" $P(x) = v'u(x)$, where $v \in \mathbb{R}^{m+1}$.

Lemma 2.1.1: Every non-trivial ($v \neq 0$) polynomial $P(x) = v'u(x)$ has m or fewer zeroes if and only if either $u_0, u_1, \ldots, u_m$ or $-u_0, u_1, \ldots, u_m$ comprise a Tchebycheff system.

Proof: A proof of this lemma may be found in Karlin and Studden ('66a).

Note that the polynomials $u_0(x) = 1, u_1(x) = x, \ldots, u_m(x) = x^m$ form a Tchebycheff system on any closed interval. In fact, the Vandermonde determinant $|u(x_0), \ldots, u(x_m)| = \prod_{0 \leq i < j \leq m} (x_j - x_i) > 0$ whenever $a \leq x_0 < \ldots < x_m \leq b$.

For the remainder of this chapter it will be assumed that both $u_0, \ldots, u_m$ and $u_0, \ldots, u_m, u_{m+1}$ are Tchebycheff systems on $[a', b]$, where $a' < a$. The notation $\mu_k = \int_a^b u_k(x) d\sigma(x)$ for $k = 0, \ldots, m+1$ will be employed. In this setting, an extension of the admissibility problem is the following.

Problem 2.1: Characterize those measures σ such that μ_{m+1} attains the largest (or smallest) possible value among all measures sharing the same values of $\mu_0, \ldots, \mu_m$.

The admissibility problem corresponds to the case that $u_k(x) = x^k$ for $k = 0, \ldots, m = 2n-1$, $\mu_0 = 1$, and the maximum value of μ_{m+1} is desired.

In the following sections of this chapter, problem 2.1 will be treated for the cases that Ξ arises from the E-problem, the (ϕ,ψ)-problem, and the p-problem on [a,b]. The E-problem results obtained will correct some results of Krein and Nudelman ('77). The (ϕ,ψ)-problem results obtained will generalize the results of Karlin and Studden ('66a) and Krein and Nudelman ('77). The p-problem results obtained will be entirely new.

2.2 The E-Problem

Recall that E denotes an arbitrary compact set contained in $\mathcal{X} = [a,b]$. It may be assumed without loss of generality that $a, b \in E$. For purposes of notation, let Σ denote the set of all finite positive measures on E and $\mathcal{M}_{m+1} = \{\mu | \mu = \int_E u(x)d\sigma(x);\ \sigma \in \Sigma\}$. In order to solve problem 2.1, some preliminary results are needed. The following lemma establishes the existence of a strictly positive polynomial which will be needed for subsequent proofs.

Lemma 2.2.1: There exists a polynomial $P(x) = \pi' u(x) > 0$ for all $x \in [a,b]$.

Proof: Let $a' \leq x_1 < x_2 < \ldots < x_m < a$ and $P(x) = |u(x_1),\ldots,u(x_m),\ u(x)|$. Then the polynomial $P(x) = \pi' u(x) > 0$ for all $x \in (x_m, b] \supset [a,b]$ by definition 2.1.2.

The guaranteed existence of such a positive polynomial enables the following theorem to be proven.

Theorem 2.2.1: $\mathcal{M}_{m+1}$ is the closed convex cone generated by u(E).

Proof: The proof is essentially that of Karlin and Studden ('66a). It is included here for the sake of completeness.

The fact that $\mathcal{M}_{m+1}$ is a convex cone is obvious.

To show that $\mathcal{M}_{m+1}$ is closed, consider a sequence $\{\mu^{(k)}\} \subset \mathcal{M}_{m+1}$ such that $\mu^{(k)} \to \mu$ and $\mu^{(k)} = \int_E u(x)d\sigma_k(x)$ for $\sigma_k \in \Sigma$. According to lemma 2.2.1, there exists $P(x) = \pi'u(x) > 0$ for all $x \in [a,b] \supset E$. Hence $\pi'\mu^{(k)} = \int_E P(x)d\sigma_k(x) \geq \sigma_k(E) \min_E P(x)$ which implies that $\sigma_k(E) \leq \pi'\mu^{(k)}/\min_E P(x)$. Because $\mu^{(k)} \to \mu$, this inequality yields the existence of $\sigma_0 \in [0,\infty)$ such that $\sigma_k(E) \leq \sigma_0$ for all k. Hence there must exist a subsequence of $\{\sigma_k\}$ which converges weakly to $\sigma \in \Sigma$. This implies that $\mu = \int_E u(x)d\sigma(x)$ and hence that $\mathcal{M}_{m+1}$ is closed.

By Carathéodory's theorem, the convex cone generated by u(E) is

$$W = \{w \mid w = \sum_{i=1}^{m+2} \rho_i u(x_i);\ \rho_i \geq 0,\ x_i \in E\}.$$

This convex cone is trivially contained in $\mathcal{M}_{m+1}$. The containment could be strict only if there exists $\mu = \int_E u(x)d\sigma(x) \in \mathcal{M}_{m+1} - W$ and a hyperplane separating μ from W. That is, there exists $\pi \in \mathbb{R}^{m+1}$ such that $\pi'\mu < \pi'w$ for all $w \in W$. Now for each $x \in E$, $w = \sigma(E)u(x) \in W$ and hence $\pi'\mu < \sigma(E)\pi'u(x)$. Integrating both sides of this inequality with respect to σ yields $\sigma(E)\pi'\mu < \sigma(E)\pi'\mu$. This contradiction implies that no such $\mu \in \mathcal{M}_{m+1} - W$ could exist and thus the theorem is proved.

The relevant properties of $\mathcal{M}_{m+1}$ may be stated in terms of non-negative polynomials $P(x) = \pi'\mu(x)$, where $\pi \in \mathbb{R}^{m+1}$. The set which generates such polynomials will be denoted by

$\mathcal{P}_{m+1} = \{\pi \mid \pi'u(x) \geq 0 \text{ for all } x \in E\}$.

Theorem 2.2.2: i. $\mu \in \mathcal{M}_{m+1}$ if and only if $\pi'\mu \geq 0$ for all $\pi \in \mathcal{P}_{m+1}$.

ii. $\mu \in \partial \mathcal{M}_{m+1}$ if and only if there exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0$.

iii. $\mu \in \text{Interior}(\mathcal{M}_{m+1})$ if and only if $\pi'\mu > 0$ for all $\pi \in \mathcal{P}_{m+1} - \{0\}$.

Proof: i. For any closed convex cone $C \subset \mathbb{R}^{m+1}$, let $C^+ = \{w | w'y \geq 0 \text{ for all } y \in C\}$. Note first that C^+ is a closed convex cone. It will now be shown that $C^{++} = C$. For any $y \in C$, $w'y \geq 0$ for all $w \in C^+$. Thus $C \subset C^{++}$. Suppose that there exists $y_0 \in C^{++} - C$. Then there must exist a hyperplane which separates y_0 from C. That is, there exists $w \in \mathbb{R}^{m+1}$ such that $w'y_0 < w'y$ for all $y \in C$. Now it must be true that $w \in C^+$. Otherwise, there exist $\tilde{y} \in C$ and $\gamma > 0$ such that $w'\tilde{y} < 0$ and $w'(\gamma\tilde{y}) < w'y_0 < w'(\gamma\tilde{y})$. Then $w \in C^+$ implies that $0 \leq w'y_0 < w'(0) = 0$. This contradiction forces $C^{++} - C = \phi$ so that $C^{++} = C$.

Now consider any $\pi \in \mathcal{M}^+_{m+1}$. Then choosing $\mu = u(x)$ yields $0 \leq \pi'\mu = \pi'u(x)$ for any $x \in E$. Thus $\mathcal{M}^+_{m+1} \subset \mathcal{P}_{m+1}$. Consider next any $\pi \in \mathcal{P}_{m+1}$. Then $\pi'\mu = \int_E \pi'u(x)d\sigma(x) \geq 0$ for any $\mu = \int_E u(x)d\sigma(x) \in \mathcal{M}_{m+1}$. Thus $\mathcal{P}_{m+1} \subset \mathcal{M}^+_{m+1}$. Therefore, $\mathcal{M}^+_{m+1} = \mathcal{P}_{m+1}$ which implies that

$\mathcal{M}_{m+1} = \mathcal{M}^{++}_{m+1} = \mathcal{P}^+_{m+1} = \{\mu | \pi'\mu \geq 0 \text{ for all } \pi \in \mathcal{P}_{m+1}\}$.

ii. Assume first that there exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0 \leq \pi'\tilde{\mu}$ for all $\tilde{\mu} \in \mathcal{M}_{m+1}$. Thus there is a supporting hyperplane to $\mathcal{M}_{m+1}$ at μ which means that $\mu \in \partial \mathcal{M}_{m+1}$.

Assume next that $\mu \in \partial \mathcal{M}_{m+1}$. Then there must exist a supporting hyperplane to $\mathcal{M}_{m+1}$ at μ. That is, there must exist $\pi \neq 0$ such that $\pi'\mu \leq \pi \tilde{\mu}$ for all $\tilde{\mu} \in \mathcal{M}_{m+1}$. Now it must be true that $\pi \in \mathcal{P}_{m+1}$. Otherwise there exist $\mu^* \in \mathcal{M}_{m+1}$ and $\gamma > 0$ such that $\pi'\mu^* < 0$ and

$\pi'(\gamma\mu^*) < \pi'\mu \leq \pi'(\gamma\mu^*)$. Then $\pi \in \mathcal{P}_{m+1}$ implies that $0 \leq \pi'\mu \leq \pi'(0)$.
iii. The last part of the theorem is equivalent to part ii. Thus the proof of the theorem is now complete.

The following lemma covers the trivial case that $\mathcal{M}_{m+1}$ has no interior points.

Lemma 2.2.2: $\mathcal{M}_{m+1}$ has a non-empty interior if and only if E contains at least m+1 points.

Proof: $\mathcal{M}_{m+1}$ has no interior points if and only if it is contained in some hyperplane which passes through the origin. That is, Interior$(\mathcal{M}_{m+1}) = \phi$ if and only if there exists $v \neq 0$ such that $0 = v'\mu = \int_E v'u(x)d\sigma(x)$ for every $\mu = \int_E u(x)d\sigma(x) \in \mathcal{M}_{m+1}$. This relationship holds for every positive measure σ on E if and only if E is contained in the set of zeroes of $P(x) = v'u(x)$. According to lemma 2.1.1, this implies that E includes m or fewer points. Conversely, suppose that $E = \{x_1,\ldots,x_k\}$ with $k \leq m$. Now let

$$P(x) = \begin{vmatrix} u_0, & u_1,\ldots,u_k \\ x, & x_1,\ldots,x_k \end{vmatrix} = v'u(x)$$

and note that $v \neq 0$. Then $E \subset P^{-1}(\{0\})$ so that the proof is complete.

Theorem 2.2.2 serves to relate the properties of $\mathcal{M}_{m+1}$ to those of $\mathcal{P}_{m+1}$. The following notions are needed to develop the relevant properties of $\mathcal{P}_{m+1}$.

Definition 2.2.1: For any real-valued function g on E which is not identically zero, $S^+(g)$ will denote the number of sign changes of g(x) on E. The convention applied to zeroes will be that they are declared positive or negative so as to maximize this number of sign changes. In the case that $g(x) = 0$ on E, then $S^+(g)$ is defined to be the number of elements of E.

Note that $S^+(g)$ may be infinite.

<u>Definition 2.2.2</u>: The <u>index of a set</u> $A \subset E$ is $I_0(A) = S^+(1-\chi_A)$, where χ_A denotes the characteristic function of A. Equivalently, $I_0(A) = S^+(\chi_{A^c})$.

<u>Definition 2.2.3</u>: The <u>index of a measure</u> $\sigma \in \Sigma$ is $I(\sigma) = I_0(\text{Support }(\sigma))$.

The following will serve as an example of the index calculation.

<u>Example 2.2.1</u>: Let $E = [0,2] \cup \{3,4,5,6\}$ and let $A = \text{Support}(\sigma) = \{0,1,2,3,5,6\}$. Then for purposes of evaluating $S^+(1-\chi_A)$, the sign of $1-\chi_A(x)$ may be declared positive for $x \in \{2,6\}$ and negative for $x \in \{0,1,3,5\}$. Hence $I_0(A) = I(\sigma) = S^+(1-\chi_A) = 7$. Note that the points $x = 2$ and 3 in this example illustrate the possibility that the assignment of signs to the zeroes of $1-\chi_A$ need not be unique.

Now the essence of the index concept is that it provides an appropriate way to count the support points of a measure q. The following notions will make explicit the means of counting.

<u>Definition 2.2.4</u>: A <u>block</u> of $A \subset E$ is a maximal subset of A with the property that no two of its elements are separated by an element of $E - A$.

The role that the blocks of a set play in the calculation of its index is made apparent by the following lemma.

Lemma 2.2.3: If A consists of blocks $B_1,\ldots,B_p$ then $I_0(A) = I_0(B_1) +\ldots+ I_0(B_p)$.

Proof: According to definition 2.2.4, adjacent blocks of A are separated by one or more points of E - A. At these points the sign of $1-\chi_A$ is positive. Hence the number of sign changes of $1-\chi_A$ is the sum of the number of sign changes through each block. That is, $I_0(A) = I_0(B_1) +\ldots+ I_0(B_p)$.

According to this lemma, the index of a set is calculated block-wise. The following definition and lemma establish the means of calculating the index of an individual block.

Definition 2.2.5: Let B be a block of $A \subset E$.

i. If a or b $\in$ B, then B adjoins a or b (respectively).

ii. Otherwise, B is interior.

Lemma 2.2.4: If B is a block containing k elements, then

$$I_0(B) = \begin{cases} k & \text{B adjoins a or b} \\ 2[(k+1)/2] & \text{B is interior,} \end{cases}$$

where $[\cdot]$ denotes the greatest integer function.

Proof: Assume first that B adjoins b and its elements are denoted by $x_1 <\ldots< x_{k-1} < x_k = b$. Then $1-\chi_B(x) = 0$ for $x \in \{x_1,\ldots,x_k\}$. According to definition 2.2.1, the sign of $1-\chi_B(x_i)$ should be declared to be $(-1)^i$ for the purpose of calculating $S^+(1-\chi_B)$. Thus $I_0(B) = S^+(1-\chi_B) = k$.

Similarly, $I_0(B) = k$ if B adjoins a.

Assume now that B is an interior block and its elements are denoted by $a < x_1 < \ldots < x_k < b$. As before, the sign of $1-\chi_B(x_i)$ may be declared to be $(-1)^i$ for the purpose of calculating $S^+(1-\chi_B)$. (Note that if k is even, then $(-1)^{i+1}$ is equally valid.) Thus

$$I_0(B) = S^+(1-\chi_B) = \begin{cases} k+1 & k \text{ odd} \\ k & k \text{ even} \end{cases}$$

and the lemma is proved.

An illustration of these concepts of counting is provided by example 2.2.1. The set A specified by that example is comprised of a block $B_1 = \{0\}$ which adjoins 0, an odd interior block $B_2 = \{1\}$, an even interior block $B_3 = \{2,3\}$, and a block $B_4 = \{5,6\}$ which adjoins 6. Hence lemma 2.2.4 yields $I_0(B_1) = 1$, $I_0(B_2) = 1 + 1 = 2$, $I_0(B_3) = 2$, and $I_0(B_4) = 2$ so that lemma 2.2.3 confirms $I_0(A) = 7$.

The following properties will be needed for the proof of subsequent results.

Corollary 2.2.1: i. Let K denote the number of elements of $A \subset E$ and $\tilde{p}$ denote the number of interior blocks which are odd. Then $I_0(A) = K + \tilde{p}$.

ii. $I_0(A) < \infty$ if and only if A is finite.

iii. If $A_1 \subset A_2$, then $I_0(A_1) \leq I_0(A_2)$.

Proof: i. The proof of the first part is an immediate consequence of lemmas 2.2.3 and 2.2.4.

ii. According to part i., $K \leq I_0(A) \leq K + K = 2K$. Thus $I_0(A) < \infty$ if and only if $K < \infty$.

iii. $A_1 \subset A_2$ implies that $0 \leq 1-\chi_{A_2}(x) \leq 1-\chi_{A_1}(x) \leq 1$ for all $x \in E$. Then according to definition 2.2.1, it must be true that $S^+(1-\chi_{A_2}) \geq S^+(1-\chi_{A_1})$. That is, $I_0(A_2) \geq I_0(A_1)$.

The following corollary will be needed to prove theorem 2.1.2.

Corollary 2.2.2: If $E = [a,b]$, then $I_0(A)$ is the number of endpoints contained in A plus twice the number of interior points contained in A.

Proof: In this setting every interior point is an odd block of size one and index two. Thus the proof is an immediate consequence of lemmas 2.2.3 and 2.2.4.

The following theorem establishes the correspondence between the index of a set and the zeroes of a non-netative polynomial. It will be important to the proof of subsequent results.

Theorem 2.2.3: Let $A \subset E$. There exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $P(x) = \pi' u(x)$ vanishes precisely on A if and only if $I_0(A) \leq m$.

Proof: First note that A must be finite. If the indicated polynomial exists, A must contain m or fewer points according to lemma 2.1.1. Conversely, if $I_0(A) \leq m$ then A must contain m or fewer points according to corollary 2.2.1(i). Thus $A = \{x_1,\ldots,x_k\}$ where $k \leq m$ and $x_1 < \ldots < x_k$.

Assume first that $I_0(A) \leq m$. Let the points $t_1 < \ldots < t_r$ be obtained from A by preceding the left endpoint x_i of any odd interior block by $x_i - \varepsilon$, where $0 < \varepsilon < \min_i(x_i - x_{i-1})$. According to corollary

2.2.1(i), $r = I_0(A) \leq m$. If $r < m$, then select $m - r$ additional points from $[a', a)$. Let $s_1 < \ldots < s_m$ denote these points and the t_i's. Also let ℓ denote the number of elements in the block of A adjoining b and $C = (-1)^{\ell}$. (It may be that $\ell = 0$.)

Now consider the polynomial

$P_\varepsilon(x) = C|u(s_1),\ldots,u(s_m), u(x)| = v_\varepsilon' u(x)$. Note that

$P_\varepsilon^{-1}([0,\infty)) = E \cap([s_m, b] \cup [s_{m-2}, s_{m-1}] \cup [s_{m-4}, s_{m-3}] \cup \ldots)$ and

$P_\varepsilon^{-1}((-\infty,0]) = E \cap ([s_{m-1}, s_m] \cup [s_{m-3}, s_{m-2}] \cup \ldots)$ due to the alternating property of the determinant and definition 2.1.2. Thus P_ε is nearly the desired polynomial; its only shortcoming might be an unwanted "dip" to the left of the left endpoint of an odd interior block. To alleviate this possible drawback, consider the limit polynomial

$$P(x) = \lim_{\varepsilon \downarrow 0} v_\varepsilon' u(x)/\sqrt{v_\varepsilon' v_\varepsilon} = \lim_{\varepsilon \downarrow 0} w_\varepsilon' u(x) = \pi' u(x).$$

The fact that the w_ε lie on the unit sphere guarantees that a limit point exists. π may be taken to be any such limit point. The polynomial P vanishes precisely on A by the construction of the sequence $s_1,\ldots,s_m$, lemma 2.1.1, and the limiting process. The non-negativity of P is due to the value of C, the construction of the sequence $s_1,\ldots,s_m$, and the limiting process.

Assume now that $P(x) = \pi' u(x)$, that $\pi \in \mathcal{P}_{m+1} - \{0\}$, and that $P^{-1}(\{0\}) = A = \{x_1,\ldots,x_k\}$. As already noted, $k \leq m$. Hence corollary 2.2.1(i) implies that $I_0(A) \leq m$ unless A contains at least one odd interior block. Thus it may be supposed that $I_0(A) > m$ and A has at least one odd interior block. For each such block there exists an element of E - A which separates it from any other block to its left. For the right-hand-most odd interior block there exists an element of

E - A which separates it from any other block to its right. Let T denote the set comprised of all such "separating elements" and the elements of A. By corollary 2.2.1(i), the number of elements of T is $I_0(A) + 1 > m + 1$. Let $t_0 < \ldots < t_{m+1}$ denote the m + 2 smallest elements of T. Because $P(x) = \sum_{i=0}^{m} \pi_i U_i(x)$ it must be true that

$$0 = \begin{vmatrix} u_0, \ldots, u_m, & P \\ t_0, \ldots, t_m, & t_{m+1} \end{vmatrix} = \sum_{i=0}^{m+1} P(t_i)(-1)^{m+1-i} \Delta_i, \text{ where}$$

$$\Delta_i = \begin{vmatrix} u_0, & \ldots & , u_m \\ t_0, \ldots, t_{i-1}, & t_{i+1}, \ldots, & t_{m+1} \end{vmatrix} > 0$$

by definition 2.1.2. Now for all "separating elements" $t_i \in E - A$, $P(t_i) > 0$ holds. Also, the construction of T implies that i is even or odd for all of these separating elements (depending on whether there is or is not an odd block adjoining A). Hence all non-zero terms in the preceding summation have the same sign which contradicts the zero value of the summation. Thus $I_0(A) > m$ cannot hold and the proof of the theorem is complete.

<u>Definition 2.2.6</u>: $\sigma \in \Sigma$ <u>represents</u> $\mu \in \mathcal{M}_{m+1}$ if and only if $\mu = \int_E u(x)d\sigma(x)$.

The following important theorem provides two characterizations of the boundary of $\mathcal{M}_{m+1}$.

<u>Theorem 2.2.4</u>: i. $\mu = \int_E u(x)d\sigma(x) \in \partial\, \mathcal{M}_{m+1}$ if and only if $I(\sigma) \leq m$.

ii. If $\mu \in \partial\, \mathcal{M}_{m+1}$, then its representation is unique. The converse holds if E contains at least m + 2 points.

Proof: i. Assume first that $\mu = \int_E u(x)d\sigma(x) \in \partial \mathcal{M}_{m+1}$. Theorem 2.2.2(ii) implies that there exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $0 = \pi'\mu = \int_E \pi' u(x)d\sigma(x) = \int_E P(x)d\sigma(x) \geq 0$. Hence it must be true that $\text{Support}(\sigma) \subset P^{-1}(\{0\})$. Thus application of corollary 2.2.1(iii) and theorem 2.2.3 yields $I(\sigma) = I_0(\text{Support}(\sigma)) \leq I_0(P^{-1}(\{0\})) \leq m$.

Assume next that $\mu = \int_E u(x)d\sigma(x)$ and $I(\sigma) \leq m$. Theorem 2.2.3 implies that there exists a polynomial $P(x) = \pi' u(x)$ such that $\pi \in \mathcal{P}_{m+1} - \{0\}$ and $P^{-1}(\{0\}) = \text{Support}(\sigma)$. Hence $\pi'\mu = \int_E \pi' u(x)d\sigma(x) = 0$ and theorem 2.2.2 implies that $\mu \in \partial \mathcal{M}_{m+1}$.

ii. Assume now that $\mu = \int_E u(x)d\sigma(x) \in \partial \mathcal{M}_{m+1}$. Part i. implies that $I(\sigma) \leq m$. Hence theorem 2.2.3 implies that there exists a non-negative polynomial P such that $\text{Support}(\sigma) = P^{-1}(\{0\}) = \{x_1,\ldots,x_k\}$. Note that corollary 2.2.1(i) yields $k \leq I(\sigma) \leq m$. Now set $\sigma_i = \sigma(\{x_i\})$ for $i = 1,\ldots,k$ and consider the linear system $\mu = \int_E u(x)d\sigma(x) = \sum_{i=1}^{k} u(x_i)\sigma_i$. Note that definition 2.1.2 implies that the vectors $u(x_1),\ldots,u(x_k)$ are linearly independent. This guarantees a unique solution for $\sigma_1,\ldots,\sigma_k$. Thus the support and mass distribution of σ are uniquely determined.

Assume next that E contains at least m + 2 points and $\mu \in \text{Interior}(\mathcal{M}_{m+1})$. Recall now the earlier assumption that there is a function u_{m+1} such that $u_0,\ldots,u_m,\ u_{m+1}$ comprise a Tchebycheff system. According to lemma 2.2.2, $\mathcal{M}_{m+2}$ has a non-empty interior. Hence there must exist $\underline{\mu}_{m+1} < \bar{\mu}_{m+1}$ such that $\begin{bmatrix} \mu \\ \underline{\mu}_{m+1} \end{bmatrix}$ and $\begin{bmatrix} \mu \\ \bar{\mu}_{m+1} \end{bmatrix}$ belong to $\mathcal{M}_{m+2}$. If $\underline{\sigma}$ and $\bar{\sigma}$ denote measures representing these two points of $\mathcal{M}_{m+2}$, then both $\underline{\sigma}$ and $\bar{\sigma}$ represent $\mu \in \mathcal{M}_{m+1}$. Thus the proof is complete.

The following lemma clarifies part ii. of theorem 2.2.4 by covering the trivial case that every moment point has a unique representation.

Lemma 2.2.5: If $E = \{x_0, \ldots, x_k\}$ and $k \leq m$, then every moment point has a unique representation.

Proof: Let $\mu = \int_E u(x)d\sigma(x) \in \mathcal{M}_{m+1}$ and $\sigma_i = \sigma(\{x_i\})$ for $i = 0, \ldots, k$. Then $\sigma_0, \ldots, \sigma_k$ must solve the linear system $\mu = \sum_{i=0}^{k} u(x_i)\sigma_i$. As in the proof of theorem 2.2.4(ii), definition 2.1.2 implies that $\sigma_0, \ldots, \sigma_k$ are uniquely determined.

A comparison of lemmas 2.2.2 and 2.2.5 now sheds further light on part ii. of theorem 2.2.4. Specifically, they imply that if E consists of exactly m + 1 points, then the interior of $\mathcal{M}_{m+2}$ is non-empty but every moment point is uniquely represented. Henceforth it will be assumed that E contains at least m + 2 points.

Note that at this point problem 2.1 has been solved for two cases. The first case is the trivial case covered by lemma 2.2.5: if E consists of less than m + 2 points, then every moment point $\mu \in \mathcal{M}_{m+1}$ is uniquely represented. The second case is covered by theorem 2.2.4: if $I(\sigma) \leq m$, then $\mu \in \partial \mathcal{M}_{m+1}$ and is uniquely represented. In either of these cases, σ is unique so that μ_{m+1} is already determined by the values $\mu_0, \mu_1, \ldots, \mu_m$. The next task will be to solve problem 2.1 for the case that $\mu \in \text{Interior}(\mathcal{M}_{m+1})$. To that end the following notions will be needed.

Definition 2.2.7: A measure σ is principal if and only if $I(\sigma) = m + 1$.

Definition 2.2.8: A principal measure σ is upper (lower) if and only if the "sign" of $1 - \chi_{\text{Support}(\sigma)}$ is negative (positive) at $x = b$. Here "sign" is taken to mean the sign of the function for purposes of evaluating $S^{+}(\cdot)$. Since E is assumed to contain at least $m + 2$ points and $I(\sigma) = m + 1$, definition 2.2.1 implies that Support(σ) $\neq$ E. Hence there must exist $x_0 \in E$ such that $1 - \chi_{\text{Support}(\sigma)}(x_0) = 1$ so that the "sign" of this function at $x = b$ is uniquely determined.

The preceding is not the most readily applied version of the definition. It is presented for comparison with definition 2.3.8 of the next section. The following lemma provides an alternate version of the definition and will be subsequently employed.

Lemma 2.2.6: A principal measure is upper (lower) if and only if it has an odd (even) block adjoining b.

Proof: First let $x_0 = \sup(E - \text{Support}(\sigma))$ which exists according to the remarks accompanying definition 2.2.8. Next let $x_1 < \ldots < x_k$ denote the elements of the block adjoining b (if it exists). Then for purposes of evaluating $S^{+}(\cdot)$, the sign of $1 - \chi_{\text{Support}(\sigma)}(x_i)$ is $(-1)^i$ for $i = 1,\ldots,k$. This sign is negative at $x_k = b$ if and only if k is odd. Thus, according to definition 2.2.8, the lemma is proved.

The following corollary will be used to prove theorem 2.1.2.

Corollary 2.2.3: If $E = [a,b]$, then a principal measure is upper if and only if it concentrates mass at b.

Proof: The proof follows immediately from lemma 2.2.6.

It should be noted that the result of lemma 2.2.6 differs from Krein and Nudelman's ('77) definition of an upper principal representation. They call a principal representation upper if it merely has positive mass at b. This discrepancy turns out to invalidate some of their later results.

The following theorem completes the solution of problem 2.1.

<u>Theorem 2.2.5</u>: Let $\mu \in \text{Interior}(\mathcal{M}_{m+1})$. Among those measures σ representing μ, the maximum (minimum) value of μ_{m+1} is attained if and only if σ is upper (lower) principal.

<u>Proof</u>: Let σ be any measure representing μ and $P(x) = \pi' u(x)$ be a strictly positive polynomial which exists according to lemma 2.2.1. Then $\pi'\mu = \int_E P(x)d\sigma(x) \geq \sigma(E) \min_E P(x)$. Hence

$$|\mu_{m+1}| = |\int_E u_{m+1}(x)d\sigma(x)| \leq \sigma(E) \max_E |u_{m+1}(x)| \leq \frac{\pi'\mu}{\min_E P(x)} \max_E |u_{m+1}(x)|.$$

Then let $\underline{\mu}_{m+1} < \bar{\mu}_{m+1}$ be the smallest and largest values of μ_{m+1} attained by measures which represent $\mu \in \mathcal{M}_{m+1}$. Also, let $\underline{\sigma}$ and $\bar{\sigma}$ be measures which represent $\begin{bmatrix} \mu \\ \underline{\mu}_{m+1} \end{bmatrix}$ and $\begin{bmatrix} \mu \\ \bar{\mu}_{m+1} \end{bmatrix}$ (respectively). Now these two points lie on the lower and upper boundaries of $\mathcal{M}_{m+2}$ (respectively), so theorem 2.2.4 implies that $I(\underline{\sigma}) \leq m + 1$, that $I(\bar{\sigma}) \leq m + 1$, and that both measures are unique with respect to $\mathcal{M}_{m+2}$. Since $\mu \in \text{Interior}(\mathcal{M}_{m+1})$, the same theorem implies that $I(\underline{\sigma}) > m$ and $I(\bar{\sigma}) > m$. Hence $I(\underline{\sigma}) = I(\bar{\sigma}) = m + 1$ so that $\underline{\sigma}$ and $\bar{\sigma}$ are principal representations of μ.

It will now be shown that $\underline{\sigma}$ must be lower principal. First note that since $I(\underline{\sigma}) = m + 1$, theorem 2.2.3 assures the existence of a polynomial $P(x) = \sum_{i=0}^{m+1} \pi_i u_i(x)$ which is non-negative on E and which satisfies $P^{-1}(\{0\}) = \text{Support}(\underline{\sigma})$. Recall that the construction of P(x) yields

$$\pi_{m+1} = \lim_{\varepsilon \downarrow 0} C \begin{vmatrix} u_0, \ldots, & u_m \\ s_0, \ldots, & s_m \end{vmatrix} \Bigg/ \sqrt{\sum_{i=0}^{m+1} (v_i^{\varepsilon})^2} ,$$

where $C = (-1)^{\ell}$ and $\ell \geq 0$ is the size of the block adjoining b. Suppose now that ℓ were odd. This would imply that $\pi_{m+1} \leq 0$. Also, $\pi_{m+1} \neq 0$ or else theorem 2.2.3 would imply that $m \geq I(\underline{\sigma}) = m + 1$. Then consider

$$\sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1} \underline{\mu}_{m+1} = \int_E P(x) d\underline{\sigma}(x) = 0$$

$$\leq \int_E P(x) d\bar{\sigma}(x) = \sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1} \bar{\mu}_{m+1} .$$

Because $\pi_{m+1} < 0$, the preceding inequality implies that $\underline{\mu}_{m+1} \geq \bar{\mu}_{m+1}$. This contradiction must mean that ℓ is even and hence that $\underline{\sigma}$ is lower principal.

The proof that $\bar{\sigma}$ is upper principal is exactly the same except that ℓ is assumed to be even, $\pi_{m+1} > 0$, and hence $\bar{\mu}_{m+1} < \underline{\mu}_{m+1}$. This contradiction implies that $\bar{\sigma}$ must have an odd block adjoining b. Thus $\bar{\sigma}$ is upper principal and the proof is complete.

Note that in the course of proving this theorem, the existence and uniqueness of upper and lower principal representations have been established. Also, this theorem completes the solution of problem 2.1.

The following will serve as a counter-example to the results of Krein and Nudelman ('77).

Example 2.2.2: Let $m = 3$, let $u(x) = (1, x, x^2, x^3)'$, and let $E = [-2, -1] \cup [1, 2] \cup \{-3, 3\}$. Consider the measure $\sigma = \frac{1}{4}(\delta_{-3} + \delta_{-2} + \delta_2 + \delta_3)$ which yields $\mu = (1, 0, 13/2, 0)'$ and $\mu_4 = 97/2$. Now $I(\sigma) = 4$ and so σ must be either the upper or lower principal representation of $\mu \in \mathcal{M}_4$. According to Krein and Nudelman ('77), σ must be upper principal since it concentrates positive mass at the point 3. Actually, σ must be lower principal since it has an even block adjoining 3. Consider now the measure $\bar{\sigma} = \frac{11}{32}(\delta_{-3} + \delta_3) + \frac{5}{32}(\delta_{-1} + \delta_1)$ which yields $\mu = (1, 0, 13/2, 0)'$ and $\bar{\mu}_4 = 56$. Since $I(\bar{\sigma}) = 4$, σ is a principal representation for μ. In fact, $\bar{\sigma}$ must be the upper principal representation since it has an odd block adjoining 3 and since $\bar{\mu}_4 > \mu_4$. This example points out the ambiguity of Krein and Nudelman's ('77) definition of an upper principal representation. In fact, their definition would imply that both σ and $\bar{\sigma}$ are upper principal.

At this point, problem 2.1 is solved and the results necessary to characterize the admissible designs for polynomial regression of degree n on E have been obtained.

Theorem 2.2.6: i. If E consists of 2n or fewer points, then all designs are admissible for polynomial regression of degree n on E.
ii. Otherwise, a design ξ is admissible if and only if either
a. $I(\xi) < 2n$ or b. $I(\xi) = 2n$ and ξ is upper.

Proof: The proof of this theorem is an application of moment space results now established to the case that $u_i(x) = x^i$ for $i = 0, 1, \ldots, m = 2n-1$ and $\mu_0 = 1$.

i. If E contains m + 1 = 2n or fewer points, then lemma 2.2.5 implies that every moment point $(\mu_0, \mu_1, \ldots, \mu_{2n-1})'$ has a unique representation. Hence theorem 2.1.1 implies that every design is admissible.

ii. If E contains at least m + 2 = 2n + 1 points, then the class of designs such that $I(\xi) < m + 1 = 2n$ coincides with the class of designs which uniquely represent $\mu = \int_E u(x)d\xi(x) \in \partial\mathcal{M}_{2n}$. Hence theorem 2.1.1 implies that each such design is admissible.

Finally, theorems 2.1.1 and 2.2.5 imply that the remaining admissible designs are those which are upper principal representations of their moment points $\mu \in \mathcal{M}_{2n}$. That is, $I(\xi) = m + 1 = 2n$ and ξ is upper. Thus the theorem is proved.

At this point it may be noted that theorem 2.1.2 follows immediately from theorem 2.2.6 and corollary 2.2.2.

Note also that the design $\bar{\sigma}$ given in example 2.2.2 is admissible for polynomial regression of degree $n \geq 2$ on E.

As an added consequence of the properties of $\mathcal{M}_{2n}$ and $\mathcal{M}_{2n+1}$, each of the admissible designs given by theorem 2.2.6 is unique.

Note finally that admissibility for the E-problem depends only on the support of a design. That will not turn out to be the case for the (ϕ,ψ) and p-problems.

Now that problem 2.1 has been solved, a secondary problem may be considered. Given a particular inadmissible design ξ, determine the design $\bar{\xi}$ which best improves ξ in the sense of admissibility. With $u(x) = (1, x, \ldots, x^{2n-1})'$, a more precise statement of this problem is the following.

Problem 2.2: Let $\mu = \int_E u(x)d\xi(x)$, let $\mu_{2n} = \int_E x^{2n}d\xi(x)$, and let $\bar{\mu}_{2n}$ be the largest value of the $2n^{th}$ moment attained by all measures in Ξ which represent μ. If $\mu_{2n} < \bar{\mu}_{2n}$, determine the design $\bar{\xi} \in \Xi$ which represents μ and attains $\bar{\mu}_{2n}$.

Note that theorem 2.2.6 implies that either a. $I(\xi) > 2n$ or b. $I(\xi) = 2n$ and ξ is the lower principal representation for $\mu \in \text{Interior}(\mathcal{M}_{2n})$.

In the case that $E = [0, 1]$, the following solution to problem 2.2 may be found in Karlin and Studden ('66a). The design $\bar{\eta}_0$ which solves problem 2.2 has $\text{Support}(\bar{\eta}_0) = \{0, 1, t_1, \ldots, t_{n-1}\}$, where $t_1, \ldots, t_{n-1}$ are the roots of the polynomial

$$P(t) = \begin{vmatrix} \mu_1-\mu_2 & \mu_2-\mu_3 & \cdots & \mu_{n-1}-\mu_n & 1 \\ \mu_2-\mu_3 & \mu_3-\mu_4 & \cdots & \mu_n-\mu_{n+1} & t \\ \vdots & \vdots & & \vdots & \vdots \\ \mu_n-\mu_{n+1} & \mu_{n+1}-\mu_{n+2} & \cdots & \mu_{2n-2}-\mu_{2n-1} & t^{n-1} \end{vmatrix}.$$

The distribution of mass among the support points is readily determined by solving the linear system $\mu = \int_0^1 u(t)d\bar{\eta}_0(t)$.

In the case that $E = [a,b]$, a design ξ under consideration may first be transformed to a design η on $[0,1]$ according to $t = (x-a)/(b-a)$. Once the transformed solution $\bar{\eta}_0$ has been obtained, it may be transformed back to the solution $\bar{\xi}_0$ on $E = [a,b]$. Then $\text{Support}(\bar{\xi}_0) = \{a, b, x_1, \ldots, x_{n-1}\}$, where $x_i = a + (b-a)t_i$ for $i = 1, \ldots, n-1$.

Now if $E \neq [a,b]$, it may nonetheless be true that $\text{Support}(\bar{\xi}_0) \subset E$. In this case, $\bar{\xi}_0 \in \Xi$ and thus provides the solution to problem 2.2.

Otherwise $x_i \notin E$ for $i = 1,\dots,s$ and $x_i \in E$ for $i = s+1,\dots,n-1$, where $1 \leq s \leq n-1$. That is, exactly s of $\bar{\xi}_0$'s interior support points do not fall in E. Now for each $i = 1,\dots,s$, let $\alpha_i = \max(E \cap [a,x_i))$ and $\beta_i = \min(E \cap (x_i,b])$. It is conjectured that $\bigcup_{i=1}^{s} \{\alpha_i,\beta_i\} \subset \text{Support}(\bar{\xi})$. If true, the problem reduces to determining the remainder of Support($\bar{\xi}$). For $s = n - 1$, no further analysis is needed. For $s = n - 2$, a method for determining the rest of Support($\bar{\xi}$) will be indicated. For $s < n - 2$, no method is suggested. This approach will be illustrated by examples with n = 1, 2, 3, and 4.

It might be noted that problem 2.2 could have been posed in more generality corresponding to problem 2.1. However, such an extension is beyond the scope of the present study.

In the case of linear regression, Support($\bar{\xi}_0$) = $\{a,b\} \subset E$. Thus $\bar{\xi}_0 \in \Xi$ and the case n = 1 is solved in general.

In the case of quadratic regression, Support($\bar{\xi}_0$) = $\{a, b, x_1\}$. If $x_1 \in E$, then $\bar{\xi}_0 \in \Xi$ and problem 2.2 is solved. Otherwise, the following lemma provides the general solution.

<u>Lemma 2.2.7</u>: If n = 2, Support($\bar{\xi}_0$) = $\{a, b, x_1\}$, and $x_1 \notin E$, then Support($\bar{\xi}$) = $\{a, b, \alpha_1, \beta_1\}$.

<u>Proof</u>: First note that if Support($\bar{\xi}$) included more than four points, then corollary 2.2.1(i) would imply that $I(\bar{\xi}) > 4$ and contradict theorem 2.2.6. If Support($\bar{\xi}$) contained fewer than four points, then the support points must be a, b, and x_1 because $\bar{\xi}_0$ is unique for the unrestricted problem. But $x_1 \notin E$ by assumption, so that Support($\bar{\xi}$) cannot contain fewer than four points. Now since $\bar{\xi}$ must have a odd

block adjoining a and one adjoining b, it must be true that $\text{Support}(\bar{\xi}) = \{a, b, y_1, y_2\}$. Without loss of generality, $y_1 < y_2$.

Now it will be shown to be impossible that $y_2 < x_1$ or $y_1 > x_1$. Consider the first of these possibilities, ie. $a < y_1 < y_2 < x_1 < b$. To solve for the mass distribution of $\bar{\xi}$, consider the linear system

$$\rho_0 u(a) + \rho_1 u(y_1) + \rho_2 u(y_2) + \rho_3 u(b) = \mu$$

$$= \bar{\xi}_{00} u(a) + \bar{\xi}_{01} u(x_1) + \bar{\xi}_{02} u(b),$$

where all masses are positive. Now

$$\rho_1 = \frac{|u(a),\ \mu,\ u(y_2),\ u(b)|}{|u(a),\ u(y_1),\ u(y_2),\ u(b)|}$$

$$= \bar{\xi}_{01} \frac{|u(a),\ u(x_1),\ u(y_2),\ u(b)|}{|u(a),\ u(y_1),\ u(y_2),\ u(b)|} < 0$$

since $a < y_1 < y_2 < x_1 < b$. This contradiction disallows $y_2 < x_1$. Similarly, $y_1 > x_1$ is not possible. Thus $a < y_1 < x_1 < y_2 < b$ must hold.

Furthermore, there is no case in which $(y_1, y_2) \cap E$ can be non-empty and $\bar{\xi}$ still be upper principal. Thus it must be true that $y_1 = \alpha_1$, $y_2 = \beta_1$, and the proof is complete.

An example for $n = 2$ has been provided by example 2.2.2.

In the case of cubic regression, $\text{Support}(\bar{\xi}_0) = \{a, b, x_1, x_2\}$. If $\{x_1, x_2\} \subset E$, then $\bar{\xi}_0 \in \Xi$ and problem 2.2 is solved. The following example will illustrate a method for determining $\bar{\xi}$ if x_1 or $x_2 \notin E$.

<u>Example 2.2.3</u>: Let $n = 3$, let $E = [-2, -1] \cup [1, 4]$, and let $\xi = \frac{1}{12}\delta_{-1} + \frac{1}{12}\delta_1 + \frac{1}{2}\delta_2 + \frac{1}{3}\delta_3$. The transformed version of the example

has $E = [0, \frac{1}{6}] \cup [\frac{1}{2}, 1]$, $\eta = \frac{1}{12}\delta_{1/6} + \frac{1}{12}\delta_{1/2} + \frac{1}{2}\delta_{2/3} + \frac{1}{3}\delta_{5/6}$, and $\tilde{\mu} \approx (1, 2/3, .4769, .3519, .2648, .2024)'$. Also,

$$P(t) = \begin{vmatrix} \tilde{\mu}_1-\tilde{\mu}_2 & \tilde{\mu}_2-\tilde{\mu}_3 & 1 \\ \tilde{\mu}_2-\tilde{\mu}_3 & \tilde{\mu}_3-\tilde{\mu}_4 & t \\ \tilde{\mu}_3-\tilde{\mu}_4 & \tilde{\mu}_4-\tilde{\mu}_5 & t^2 \end{vmatrix} \approx \begin{vmatrix} .1898 & .1250 & 1 \\ .1250 & .0871 & t \\ .0871 & .0624 & t^2 \end{vmatrix} .$$

The roots of this polynomial are approximately $t_1 = .31$ and $t_2 = .75$. The transformed support points are -2, 4, $x_1 \approx -.14$, and $x_2 \approx 2.5$. Note that $x_1 \notin E$ so that $\bar{\xi}_0 \notin \Xi$.

Now the form of $\bar{\xi}_0$ suggests that Support$(\bar{\xi}) = \{-2, 4, -1, +1, y\}$, where $1 < y < 4$. If this is true, then $\bar{\xi}$ will be upper principal and the only remaining requirement is that the linear system

$$\rho_1 u(-2) + \rho_2 u(-1) + \rho_3 u(1) + \rho_4 u(4) + \rho_5 u(y) = \mu \qquad (2.2.1)$$

must be satisfied. Hence it must be true that
$0 = |\mu, u(-2), u(-1), u(1), u(4), u(y)| = P(y)$. That is, y must be a root of the polynomial P. With
$\mu = \int_E u(x)d\xi(x) = (1, 2, 31/6, 13, 211/6, 97)'$, the roots of this polynomial are -2, -1, 1, 4, and the desired value $y = 48/19 \approx 2.5263$. Then the solution to (2.2.1) is approximately $\rho_1 = .0040$, $\rho_2 = .0682$, $\rho_3 = .1984$, $\rho_4 = .0238$, and $\rho_5 = .7056$. Thus $\bar{\xi} \in \Xi$ is the admissible improvement of ξ; in fact, $\mu_6 \approx 275.17 < 281.44 \approx \bar{\mu}_6$.

Now consider the same setting except that
$E = [-2, -1] \cup [1, 2] \cup [\sqrt{8}, 4]$. With this modification, $y \notin E$ which suggests that now Support$(\bar{\xi}) = \{-2, 4, -1, 1, 2, \sqrt{8}\}$. If this is true, then $\bar{\xi}$ will be upper principal and the remaining requirement is that the linear system

$\rho_1 u(-2) + \rho_2 u(-1) + \rho_3 u(1) + \rho_4 u(4) + \rho_5 u(\sqrt{8}) + \rho_6 u(2) = \mu$ must be satisfied. The solution is approximately $\rho_1 = .0012$, $\rho_2 = .0786$, $\rho_3 = .1177$, $\rho_4 = .0108$, $\rho_5 = .4065$, and $\rho_6 = .3852$. Thus $\bar{\xi} \in \Xi$ is the admissible improvement of ξ; in fact, $\mu_6 \approx 275.17 < 277.29 \approx \bar{\mu}_6$.

It is conjectured that the methods illustrated in example 2.2.3 will be sufficient to solve problem 2.2 in any setting involving cubic regression.

In the case of quartic regression, Support$(\bar{\xi}_0) = \{a, b, x_1, x_2, x_3\}$. If $\{x_1, x_2, x_3\} \subset E$, then $\bar{\xi}_0 \in \Xi$ and problem 2.2 is solved. If none of the points x_1, x_2, x_3 belong to E, it is conjectured that Support$(\bar{\xi}) = \{a, b, \alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3\}$ and the mass distribution is obtained by solving the linear system $\mu = \int_E u(x)d\bar{\xi}(x)$. If exactly one of the points x_1, x_2, x_3 belongs to E, it is proposed that the method illustrated in example 2.2.3 be employed. The difficult case is that in which exactly two of the points x_1, x_2, x_3 belong to E. The following example will illustrate such a situation.

<u>Example 2.2.4</u>: Let $n = 4$, let $E = [-4, -1] \cup [1, 4]$, and let $\xi = \frac{1}{4}(\delta_{-3} + \delta_{-2} + \delta_2 + \delta_3)$. Then $\mu = (1, 0, 13/2, 0, 97/2, 0, 793/2, 0)'$ and it may be determined that $x_1 = -x_3 \approx 2.61$ and $x_2 = 0$. Here $x_2 \notin E$ so that $\bar{\xi}_0 \notin \Xi$.

Now the symmetry of ξ and the form of $\bar{\xi}_0$ suggest that Support$(\bar{\xi}) = \{-4, 4, -1, +1, -y, y\}$, where $1 < y < 4$, and that $\bar{\xi}$ have a symmetrical mass distribution. If this is true, $\bar{\xi}$ will be upper principal and the only remaining requirement is that the even moments of $\bar{\xi}$ match those of ξ. To implement this requirement, let $\rho = \bar{\xi}(\{y, -y\})$ and $\tau = \bar{\xi}(\{4, -4\})$. Then ρ, τ, and y must solve the system

$$(1-\rho-\tau) + \rho y^2 + 16\tau = \mu_2 = 13/2,$$
$$(1-\rho-\tau) + \ y^4 + 256\tau = \mu_4 = 97/2,$$
$$(1-\rho-\tau) + \rho y^6 + 4096\tau = \mu_6 = 793/2.$$

The solution is approximately $y = 2.6540$, $\rho = .8498$, and $\tau = .0243$ so that $\bar{\xi}(\{1, -1\}) \approx .1259$. Thus $\bar{\xi} \in \Xi$ is the admissible improvement of ξ; in fact, $\mu_8 = 3408.5 < 3684.5 \approx \bar{\mu}_8$.

What made $\bar{\xi}$ so easy to obtain here was the symmetry of ξ and E. To illustrate the difficulties otherwise, consider the same setting except that $E = [-4, -2] \cup [1, 4]$. With this modification, it is proposed that Support$(\bar{\xi}) = \{-4, 4, -2, 1, y_1, y_2\}$, where $-4 < y_1 < -2$ and $1 < y_2 < 4$. The open question is how to determine y_1 and y_2 in such a situation.

The preceding examples should serve to illustrate some methods that may be used to solve problem 2.2 for general n and some difficulties that arise for $n \geq 4$.

2.3 The (φ,ψ)-Problem

Recall that φ and ψ are finite measures on $\mathcal{X} = [a,b]$ such that $d\varphi \leq d\psi$. Recall also that, according to remark 1.3.1, it suffices to consider the $(0,\nu)$-problem on $\mathcal{X} = [a,b]$. It may be assumed without loss of generality that $a, b \in$ Support(ν). For purposes of notation let $E =$ Support(ν), let $\Sigma = \{\sigma | d\sigma \leq d\nu\}$, and let $\mathcal{M}_{m+1} = \{\mu | \mu = \int_E u(x)d\sigma(x);\ \sigma \in \Sigma\}$. In order to solve problem 2.1, some preliminary results are needed. The following lemma is analogous to theorem 2.2.1. The moment space $\mathcal{M}_{m+1}$ of this section may be viewed as a truncation of the convex cone $\mathcal{M}_{m+1}$ of section 2.2.

Lemma 2.3.1: $\mathcal{M}_{m+1}$ is convex, bounded, and closed.

Proof: The convexity of $\mathcal{M}_{m+1}$ is obvious.

To show that $\mathcal{M}_{m+1}$ is bounded, note that the Euclidean norm $||\mu|| \leq \sigma(E) \max_{a \leq x \leq b} ||u(x)|| \leq \nu(E) \max_{a \leq x \leq b} ||u(x)||$ for all $\mu \in \mathcal{M}_{m+1}$.

To show that $\mathcal{M}_{m+1}$ is closed, consider a sequence $\{\mu^{(k)}\} \subset \mathcal{M}_{m+1}$ such that $\mu^{(k)} \to \mu$ and $\mu^{(k)} = \int_E u(x) d\sigma_k(x)$ for $\sigma_k \in \Sigma$. Now each $\sigma_k(E) \leq \nu(E)$. Hence there must exist a subsequence of $\{\sigma_k\}$ which converges weakly to $\sigma \in \Sigma$. This implies that $\mu = \int_E u(x) d\sigma(x)$ and hence that $\mathcal{M}_{m+1}$ is closed, completing the proof.

The following lemma dispatches the trivial case that Interior$(\mathcal{M}_{m+1}) = \phi$ so that theorem 2.3.1(iii) may be proved. Note that this lemma corresponds exactly to lemma 2.2.2.

Lemma 2.3.2: $\mathcal{M}_{m+1}$ has a non-empty interior if and only if E contains at least m+1 points.

Proof: First note that $0 \in \Sigma$ implies that $0 \in \mathcal{M}_{m+1}$. At this point, the proof is that of lemma 2.2.2 verbatim.

The relevant important properties of $\mathcal{M}_{m+1}$ may be stated in terms of polynomials $P(x) = v'u(x)$, where $v \in \mathbb{R}^{m+1}$. For purposes of notation, let

$$P_-(x) = \begin{cases} -P(x) & P(x) \leq 0 \\ 0 & P(x) \geq 0 \end{cases}$$

and $P_+(x) = P(x) + P_-(x)$ denote the negative and positive parts of $P(x)$.

The following theorem relates such polynomials to the properties of $\mathcal{M}_{m+1}$. Its role is analogous to that of theorem 2.2.2.

<u>Theorem 2.3.1</u>: i. $\mu \in \mathcal{M}_{m+1}$ if and only if

$$v'\mu \geq -\int_E P_-(x)d\nu(x) \tag{2.3.1}$$

for all polynomials $P(x) = v'u(x)$.

ii. $\mu \in \partial\, \mathcal{M}_{m+1}$ if and only if $\mu \in \mathcal{M}_{m+1}$ and there exists $v \neq 0$ such that $v'\mu = -\int_E P_-(x)d\nu(x)$.

iii. If $\mu \in \partial\, \mathcal{M}_{m+1}$, then μ has a unique representation. If E contains at least m+2 points, the converse also holds.

<u>Proof</u>: i. Assume first that $\mu = \int_E u(x)d\sigma(x) \in \mathcal{M}_{m+1}$. Then

$v'\mu = \int_E P(x)d\sigma(x) = \int_E P_+(x)d\sigma(x) - \int_E P_-(x)d\sigma(x) \geq 0 - \int_E P_-(x)d\nu(x)$

for all polynomials $P(x) = v'u(x)$.

Assume next that $\mu \notin \mathcal{M}_{m+1}$. Then there must exist a separating hyperplane such that $v'\mu < v'w$ for all $w \in \mathcal{M}_{m+1}$, where $v \neq 0$. Now let $P(x) = v'u(x)$, let σ be some measure such that

$$\frac{d\sigma}{d\nu}(x) = \begin{cases} 1 & P(x) < 0 \\ 0 & P(x) > 0, \end{cases}$$

and let $\tilde{w} = \int_E u(x)d\sigma(x)$. Then

$$\begin{aligned} v'\mu < v'\tilde{w} &= \int_E P(x)d\sigma(x) \\ &= \int_E P_+(x)\frac{d\sigma}{d\nu}(x)d\nu(x) - \int_E P_-(x)\frac{d\sigma}{d\nu}(x)d\nu(x) \\ &= -\int_E P_-(x)d\nu(x) \end{aligned}$$

so that (2.3.1) is not satisfied.

ii. Assume now that $\mu \in \partial \mathcal{M}_{m+1}$. Because $\mathcal{M}_{m+1}$ is convex, there must exist a supporting hyperplane to $\mathcal{M}_{m+1}$ at μ such that $v'\mu \leq v'w$ for all $w \in \mathcal{M}_{m+1}$, where $v \neq 0$. If $P(x)$, σ, and $\tilde{w}$ are defined as in part i., then $v'\mu \leq v'\tilde{w} = -\int_E P_-(x)d\ (x)$. Thus equality must hold in (2.3.1).

Assume now that $\mu \in \text{Interior}(\mathcal{M}_{m+1})$. Then for any $v \neq 0$, there must exist $w \in \mathcal{M}_{m+1}$ such that $v'\mu > v'w \geq -\int_E P_-(x)d\nu(x)$, where $P(x) = v'u(x)$. That is , equality does not hold in (2.3.1) for any $v \neq 0$.

iii. Assume now that $\mu \in \partial \mathcal{M}_{m+1}$. According to part ii., there exists a non-trivial polynomial $P(x) = v'u(x)$ such that $v'\mu = -\int_E P_-(x)d\nu(x)$ for any measure σ which represents μ. Hence (2.3.1) implies that σ must satisfy

$$\frac{d\sigma}{d\nu}(x) = \begin{cases} 1 & P(x) < 0 \\ 0 & P(x) > 0. \end{cases}$$

Thus σ is uniquely determined except possibly for its distribution of mass among the zeroes of $P(x)$. If $x_1,\ldots,x_k$ denote these zeroes, then lemma 2.1.1 implies that $k \leq m$. Now let $\sigma_i = \sigma(\{x_i\})$ for $i = 1,\ldots,k$, let $A = E \cap P^{-1}((-\infty,0))$, and note that $\mu = \int_A u(x)d\nu(x) + \sum_{i=1}^{k} u(x_i)\sigma_i$. Definition 2.1.2 implies that $u(x_1),\ldots,u(x_k)$ are independent and hence that the masses $\sigma_1,\ldots,\sigma_k$ must be unique.

The proof of the converse of part iii. follows the proof of the converse of theorem 2.2.4(ii) verbatim except that lemma 2.3.2 replaces lemma 2.2.2. The proof of the theorem is now complete.

The following lemma clarifies part iii. of theorem 2.3.1 by covering the trivial case that every moment point has a unique representation. Its statement and proof are essentially the same as lemma 2.2.5.

Lemma 2.3.3: If $E = \{x_0, \ldots, x_k\}$ and $k \leq m$, then every moment point has a unique representation.

Proof: The proof follows the proof of lemma 2.2.5 verbatim except that theorem 2.3.1(iii) replaces theorem 2.2.4(ii).

A comparison of lemmas 2.3.2 and 2.3.3 sheds further light on part iii. of theorem 2.3.1. Specifically, they imply that if E consists of exactly m+1 points, then the interior of $\mathcal{M}_{m+2}$ is non-empty but every moment point is uniquely represented. Henceforth it will be assumed that E contains at least m+2 points.

Thus far, theorem 2.3.1(ii) and (iii) provide characterizations of the boundary of $\mathcal{M}_{m+1}$. However, implementation of either may be less than straightforward. Specifically, part ii.'s criterion requires the construction of a certain polynomial. Implementation of part iii. would require that uniqueness be proved. The following notions will be used to provide a more applicable characterization of those measures representing moment points on the boundary of $\mathcal{M}_{m+1}$. The characterization obtained will lead to a solution of problem 2.1 in the present setting.

Definition 2.3.1: Let $q(x;\sigma) = \chi_{\text{Support}(\nu-\sigma)}(x) - \chi_{\text{Support}(\sigma)}(x)$ for $\sigma \in \Sigma$. Then the index of σ is $I(\sigma) = S^+(q)$, where the operation $S^+(\cdot)$ is performed as required by definition 2.2.1. That is, the sign changes of q are calculated relative to $E = \text{Support}(\nu)$.

Note the correspondence between this definition and definitions 2.2.2 and 2.2.3. The following will serve as an example of the index calculation.

Example 2.3.1: Let $[a,b] = [0,10]$ and

$$d\nu(x) = \begin{cases} 0 & x \in (3,4) \cup (5,6) \cup (7,8) \cup (9,10) \\ dx & x \in (0,1) \cup (1,2) \cup (2,3) \cup (4,5) \cup (6,7) \cup (8,9) \\ 1/5 & x \in \{0, 1, 2, 3, \ldots, 10\}. \end{cases}$$

That is, ν has eleven point masses superimposed on an absolutely continuous measure which equals either zero or Lebesque measure. Also, $E = \text{Support}(\nu) = [0,3] \cup [4,5] \cup [6,7] \cup [8,9] \cup \{10\}$. Now consider the measure defined on E by

$$d\sigma(x) = \begin{cases} 0 & x \in [0,1) \cup (4,5) \cup (6,7] \\ dx & x \in (1,2) \cup (2,3) \cup (8,9) \\ 1/5 & x \in \{8,9\} \\ 1/10 & x \in \{1, 2, 3, 4, 5, 6, 10\}. \end{cases}$$

Then $\text{Support}(\nu-\sigma) = [0,1] \cup \{2,3\} \cup [4,5] \cup [6,7] \cup \{10\}$ and $\text{Support}(\sigma) = [1,3] \cup \{4,5,6\} \cup [8,9] \cup \{10\}$. Hence

$$q(x;\sigma) = \begin{cases} -1 & x \in (1,2) \cup (2,3) \cup [8,9] \\ 0 & x \in \{1, 2, 3, 4, 5, 6, 10\} \\ +1 & x \in [0,1) \cup (4,5) \cup (6,7) \end{cases}$$

for $x \in E$. Thus $I(\sigma) = S^+(q) = 10$.

The essence of the index concept is that it provides an appropriate way to count the number of times that a measure alternates between 0 and ν. This "counting" is based on the sign of $q(x;\sigma)$ for $x \in E = \text{Support}(\nu)$. The following notions will make explicit the means of counting. They will be illustrated following corollary 2.3.1. These concepts will also be very much analogous to those introduced in section 2.2.

First assume that $A = q^{-1}(\{0\}) \cap E$ is finite and $S^+(q) < \infty$.

Definition 2.3.2: Let B be a block of A (as in definition 2.2.4) and $x_1 < \ldots < x_k$ denote its elements.

i. If $x_0 = \max([a,x_1) \cap E)$ exists, then the pre-sign of the block B is

$$\mathcal{S}(B-) = \begin{cases} \text{sgn}[q(x_0)] & x_0 < x_1 \\ \text{sgn}[q(x_1-)] & x_0 = x_1. \end{cases}$$

(Here $\text{sgn}[y] = y/|y|$ for $y \neq 0$ and $\text{sgn}[0] = 0$.)

ii. If $x_{k+1} = \min((x_k,b] \cap E)$ exists, then the post-sign of B is

$$\mathcal{S}(B+) = \begin{cases} \text{sgn}[q(x_{k+1})] & x_{k+1} > x_k \\ \text{sgn}[q(x_k+)] & x_{k+1} = x_k. \end{cases}$$

Definition 2.3.3: An interior block B is nodal (non-nodal) if and only if $\mathcal{S}(B-)$ and $\mathcal{S}(B+)$ are equal (unequal).

The relationship between these concepts and those of section 2.2 may be noted. An interior block for the E-problem corresponds to a non-nodal block with $\mathcal{S}(B-) = \mathcal{S}(B+) = 1$.

To consider the sign changes of $q(x)$ on E, note that the following definition describes the only way $q(x)$ can change sign without attaining an intervening zero.

Definition 2.3.4: If there exist α and β such that $[\alpha,\beta] \cap E = \{\alpha,\beta\}$, $q(\alpha) \neq 0$, $q(\beta) \neq 0$, and $\text{sgn}[q(\alpha)] \neq \text{sgn}[q(\beta)]$, then then $G = (\alpha,\beta)$ is a gap of A.

Note that a gap might be regarded as a nodal interior block of size 0.

According to lemma 2.3.3, $S^+(q)$ will be obtained by adding the indices of A's blocks and gaps. The following definition prescribes how to calculate these individual indices.

Definition 2.3.5: The index of a block B of A which contains k elements is $I_0(B) = k+1$ if B is an odd non-nodal interior block or an even nodal interior block. Otherwise, $I_0(B) = k$.

If all interior blocks in section 2.2 are considered to be non-nodal, then definition 2.3.5 corresponds nicely to lemma 2.2.4.

Lemma 2.3.4: Let $B_1,\ldots,B_p$ and $G_1,\ldots,G_s$ comprise all the blocks and gaps of $A = q^{-1}(\{0\}) \cap E$. Also, assume that $S^+(q) < \infty$. Then $S^+(q) = I_0(B_1) +\ldots+ I_0(B_p) + s$.

Proof: The proof follows readily from definition 2.2.3 and definitions 2.3.2 through 2.3.6.

This lemma plays the role in section 2.3 that lemma 2.2.3 played in the preceeding section. The following corollary is analogous to corollary 2.2.1(i).

Corollary 2.3.1: Let K denote the number of elements of $A = q^{-1}(\{0\}) \cap E$, let $\tilde{p}$ denote the number of its interior blocks which are odd non-nodal or even nodal, and s denote the number of its gaps. Then $S^+(q) = K + \tilde{p} + s$.

Proof: The proof is an immediate consequence of definition 2.3.5 and lemma 2.3.4.

A second look at example 2.3.1 may serve to help illuminate the developments from definition 2.3.2 to corollary 2.3.1. First note that for the example, $A = q^{-1}(\{0\}) \cap E = \{1, 2, 3, 4, 5, 6, 10\}$. Its blocks relative to E and $B_1 = \{1\}$, $B_2 = \{2\}$, $B_3 = \{3,4\}$, $B_4 = \{5,6\}$, and $B_5 = \{10\}$. Here B_1 is an odd nodal, B_2 an odd non-nodal, B_3 an even nodal, and B_4 an even non-nodal interior block. Thus $I_0(B_1) = 1$, $I_0(B_2) = 1 + 1 = 2$, $I_0(B_3) = 2 + 1 = 3$, $I_0(B_4) = 2$, and $I_0(B_5) = 1$ according to definition 2.3.5. Now the one gap of A relative to E is $G_1 = (\alpha_1,\beta_1) = (7,8)$. Thus lemma 2.3.4 confirms that $I(\sigma) = S^+(q) = I_0(B_1) +\ldots+ I_0(B_5) + 1 = 10$.

The following theorem establishes the conditions under which a polynomial exists which has the same alternating behavior as g(x). This theorem will be important to the proof of subsequent results.

Theorem 2.3.2: There exists a non-trivial polynomial $P(x) = v'u(x)$ such that the properties

i. $P(x) \geq 0$ whenever $q(x) \geq 0$ and

ii. $P(x) \leq 0$ whenever $q(x) \leq 0$

hold for all $x \in E$ if and only if $S^+(q) \leq m$.

Proof: Assume first that the polynomial exists. Note that properties i. and ii. imply that $P(x) = 0$ whenever $q(x) = 0$ for $x \in E$. Hence $S^+(q) \leq S^+(P)$. Now the number of sign changes of P(x) on [a,b] is no more than m as a consequence of lemma 2.1.1. Thus the same must be true for the number of sign changes on $E \subset [a,b]$ so that $S^+(q) \leq S^+(P) \leq m$.

Assume now that $S^+(q) \leq m$. Then the set $A = q^{-1}(\{0\}) \cap E$ must be finite so that its elements may be denoted by $x_1 <\ldots< x_k$. Also, corollary 2.3.1 implies that $K \leq S^+(q) \leq m$. At this point, the construction of the polynomial $P(x)$ is similar in spirit to the construction of the polynomial in theorem 2.2.3.

First let the sequence $t_1 < \ldots < t_r$ be obtained from A by preceding the left endpoint x_i of any interior block which is odd non-nodal or even nodal by $x_i - \varepsilon$ (where $0 < \varepsilon < \min_i(x_i - x_{i-1})$) and by including $(\alpha_i + \beta_i)/2$ for every gap $G_i = (\alpha_i, \beta_i)$ of A. Corollary 2.3.1 implies that $r = S^+(q) \leq m$. If $r < m$, then select $m - r$ additional points from $[a', a)$. Let $s_1 < \ldots < s_m$ denote these points and the t_i's. Now there exists $x_0 \in E$ such that $q(x_0) \neq 0$ and $q(x)/q(x_0) \geq 0$ for all $x \in (x_0, b] \cap E$. (If such an x_0 did not exist, then it would be the case that $q(x) = 0$ on E. But then $S^+(q) \geq m + 2$ according to the assumption that E contains at least $m + 2$ points and corollary 2.3.1. This contradiction ensures that such an x_0 exists.) Then let ℓ denote the size of the block adjoining b and $C = (-1)^{\ell} \text{sgn}[q(x_0)]$. (It may be that $\ell = 0$.)

Now consider the polynomial

$P_\varepsilon(x) = C|u(s_1), \ldots, u(s_m), u(x)| = v_\varepsilon' u(x)$. Note that

$P_\varepsilon^{-1}([0,\infty)) = E \cap ([s_m, b] \cup [s_{m-2}, s_{m-1}] \cup [s_{m-4}, s_{m-3}] \cup \ldots)$ and

$P_\varepsilon^{-1}((-\infty,0]) = E \cap ([s_{m-1}, s_m] \cup [s_{m-3}, s_{m-2}] \cup \ldots)$ due to the alternating property of the determinant and definition 2.1.2. Thus P_ε is nearly the desired polynomial; its only shortcoming might be an unwanted change of sign to the left of the left endpoint of an odd non-nodal or even nodal interior block. To alleviate this possible drawback, consider the limit polynomial

$P(x) = \lim_{\varepsilon \downarrow 0} v_\varepsilon' u(x)/\sqrt{v_\varepsilon' v_\varepsilon} = \lim_{\varepsilon \downarrow 0} w_\varepsilon' u(x) = v'u(x)$. The fact that the

w_ε lie on the unit sphere guarantees that such a limit point exists. v may be taken to be any such limit point. Because of the construction of the sequence $s_1,\ldots,s_m$, lemma 2.1.1, and the limiting process, $P^{-1}(\{0\}) \cap E = A = q^{-1}(\{0\}) \cap E$. The fact that $P(x) \geq 0$ whenever $q(x) \geq 0$ and $x \in E$ is due to the value of C, the construction of $s_1,\ldots,s_m$, and the limiting process. Likewise $P(x) \leq 0$ whenever $q(x) \leq 0$ and $x \in E$. The proof of the theorem is now complete.

Clearly this theorem plays the same role in section 2.3 as did theorem 2.2.3 in the preceding section. Note also that the developments from definition 2.3.2 to theorem 2.3.2 did not exploit any special properties of $q(x;\sigma)$ or $E = \text{Support}(\nu)$. In fact, they apply equally well to any function $q(x)$ defined on any compact set $E \subset [a,b]$.

At this point, a more applicable characterization of those measures representing moment points on the boundary of $\mathcal{M}_{m+1}$ may be proved.

Theorem 2.3.3: $\mu \in \partial\, \mathcal{M}_{m+1}$ if and only if its representation σ has index $I(\sigma) \leq m$.

Proof: Assume first that $\mu = \int_E u(x)d\sigma(x) \in \partial\, \mathcal{M}_{m+1}$. Then according to theorem 2.3.1(ii), there exists a non-trivial polynomial $P(x) = v'u(x)$ such that

$$\frac{d\sigma}{d\nu}(x) = \begin{cases} 1 & P(x) < 0 \\ 0 & P(x) > 0. \end{cases}$$

Now $I(\sigma) = S^+(q)$, where the function $q(x;\sigma) = \chi_{\text{Support}(\nu-\sigma)}(x) - \chi_{\text{Support}(\sigma)}(x)$. Thus $P(x) < 0$ implies that $q(x;\sigma) = -1$ and $P(x) > 0$ implies that $q(x;\sigma) = +1$ for $x \in E$. Application of theorem 2.3.2 immediately implies that $S^+(q) \leq m$.

Assume next that $\mu = \int_E u(x)d\sigma(x)$ and that $m \geq I(\sigma) = S^+(q)$. Then theorem 2.3.2 implies that there exists a non-trivial polynomial $P(x) = v'u(x)$ such that $P(x) \geq 0$ whenever $q(x) \geq 0$ and $P(x) \leq 0$ whenever $q(x) \leq 0$ for $x \in E$. Hence $q(x) < 0$ implies that $P(x) < 0$ and $q(x) > 0$ implies that $P(x) > 0$ for $x \in E$. Thus

$$\frac{d\sigma}{d\eta}(x) = \begin{cases} 1 & P(x) < 0 \\ 0 & P(x) > 0 \end{cases}$$

so that $v'\mu = \int_E P(x)d\sigma(x) = -\int_E P_-(x)d\nu(x)$. Then theorem 2.3.1(ii) implies that $\mu \in \partial\, \mathcal{M}_{m+1}$ and the theorem is proved.

Note that at this point, problem 2.1 has been solved for two cases. The first case is the trivial case covered by lemma 2.2.3: if E = Support(ν) consists of less than m+2 points, then every moment point $\mu \in \mathcal{M}_{m+1}$ is uniquely represented. The second case is covered by theorems 2.3.3 and 2.3.1(iii): if $I(\sigma) \leq m$, then $\mu \in \partial\, \mathcal{M}_{m+1}$ and is uniquely represented. In either of these cases, σ is unique so that μ_{m+1} is already determined by the values $\mu_0, \mu_1, \ldots, \mu_m$. The next task will be to solve problem 2.1 for the case that $\mu \in$ Interior($\mathcal{M}_{m+1}$). To that end the following notions will be needed.

Definition 2.3.6: A measure $\sigma \in \Sigma$ is principal if and only if $I(\sigma) = m + 1$.

Definition 2.3.7: A principal measure σ is upper (lower) if and only if the "sign" of $q(x;\sigma)$ is negative (positive) at $x = b$. Here "sign" is taken to mean the sign of the function for purposes of evaluating $S^{+}(\cdot)$. Since E is assumed to contain at least m + 2 points and $I(\sigma) = m + 1$, definition 2.2.3 implies that $q(x;\sigma)$ is not identically zero on E. Hence the "sign" of this function at x = b is uniquely determined.

Note the correspondence between definitions 2.3.6 and 2.2.7 as well as that between definitions 2.3.7 and 2.2.8. The following theorem completes the solution of problem 2.1 in the present setting. Note especially that the solution it provides reads exactly the same as that of theorem 2.2.5. Of course the terms in the two statements carry different but analogous meanings.

Theorem 2.3.4: Let $\mu \in \text{Interior}(\mathcal{M}_{m+1})$. Among those measures σ representing μ, the maximum (minimum) value of μ_{m+1} is attained if and only if σ is upper (lower) principal.

Proof: Recall that $\mathcal{M}_{m+2}$ is bounded according to lemma 2.3.1. At this point, the proof is nearly the same as that of theorem 2.2.5. Let $\underline{\mu}_{m+1} < \bar{\mu}_{m+1}$ be the smallest and largest values of μ_{m+1} attained by measures which represent $\mu \in \mathcal{M}_{m+1}$. Also, let $\underline{\sigma}$ and $\bar{\sigma}$ be measures which represent $\begin{bmatrix} \mu \\ \underline{\mu}_{m+1} \end{bmatrix}$ and $\begin{bmatrix} \mu \\ \bar{\mu}_{m+1} \end{bmatrix}$ (respectively). Exactly as proved in theorem 2.2.5, both $\underline{\sigma}$ and $\bar{\sigma}$ are unique and principal.

It will now be shown that $\underline{\sigma}$ must be lower principal. First note that since $I(\sigma) = m + 1$, theorem 2.3.2 assures the existence of a polynomial $P(x) = \sum_{i=0}^{m+1} v_i u_i(x)$ such that $P(x) \geq 0$ whenever $q(x;\sigma) \geq 0$ and $P(x) \leq 0$ whenever $q(x;\underline{\sigma}) \leq 0$ for $x \in E$. Recall that the construction of P(x) yields

$$v_{m+1} = \lim_{\varepsilon \downarrow 0} C \begin{vmatrix} u_0, \ldots, u_m \\ s_0, \ldots, s_m \end{vmatrix} \Big/ \sqrt{\sum_{i=0}^{m+1} (v_i^{\varepsilon})^2},$$

where $C = (-1)^{\ell} \operatorname{sgn}[q(x_0)]$. That is, C is the "sign" of $q(x;\underline{\sigma})$ for purposes of calculating $S^+(q)$. Suppose now that this "sign" were negative. This would imply that $v_{m+1} \leq 0$. Also, $v_{m+1} \neq 0$ or else theorems 2.3.1(ii) and 2.3.3 would imply that $m \geq I(\underline{\sigma}) = m + 1$. Then consider

$$\sum_{i=0}^{m} v_i \mu_i + v_{m+1} \underline{\mu}_{m+1} = -\int_E P_-(x) d\underline{\sigma}(x) \leq \sum_{i=0}^{m} v_i \mu_i + v_{m+1} \bar{\mu}_{m+1}.$$

Because $v_{m+1} < 0$, this inequality implies that $\underline{\mu}_{m+1} \geq \bar{\mu}_{m+1}$. This contradiction must mean that C is positive and hence that $\underline{\sigma}$ is lower principal.

The proof that $\bar{\sigma}$ must be upper principal is exactly the same except that C is assumed to be positive, $v_{m+1} > 0$, and hence $\bar{\mu}_{m+1} \leq \underline{\mu}_{m+1}$. This contradiction implies that C must be negative. Thus $\bar{\sigma}$ is upper principal and the proof is complete.

Note that in the course of proving this theorem, the existence and uniqueness of upper and lower principal representations have been established. Also, this theorem completes the solution of problem 2.1.

At this point, the results necessary to characterize the admissible designs for polynomial regression of degree n have been obtained.

<u>Theorem 2.3.5</u>: Let $\nu = \psi - \varphi$ and $E = \text{Support}(\nu)$.

i. If E consists of 2n or fewer points, then all designs are admissible.

ii. Otherwise, a design ξ is admissible if and only if either

a. $I(\xi-\varphi) < 2n$ or b. $I(\xi-\varphi) = 2n$ and $\xi-\varphi$ is upper.

Proof: The proof of this theorem is an application of moment space results now established to the case that $u_i(x) = x^i$ for $i = 0, 1, \ldots, m = 2n - 1$ and $\mu_0 = 1 - \varphi([a,b])$.

The remainder of the proof is the same as that of theorem 2.2.6 except the results of section 2.2 that are cited should be replaced by their counterparts in section 2.3.

Note that as an added consequence of the properties of $\mathcal{M}_{2n}$ and $\mathcal{M}_{2n+1}$, each of the admissible designs given by theorem 2.3.5 is unique.

The following will serve as an example of the application of theorem 2.3.5.

Example 2.3.2: Let $n = 2$, let $[a,b] = [-3,3]$, let $d\varphi = 0$, and let

$$d\psi(x) = \begin{cases} \frac{1}{3}dx & -3 \leq x < -2 \\ \frac{1}{2} & x = -2 \\ -\frac{1}{3}(x+1)dx & -2 < x \leq -1 \\ 0 & -1 \leq x < 0 \\ 1 & x = 0 \\ d\psi(-x) & 0 \leq x \leq -3. \end{cases}$$

That is, ψ is a symmetrical measure which has 3 point masses superimposed on an absolutely continuous measure. Now consider the design

$$d\xi(x) = \begin{cases} \frac{1}{3} & x = -2 \\ -\frac{1}{3}(x+1)dx & -2 < x \leq -1 \\ 0 & x \in [-3,-2) \cup [-1,0] \\ d\xi(-x) & 0 \leq x \leq 3. \end{cases}$$

Now $I(\xi) = S^{+}(q) = 4$ according to definition 2.3.1. Also, $q(3;\xi) = 1$ implies that ξ is a lower principal representation of $\mu = (1, 0, 65/18, 0)'$ with $\mu_4 = 203/15$. Hence theorem 2.3.5 implies that ξ is inadmissible for quadratic regression. Consider instead the design

$$d\bar{\xi}(x) = \begin{cases} \frac{1}{3}dx & -3 \leq x \leq -w \\ 0 & -w \leq x < 0 \\ 1-\frac{2}{3}(3-w) & x = 0 \\ d\bar{\xi}(-x) & 0 \leq x \leq 3, \end{cases}$$

where $w = (43/4)^{1/3}$. This design yields $\mu = (1, 0, 65/18, 0)'$ and $\bar{\mu}_4 = \frac{2}{15}(243-w^5) > \mu_4$. Thus $M(\bar{\xi}) \geq M(\xi)$ so that $\bar{\xi}$ is better than ξ in the sense of admissibility. In fact, $\bar{\xi}$ is an upper principal representation of μ so theorem 2.3.5 implies that $\bar{\xi}$ is admissible.

In the present setting, problem 2.2 will not be treated. Results would be much harder to obtain under the (φ,ψ) restriction than under the E restriction. The reason for the increased difficulty is that the E-problem is more nicely related to the unrestricted version of problem 2.2 than is the (φ,ψ)-problem.

2.4 The p-Problem

Recall that $\{G_\omega | \omega \in \Omega\}$ is a collection of disjoint open sets, that $p_\omega \in (0,1)$ for each $\omega \in \Omega$, and that $\Xi = \{\xi | \xi(G_\omega) \leq p_\omega;\ \omega \in \Omega\}$. For $\mathcal{X} = [a,b]$, it may be assumed without loss of generality that each G_ω is an interval which is open in [a,b]. It will be seen that the further assumption that $\Omega = \{1,\ldots,s\}$ imposes no loss of generality. It will be helpful to let $G_0 = [a,b] - \bigcup_{i=1}^{s} G_i$. For technical reasons, neither a nor b is allowed to belong to $\bigcup_{i=1}^{s} \partial G_i$. To obtain the results of this section, it will also be assumed that $u(x) = (1, x,\ldots,x^m)'$. For purposes of notation, let $M_{m+1} = \{\mu | \mu = \int_a^b u(x)d\xi(x);\ \xi \in \Xi\}$, let $\Sigma = \{\gamma\xi | \xi \in \Xi,\ \gamma \geq 0\}$, and let $\mathcal{M}_{m+1} = \{\mu | \mu = \int_a^b u(x)d\sigma(x);\ \sigma \in \Sigma\}$. As in sections 2.2 and 2.3, M_{m+1} is convex and closed. Note also that $\mathcal{M}_{m+1}$ is the closed convex cone generated by M_{m+1}. Furthermore, $\mathcal{M}_{m+1}$ may be viewed as a truncation of the (unrestricted) convex cone generated by u([a,b]).

A convention that will be adopted in this section is that ∂M_{m+1} denotes the boundary of M_{m+1} relative to $\{1\} \times \mathbb{R}^m$. That is, $\partial M_{m+1} = M_{m+1} \cap \partial \mathcal{M}_{m+1}$. A further convention will be that $\text{Interior}(M_{m+1}) = M_{m+1} - \partial M_{m+1}$.

The first aim of this section is to provide necessary conditions for the solution of problem 2.1 when $s \geq 1$. It will then be shown that these conditions are also sufficient for the solution of problem 2.1 in two special cases. The first case is that s = 1 and $G_1 = (\alpha_1,b]$ for $\alpha_1 > a$ (or $G_1 = [a,\beta_1)$ for $\beta_1 < b$). The second case is that $s = 2$, that $G_1 = [a,\beta_1)$ for $\beta_1 < b$, and that $G_2 = (\alpha_2,b]$ for $\beta_1 < \alpha_2$.

The results for $s \geq 1$ will first be derived for the case that $\xi(G_i) = p_i$ for $i = 1,\ldots,s$ (and $p_1 + \ldots + p_s \leq 1$). The results for general designs $\xi \in \Xi$ will then follow readily by an induction argument. Now let the "dual cone" of $\mathcal{M}_{m+1}$ be denoted by $\mathcal{P}_{m+1} = \mathcal{M}^+_{m+1} = \{\pi | \pi'\mu \geq 0 \text{ for all } \mu \in \mathcal{M}_{m+1}\}$.

<u>Lemma 2.4.1</u>: Let $\xi(G_i) = p_i$ for $i = 1,\ldots,s$ and let

$\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$.

i. There exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0$.

ii. Let $\lambda_i = \inf_{G_i} \pi'u(x_i)$ for $i = 0,\ldots,s$. Then

$$\text{Support}(\xi) \subset \bigcup_{i=0}^{s} \{x_i \in G_i | \pi'u(x_i) = \lambda_i\}. \qquad (2.4.1)$$

<u>Proof</u>: i. First note that $\mu \in \partial M_{m+1}$ implies that $\mu \in \partial \mathcal{M}_{m+1}$. The conclusion of this part then follows immediately from theorem 2.2.2(ii).

ii. To establish the second conclusion, it will first be shown that

$$\sum_{i=1}^{s} p_i \lambda_i + (1 - \sum_{i=1}^{s} p_i)\lambda_0 \geq 0. \qquad (2.4.2)$$

If this were not true, then there would exist $x_i \in G_i$ for $i = 0,\ldots,s$ such that $\sum_{i=1}^{s} p_i \pi'u(x_i) + (1 - \sum_{i=1}^{s} p_i)\pi'u(x_0) < 0$. Then set

$\tilde{\xi} = \sum_{i=1}^{s} p_i \delta_{x_i} + (1 - \sum_{i=1}^{s} p_i)\delta_{x_0}$ and note that

$\pi'\tilde{\mu} = \int_a^b \pi'u(x)d\tilde{\xi}(x) = \sum_{i=1}^{s} p_i\pi'u(x_i) + (1 - \sum_{i=1}^{s} p_i)\pi'u(x_0) < 0$. But $\pi \in \mathcal{P}_{m+1}$ so that this contradiction implies that (2.4.2) does hold. Thus,

$$0 = \pi'\mu = \int_a^b \pi'u(x)d\xi(x) \geq \sum_{i=1}^{s} \xi(G_i)\lambda_i + [1 - \sum_{i=1}^{s} \xi(G_i)]\lambda_0$$

$$= \sum_{i=1}^{s} p_i\lambda_i + (1 - \sum_{i=1}^{s} p_i)\lambda_0 \geq 0.$$

Furthermore, equality can hold only if (2.4.1) is valid. Thus the lemma is proved.

The following will be analogous to definition 2.2.3.

Definition 2.4.1: Let $\xi(G_i) = p_i$ for all $i \in W \subset \Omega = \{1,\ldots,s\}$ and $\xi(G_i) < p_i$ otherwise. Then the index of the measure $\xi \in \Xi$ is $I(\xi) = I_0(\text{Support}(\xi) - \bigcup_{i \in W} \partial G_i)$. Here $I_0(\cdot)$ is as defined by corollary 2.2.2.

It might be emphasized at this point that if $G_i = (\alpha_i, b]$ then $\partial G_i = \{\alpha_i\}$, if $G_i = [a, \beta_i)$ then $\partial G_i = \{\beta_i\}$, and if $G_i = (\alpha_i, \beta_i) \subset (a,b)$ then $\partial G_i = \{\alpha_i, \beta_i\}$. The following example will illustrate the index calculation.

Example 2.4.1: Let $[a,b] = [-2,1]$, let $s = 1$, let $G_1 = (0,1]$, let $p_1 = 1/5$, and note that $\partial G_1 = \{0\}$. Consider first the measure $\xi = 1/6\delta_{-1} + 2/3\delta_0 + 1/6\delta_1$. Then $\xi(G_1) = 1/6 < p_1$ so that $I(\xi) = I_0(\text{Support}(\xi)) = 5$. Consider next the measure $\bar{\xi} = 5/210\delta_{-2} + 128/210\delta_{-1/4} + 35/210\delta_0 + 1/5\delta_1$. Then $\bar{\xi}(G_1) = 1/5 = p_1$ so that $I(\bar{\xi}) = I_0(\text{Support}(\bar{\xi}) - \{0\}) = 4$. This illustrates the essence of definition 2.4.1: if ξ assigns full mass to a G_i, then G_i's boundary is excluded from the index calculation. It is also seen that $I(\xi) > I(\bar{\xi})$ even though $I_0(\text{Support}(\xi)) < I_0(\text{Support}(\bar{\xi}))$.

The following theorem provides a necessary condition for $\mu \in \partial M_{m+1}$ in the special case at hand. In that sense, it is analogous to theorems 2.2.4(i) and 2.3.3.

Theorem 2.4.1: Let $\xi(G_i) = p_i$ for $i = 1,\ldots,s$. If $\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$, then $I(\xi) \leq m$.

Proof: A case by case analysis for m even and m odd will show that $I(\xi) > m$ would contradict lemma 2.4.1.

Suppose first that $m = 2k$ and $I(\xi) > m$. Note that this can happen only if $\text{Support}(\xi) - \bigcup_{i=1}^{s} \partial G_i$ includes more than k interior points or it contains exactly k interior points and at least one endpoint. Now let $P(x) = \pi'u(x)$ be the polynomial given by lemma 2.4.1. Observe that (2.4.1) can hold only if $P'(x) = 0$ at every support point belonging to $(a,b) - \bigcup_{i=1}^{s} \partial G_i$ and $P'(x) = 0$ has at least one root between every pair of support points which do not belong to $\bigcup_{i=1}^{s} \partial G_i$. Thus if $\text{Support}(\xi) - \bigcup_{i=1}^{s} \partial G_i$ includes $k + 1$ interior points, then $P'(x)$ must have at least $(k+1) + k = m + 1$ roots. This is impossible for a non-trivial polynomial of degree $m - 1$. (If $P(x) = \pi'u(x)$ were constant, then $\pi'\mu = 0$ for $\pi \neq 0$ cannot hold.) Similarly, if

$\text{Support}(\xi) - \bigcup_{i=1}^{s} \partial G_i$ contains exactly k interior points and at least one endpoint, then P'(x) must have at least k + k = m roots which is impossible. Hence $I(\xi) > m$ cannot hold for $m = 2k$.

Suppose now that $m = 2k + 1$ and $I(\xi) > m$. Note that this can happen only if $\text{Support}(\xi) - \bigcup_{i=1}^{s} \partial G_i$ includes more than k interior points or it contains exactly k interior points and both endpoints. For the first possibility, P'(x) must have at least $(k+1) + k = m$ roots. For the second, P'(x) must have at least $k + (k+1) = m$ roots. Since this is impossible, $I(\xi) > m$ cannot hold for $m = 2k + 1$ and the theorem is proved.

An indication of why $I(\xi) \leq m$ does not necessarily imply $\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$ will be provided with the results for the special case $s = 1$. This discussion will show that the implication does not hold only if there is an interior interval $G_i = (\alpha_i, \beta_i) \subset (a,b)$ such that $\xi(G_i) = p_i$.

The next task will be to provide a necessary condition for the solution of problem 2.1 when $\mu \in \text{Interior}(M_{m+1})$. To that end, the following notions will be needed.

<u>Definition 2.4.2</u>: A measure $\xi \in \Xi$ is <u>principal</u> if and only if $I(\xi) = m + 1$.

<u>Definition 2.4.3</u>: A principal measure $\xi \in \Xi$ is <u>upper</u> (<u>lower</u>) if and only if the point b belongs (does not belong) to Support(ξ).

Note that definition 2.4.2 is analogous to definitions 2.2.7 and 2.3.6. The only differences are the meanings of index in the different sections. Note also that definition 2.4.3 is exactly the same as that given by corollary 2.2.3.

It might be noted that the assumption that neither a nor b belongs to $\bigcup_{i=1}^{s} \partial G_i$ is needed for the following theorem to be proven.

<u>Theorem 2.4.2</u>: Let $\xi(G_i) = p_i$ for $i = 1,\ldots,s$ and let $\mu \in \text{Interior}(M_{m+1})$. If $\xi \in \Xi$ attains the maximum (minimum) value of μ_{m+1} among all measures from Ξ which represent μ, then either $I(\xi) \leq m$ or ξ is upper (lower) principal.

<u>Proof</u>: Note first that $|\mu_{m+1}| \leq \max_{a \leq x \leq b} |u_{m+1}(x)|$. Then let $\underline{\mu}_{m+1} < \bar{\mu}_{m+1}$ be the smallest and largest values of μ_{m+1} attained by measures from Ξ which represent $\mu \in M_{m+1}$. Also, let $\underline{\xi}$ and $\bar{\xi}$ be measures which represent $\begin{bmatrix} \mu \\ \underline{\mu}_{m+1} \end{bmatrix}$ and $\begin{bmatrix} \mu \\ \bar{\mu}_{m+1} \end{bmatrix}$ (respectively). Now these two points lie on the lower and upper boundaries of M_{m+2} (respectively), so theorem 2.4.1 implies that $I(\underline{\xi}) \leq m + 1$ and $I(\bar{\xi}) \leq m + 1$. Thus the only way for the conclusion of the theorem to be violated is for $\underline{\xi}$ to be upper principal or $\bar{\xi}$ to be lower principal.

It will now be shown that $\underline{\xi}$ cannot be upper principal. According to lemma 2.4.1(i), there exists $(\pi_0,\ldots,\pi_m, \pi_{m+1})' \in \mathcal{P}_{m+2} - \{0\}$ such that $\sum_{i=0}^{m} \pi_i\mu_i + \pi_{m+1}\underline{\mu}_{m+1} = 0$. Here $\pi_{m+1} \neq 0$ or else $\mu \in \partial M_{m+1}$. Suppose that $b \in \text{Support}(\underline{\xi})$. In the case that $m = 2k$, $I(\underline{\xi}) = m + 1$ implies that $\text{Support}(\underline{\xi}) - \bigcup_{i=1}^{s} \partial G_i$ consists of the point b and k points from

$(a,b) - \bigcup_{i=1}^{s} \partial G_i$. Then with $P(x) = \sum_{i=0}^{m+1} \pi_i u_i(x)$, it must be true that $\pi_{m+1} < 0$ or else $P'(x)$ has $k + k + 1 = m + 1$ roots which is impossible for a non-trivial polynomial of degree m. In the case that $m = 2k + 1$, $I(\underline{\xi}) = m + 1$ implies that $\text{Support}(\underline{\xi}) - \bigcup_{i=1}^{s} \partial G_i$ consists of the points a and b as well as k points from $(a,b) - \bigcup_{i=1}^{s} \partial G_i$. As in the previous case, $\pi_{m+1} < 0$ or else $P'(x)$ has $k + (k+1) + 1 = m + 1$ roots. Now by definition of $\mathcal{P}_{m+2}$, $0 \leq \sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1}\mu_{m+1}$ for all $\begin{bmatrix} \mu \\ \mu_{m+1} \end{bmatrix} \in M_{m+2}$. In particular,

$$\sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1}\underline{\mu}_{m+1} = 0 \leq \sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1}\bar{\mu}_{m+1}.$$

Since $\pi_{m+1} < 0$, this inequality implies that $\mu_{m+1} \geq \bar{\mu}_{m+1}$ which is a contradiction. Thus $b \in \text{Support}(\underline{\xi})$ cannot hold so that $\underline{\xi}$ cannot be upper principal.

The proof that $\bar{\xi}$ cannot be lower principal proceeds with only minor modifications and will therefore be omitted. Thus the theorem is proved.

An indication of why the converse of this theorem need not hold will be provided with the results for the special cases s = 1 and 2. The problem will be seen to occur only if there is an interior interval G_i such that $\xi(G_i) = p_i$.

At this point, note that theorems 2.4.1 and 2.4.2 have established necessary conditions for the solution of problem 2.1 in the case that $\xi(G_i) = p_i$ for $i = 1,\ldots,s$.

To establish necessary conditions for the solution of problem 2.1 in the general case, the following notation will prove useful. For any set $W \subset \Omega = \{1,\ldots,s\}$, let

$$\Xi(W) = \{\xi | \xi(G_i) \leq p_i \text{ for all } i \in W\}.$$

That is, $\Xi(W)$ is the set of allowable design measures for a "sub-problem" with (possibly) fewer design restrictions. The set W prescribes which of the restrictions are in effect for the sub-problem. Note that $\Xi(\Omega) = \Xi$ and that $\Xi(\phi) = \Xi_0$, the set of all probability measures on $\mathcal{X} = [a,b]$. Note also that $W_1 \subset W_2 \subset \Omega$ implies that $\Xi(W_1) \supset \Xi(W_2)$.

The following lemma relates such sub-problems to Ξ in the context of problem 2.1.

<u>Lemma 2.4.2</u>: Let $W \subset \Omega$, let $\xi \in \Xi$, and let $\xi(G_i) < p_i$ for all $i \in \Omega - W$. If ξ solves problem 2.1 among all designs from Ξ, then ξ solves problem 2.1 among all designs from $\Xi(W)$.

<u>Proof</u>: Recall that ξ solves problem 2.1 if and only if it attains the largest (or smallest) value of μ_{m+1} among all design measures sharing the same values of $\mu_0,\ldots,\mu_m$. Without loss of generality, it may be assumed that μ_{m+1} is to be maximized.

Then suppose that there exists $\bar{\xi} \in \Xi(W)$ such that

$\int_a^b u(x)d\bar{\xi}(x) = \mu = \int_a^b u(x)d\xi(x)$ but with

$\int_a^b u_{m+1}(x)d\bar{\xi}(x) = \bar{\mu}_{m+1} > \mu_{m+1} = \int_a^b u_{m+1}(x)d\xi(x)$. Now since $\xi(G_i) < p_i$ for all $i \in \Omega - W$, there exists $\rho \in (0,1]$ such that

$p_i \geq (1-\rho)\xi(G_i) + \rho\bar{\xi}(G_i)$ for all $i \in \Omega - W$. Thus $\tilde{\xi} = (1-\rho)\xi + \rho\bar{\xi} \in \Xi$. Also, $\int_a^b u(x)d\tilde{\xi}(x) = (1-\rho)\mu + \rho\mu = \mu$ but

$\tilde{\mu}_{m+1} = \int_a^b u_{m+1}(x)d\tilde{\xi}(x) = (1-\rho)\mu_{m+1} + \rho\bar{\mu}_{m+1} > \mu_{m+1}$. That is, ξ does not attain the largest value of μ_{m+1} among all designs from Ξ which share the same values of $\mu_0,\ldots,\mu_m$. This contradiction must mean that no such $\bar{\xi} \in \Xi(W)$ exists. Thus the lemma is proved.

The following theorem provides a necessary condition for the solution of problem 2.1 in the general case.

Theorem 2.4.3: If $\xi \in \Xi$ attains the maximum (minimum) value of μ_{m+1} among all measures from Ξ which represent μ, then either $I(\xi) \leq m$ or ξ is upper (lower) principal.

Proof: The proof will proceed by induction on s, the number of restricting intervals.

If $s = 0$, then $\Xi = \Xi_0$, the unrestricted set of design measures. In this case, the conclusion follows immediately from theorems 2.2.4(i) and 2.2.5 with $E = [a,b]$.

Suppose now that the conclusion holds for $s = 0, 1,\ldots,r-1$ and consider $s = r$. If $\xi(G_i) = p_i$ for $i = 1,\ldots,s$, then the conclusion follows immediately from theorems 2.4.1 and 2.4.2. Thus it may now be assumed that there exists a set W which is strictly contained in $\Omega = \{1,\ldots,r\}$ such that $\xi(G_i) = p_i$ for all $i \in W$ and $\xi(G_i) < p_i$ for all $i \in \Omega - W$. Then lemma 2.4.2(ii) implies that ξ must solve problem 2.1 among all designs from $\Xi(W)$. According to the induction hypothesis, the conclusion is now established in this "reduced" case. Thus the induction is complete and the theorem proved.

The following theorem is the application of theorem 2.4.3 to the admissibility problem.

Theorem. 2.4.4: If a design $\xi \in \Xi$ is admissible for polynomial regression of degree n, then either a. $I(\xi) < 2n$ or b. $I(\xi) = 2n$ and $b \in \text{Support}(\xi)$.

Proof: This theorem follows immediately from theorems 2.1.1 and 2.4.3.

The conditions of this theorem are only necessary for a design to be admissible. Thus the set of all designs which satisfy these conditions form a "complete class". In that sense, a helpful simplification of the admissibility problem has been achieved.

It will now be shown that the conditions of theorem 2.4.3 are sufficient for the solution of problem 2.1 in two special cases. The first is that $s = 1$ and $G_1 = (\alpha_1, b]$ for $\alpha_1 > a$. (Note that $G_1 = [a, \beta_1)$ for $\beta_1 < b$ is essentially the same case.) Here $G_0 = [a,b] - G_1$.

In the case that $s = 1$, the following lemma characterizes the dual cone $\mathcal{P}_{m+1}$. Note that it holds regardless of the form of G_1.

Lemma 2.4.3: Let $s = 1$, let $\pi \in \mathbb{R}^{m+1}$, and let $\lambda_i = \inf_{G_i} \pi' u(x_i)$ for $i = 0,1$. Then $\pi \in \mathcal{P}_{m+1}$ if and only if $\lambda_0 \geq 0$ and $p_1\lambda_1 + (1-p_1)\lambda_0 \geq 0$.

Proof: Assume first that $\pi \in \mathcal{P}_{m+1}$. Then setting $\sigma = \delta_{x_0} \in \Sigma$ for any $x_0 \in G_0$ yields $0 \leq \pi'\mu = \pi'u(x_0)$. Hence $\lambda_0 \geq 0$. Similarly, setting $\sigma = p_1\delta_{x_1} + (1-p_1)\delta_{x_0}$ for any $x_1 \in G_1$ and $x_0 \in G_0$ yields $0 \leq \pi'\mu = p_1\pi'u(x_1) + (1-p_1)\pi'u(x_0)$. Hence $p_1\lambda_1 + (1-p_1)\lambda_0 \geq 0$.

Assume next that $\lambda_0 \geq 0$ and $p_1\lambda_1 + (1-p_1)\lambda_0 \geq 0$. Now let $\mu = \int_a^b u(x)d\sigma(x)$ for any $\sigma = \gamma\xi \in \Sigma$ (where $\xi \in \Xi$ and $\gamma \geq 0$). Note that if $\lambda_1 \geq 0$, then $\pi'\mu = \int_a^b \pi'u(x)d\sigma(x) \geq 0$. Assume now that $\lambda_1 < 0$. Then

$$\begin{aligned}\pi'\mu = \int_a^b \pi'u(x)d\sigma(x) &\geq \sigma(G_1)\lambda_1 + \sigma(G_0)\lambda_0 \\ &= \gamma\{\xi(G_1)\lambda_1 + [1-\xi(G_1)]\lambda_0\} \\ &\geq \gamma\{p_1\lambda_1 + [1-p_1]\lambda_0\} \geq 0.\end{aligned}$$

Thus $\pi \in \mathcal{P}_{m+1}$ and the lemma is proved.

The following theorem provides a converse to theorem 2.4.1 in the case that $s = 1$, either a or $b \in G_1$, and $\xi(G_1) \leq p_1$. In order to prove the theorem, the following preliminary lemma is needed. It is essentially a statement that $u_0,\ldots,u_m$ form an extended Tchebycheff system of order 1.

Lemma 2.4.4: Let $r \leq s \leq m - r + 1$, let $\{x_1,\ldots,x_s\} \subset [a,b]$, and let $u'(x) = (0, 1, 2x,\ldots,mx^{m-1})'$ denote the derivative vector of $u(x)$. Then the vectors $u'(x_1),\ldots,u'(x_r)$, $u(x_1),\ldots,u(x_s)$ are linearly independent.

Proof: To demonstrate independence, choose additional points $x_{s+1},\ldots,x_{m-r+1}$ and consider the determinant

$$\Delta = |u(x_1),\ u'(x_1),\ldots,u(x_r),\ u'(x_r),\ u(x_{r+1}),\ldots,u(x_{m+1-r})|$$

$$= \lim_{\varepsilon\downarrow 0}|u(x_1),\ [u(x_1+\varepsilon) - u(x_1)],\ldots,u(x_r),\ [u(x_r+\varepsilon) - u(x_r)],$$

$$u(x_{r+1}),\ldots,u(x_{m+1-r})|/\varepsilon^r$$

$$= \lim_{\varepsilon\downarrow 0}|u(x_1),\ u(x_1+\varepsilon),\ldots,u(x_r),\ u(x_r+\varepsilon),\ u(x_{r+1}),\ldots,u(x_{m+1-r})|/\varepsilon^r.$$

Now let $t_1,\ldots,t_{m+1}$ denote the points

$x_1,\ x_1 + \varepsilon,\ldots,x_r,\ x_r + \varepsilon,\ x_{r+1},\ldots,x_{m+1-r}$.

Then $\Delta = \lim_{\varepsilon\downarrow 0} \prod_{1\le i<j\le m+1} (t_j - t_i)/\varepsilon^r$

$$= \prod_{1\le i<j\le r} (x_j - x_i)^2 \prod_{1\le k<\ell\le m+1-r} (x_\ell - x_k)^2 \neq 0.$$

Thus the vectors comprising the colums of Δ must be independent so that the proof is complete.

<u>Theorem 2.4.5</u>: Let $s = 1$ and either a or $b \in G_1$. If $I(\xi) \le m$, then $\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$.

<u>Proof</u>: Note first that the proof need only be carried out for $I(\xi) = m$. The conclusion will follow immediately in the case that $I(\xi) = \ell < m$. For then, $(\mu_0,\ldots,\mu_\ell)' \in \partial M_{\ell+1}$ implies that $\mu = (\mu_0,\ldots,\mu_m)' \in \partial M_{m+1}$. Now in the case that $I_0(\text{Support}(\xi)) \le m$, theorem 2.2.4(i) (with $E = [a,b]$) implies that $\mu \in \partial M_{m+1}$. Thus the theorem need only be proven for the case $I(\xi) = m < I_0(\text{Support}(\xi))$. It will be assumed that $G_1 = (\alpha_1,b]$ for $\alpha_1 > a$. The proof for $G_1 = [a,\beta_1)$ and $\beta_1 < b$ proceeds exactly the same except that β_1 replaces α_1.

By inspection, there are precisely three possible ways to achieve $I(\xi) = m < I_0(\text{Support}(\xi))$. All of them require $\xi(G_1) = p_1$. For the first possibility, $m = 2k$ and Support(ξ) consists of k points from $(a,b) - \{\alpha_1\}$ as well as the point α_1. For the second, $m = 2k$ and Support(ξ) consists of $k - 1$ points from $(a,b) - \{\alpha_1\}$, the point α_1, and both endpoints. For the final possibility, $m = 2k + 1$ and Support(ξ) consists of k points from $(a,b) - \{\alpha_1\}$, the point α_1, and exactly one endpoint.

For each of these cases it will be argued that there exists a polynomial $Q(x) = v'u(x)$ such that $Q(x) \geq 0$ on G_1 and $Q(x) \geq 1$ on $[a,b] - G_1$ with equality precisely on Support(ξ). Then setting $e_0 = (1, 0,\ldots,0)'$ and $\pi = [v + (p-1)e_0]/p$ yields $\pi \in \mathcal{P}_{m+1} - \{0\}$ (according to lemma 2.4.3) and $\pi'\mu = \int_a^b \pi'u(x)d\xi(x) = (1-p)1 + p(1-1/p) = 0$ so that $\mu \in \partial M_{m+1}$.

The existence of the polynomial Q(x) will be demonstrated only for the second possible case: $m = 2k$ and $\text{Support}(\xi) = \{a, b, \alpha_1, x_1,\ldots,x_{k-1}\}$, where $x_1 <\ldots< x_{k-1}$. The other two cases are readily proven with only minor modifications to the proof of this case.

First let r denote the number of support points contained in (a,α_1). The polynomial $Q(x) = v'u(x)$ must satisfy the equations $Q(a) = Q(\alpha_1) = Q(x_1) =\ldots= Q(x_r) = 1$, the equations $Q(b) = Q(x_{r+1}) =\ldots= Q(x_{k-1}) = 0$, and the equations $Q'(x_1) =\ldots= Q'(x_{k-1}) = 0$. Thus v must solve the linear system

$$\begin{bmatrix} u_0(a) & \cdots & u_m(a) \\ u_0(x_1) & \cdots & u_m(x_1) \\ \vdots & & \vdots \\ u_0(x_r) & \cdots & u_m(x_r) \\ u_0(\alpha_1) & \cdots & u_m(\alpha_1) \\ u_0(x_{r+1}) & \cdots & u_m(x_{r+1}) \\ \vdots & & \vdots \\ u_0(x_{k-1}) & \cdots & u_m(x_{k-1}) \\ u_0(b) & \cdots & u_m(b) \\ u_0'(x_1) & \cdots & u_m'(x_1) \\ \vdots & & \vdots \\ u_0'(x_{k-1}) & \cdots & u_m'(x_{k-1}) \end{bmatrix} \begin{bmatrix} v_0 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

According to lemma 2.4.4, the coefficient matrix of v in this system is non-singular and so the polynomial $Q(x) = v'u(x)$ does exist.

It remains to be demonstrated that $Q(x) \geq 1$ on $[a,\alpha_1]$ and $Q(x) \geq 0$ on $(\alpha_1,b]$. Suppose that there exists an x_0 which violates either inequality. This would imply that $Q'(x)$ has at least $(k-1) + (k+1) = m$ roots (counting multiplicities) which is impossible. Finally, $Q(x)$ cannot attain equality at a non-support point or else $Q'(x)$ has at least m+1 roots. Thus the proof of the theorem is now complete.

As already noted, this theorem need not hold if $G_1 = (\alpha_1,\beta_1) \subset (a,b)$. To illustrate this, consider the setting with $m = 4$, with $\text{Support}(\xi) - \{\alpha_1,\beta_1\} = \{a, b, x_1\}$ for $\alpha_1 < x_1 < \beta_1$, and with $\xi(G_1) = \xi(\{x_1\}) = p_1$. Then $I(\xi) = 4$ but $\mu \in \partial M_{m+1}$ only if there exists a polynomial $Q(x)$ as in the proof of theorem 2.4.5. That is, there must exist a polynomial $Q(x)$ such that $Q(x) \geq 1$ on $[a,\alpha_1] \cup [\beta_1,b]$ and $Q(x) \geq 0$ on (α_1,β_1) with equality precisely at the support points of ξ.

Consider first the possibility that Support$(\xi) = \{a, b, \alpha_1, x_1\}$. Then

$$Q(x) = 1 - C(x-a)(x-\alpha_1)(x-w)(x-b) = 1 - C(x-w)q(x)$$

for some $w \in (x_1,\beta_1]$. (If $w > \beta_1$ held, then $Q(\beta_1) < 1$ would hold which is not allowed.) Now solving the necessary condition that $0 = Q'(x_1)$ yields $w = x_1 + q(x_1)/q'(x_1)$. Then manipulating the requirement $w \leq \beta_1$ yields the equivalent form

$$g(x_1) = \frac{1}{x_1-a} + \frac{1}{x_1-\alpha_1} + \frac{1}{x_1-\beta_1} + \frac{1}{x_1-b} \geq 0.$$

Now note that the function g can have at most three zereos. Note also that $g(\alpha_1-) = g(\beta_1-) = g(b-) = -\infty$ and $g(a+) = g(\alpha_1+) = g(\beta_1+) = +\infty$. Thus the function g must have exactly one zero in each of the intervals (a,α_1), (α_1,β_1), and (β_1,b). Let $\tilde{x}$ denote the zero which belongs to (α_1,β_1). Then $g(x_1) \geq 0$ can hold only if $x_1 \leq \tilde{x}$. That is, a design ξ with Support$(\xi) = \{a, b, \alpha_1, x_1\}$ achieves

$\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$ only if $x_1 \leq \tilde{x}$.

Reversing the roles of α_1 and β_1 yields the conclusion that a design ξ with Support$(\xi) = \{a, b, \beta_1, x_1\}$ achieves

$\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$ only if $x_1 \geq \tilde{x}$. Furthermore, it is clear that if Support$(\xi) = \{a, b, \alpha_1, \beta_1, x_1\}$ and ξ is to represent a boundary point, then it must be true that $x_1 = \tilde{x}$.

The following theorem establishes the uniqueness of a representation of any boundary moment point. It is analogous to theorem 2.2.4(ii).

Theorem 2.4.6: Let $s = 1$. Then $\mu \in \partial M_{m+1}$ if and only if it has a unique representation $\xi \in \Xi$.

Proof: Assume first that $\mu \in \text{Interior}(M_{m+1})$. Then there must exist $\underline{\mu}_{m+1} < \bar{\mu}_{m+1}$ such that $\begin{bmatrix} \mu \\ \underline{\mu}_{m+1} \end{bmatrix}$ and $\begin{bmatrix} \mu \\ \bar{\mu}_{m+1} \end{bmatrix}$ belong to M_{m+2}. If $\underline{\xi}$ and $\bar{\xi}$ denote measures representing these two points of M_{m+2}, then both $\underline{\xi}$ and $\bar{\xi}$ represent $\mu \in M_{m+1}$.

Assume next that $\mu \in \partial M_{m+1}$ and that ξ represents μ. First consider the case that $\xi(G_1) < p_1$. Now there exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0$. Let $\lambda_i = \inf_{G_i} \pi' u(x_i)$ for $i = 0,1$. According to lemma 2.4.3, $\lambda_0 \geq 0$ and $p_1\lambda_1 + (1-p)\lambda_0 \geq 0$. If $\lambda_1 \geq 0$, then theorems 2.2.2(ii) and 2.2.4(ii) imply that ξ is unique. If $\lambda_1 < 0$, then

$$
\begin{aligned}
0 = \pi'\mu &= \int_a^b \pi' u(x) d\xi(x) \\
&\geq \xi(G_1)\lambda_1 + [1-\xi(G_1)]\lambda_0 \\
&> p_1\lambda_1 + (1-p_1)\lambda_0 \geq 0.
\end{aligned}
$$

Thus $\lambda_1 < 0$ is impossible and ξ is unique if $\xi(G_1) < p_1$.

Now consider the case that $\xi(G_1) = p_1$ and $\mu = \int_a^b u(x) d\xi(x) \in \partial M_{m+1}$. Then there exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0$. Let $\lambda_i = \inf_{G_i} \pi' u(x_i)$ for $i = 0,1$. Lemma 2.4.1(ii) implies that

$$\text{Support}(\xi) \subset \bigcup_{i=0}^{1} \{x_i \in G_i \mid \pi' u(x_i) = \lambda_i\} \qquad (2.4.3)$$

for any measure which represents μ. Let $A = \{x_0,\ldots,x_k\}$ denote the right hand side of (2.4.3). By now familiar derivative arguments, k is maximized if $G_1 = (\alpha_1,\beta_1) \subset (a,b)$ and A consists of α_1, β_1, a, b, and $[(m + 1 - 4)/2]$ points from $(a,b) - \{\alpha_1,\beta_1\}$. Thus $k \leq 3 + [(m-3)/2]$ which is less than or equal to m unless $m = 1$. In the special case that $m = 1$, the number k is seen to be 1 or less. Hence the linear system $\mu = \int_a^b u(x)d\xi(x) = \sum_{i=0}^{k} u(x_i)\xi_i$ must yield a unique solution for the masses $\xi_i = \xi(\{x_i\})$. Thus ξ is unique and the proof complete.

Note that the conclusion of this theorem holds for $s = 1$ regardless of the form of G_1. Note also that theorems 2.4.5 and 2.4.6 provide a sufficient condition for the solution of problem 2.1 in the special case at hand. If $I(\xi) \leq m$, then $\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$ and ξ must be unique. Thus the value of μ_{m+1} is already determined by the values $\mu_0,\ldots,\mu_m$.

The following theorem provides a sufficient condition for the solution of problem 2.1 otherwise.

<u>Theorem 2.4.7</u>: Let $s = 1$ and either a or $b \in G_1$. Also, let $\mu \in \text{Interior}(M_{m+1})$. If ξ is upper (lower) principal, then it attains the maximum (minimum) value of μ_{m+1} among all measures from Ξ which represent μ.

Proof: Since $I(\xi) = m + 1$, theorem 2.4.5 implies that

$\begin{bmatrix} \mu \\ \mu_{m+1} \end{bmatrix} \in \partial M_{m+2}$. Thus ξ is unique (with respect to M_{m+2}) and there exists $(\pi_0, \ldots, \pi_m, \pi_{m+1})' \in \mathcal{P}_{m+2} - \{0\}$ such that $\sum_{i=0}^{m} \pi_i \mu_i + \pi_{m+1}\mu_{m+1} = 0$. As in the proof of theorem 2.4.2, $\pi_{m+1} < 0$ if ξ is upper and $\pi_{m+1} > 0$ if ξ is lower principal. Now let $\tilde{\xi}$ be any other measure representing μ. As in the proof of theorem 2.4.2, $\tilde{\mu}_{m+1} = \int_a^b u_{m+1}(x) d\tilde{\xi}(x) \leq \mu_{m+1}$ if $\pi_{m+1} < 0$ and $\tilde{\mu}_{m+1} \geq \mu_{m+1}$ if $\pi_{m+1} > 0$. That is, ξ attains the maximum (minimum) value of μ_{m+1} among all measures from Ξ which represent μ if ξ is upper (lower) principal.

Note that at this point the existence and uniqueness of the upper and lower principal representations have been established. Also, problem 2.1 has been solved for $s = 1$ when a or $b \in G_1$. The following theorem applies these results to the admissibility problem. It serves as the converse to theorem 2.4.4.

Theorem 2.4.8: Let $s = 1$ and either a or $b \in G_1$. If either a. $I(\xi) < 2n$ or b. $I(\xi) = 2n$ and $b \in \text{Support}(\xi)$, then $\xi \in \Xi$ is admissible for polynomial regression of degree n.

Proof: The proof follows immediately from theorems 2.1.1, 2.4.5, 2.4.6, and 2.4.7.

Note that all admissible designs are unique according to theorem 2.4.6.

Note also that the designs ξ and $\bar{\xi}$ given in example 2.4.1 are inadmissible and admissible (respectively) for quadratic regression.

In fact, both designs achieve $\mu = (1, 0, 1/3, 0)'$ while $\mu_4 = 1/3$ and $\tilde{\mu}_4 = 7/12$.

The following example provides an instance in which theorem 2.4.8 does not hold for $G_1 = (\alpha_1, \beta_1) \subset (a,b)$.

Example 2.4.2: Let $[a,b] = [-2,5]$, $G_1 = (-1,1)$, $p_1 = 1/5$, and $n = 2$. Consider first the measure $\xi = \frac{1}{5}(\delta_{-2} + \delta_{-1} + \delta_0 + \delta_1 + \delta_5)$. Then $\xi(G_1) = p_1$ so that $I(\xi) = I_0(\text{Support}(\xi) - \partial G_1) = 4$ and ξ is upper principal. However, ξ is not admissible for quadratic regression. If if were, there would exist a fourth degree polynomial $Q(x)$ such that $Q(x) \geq 1$ for $x \in [-2,-1] \cup [1,5]$ and $Q(x) \geq 0$ for $x \in (-1,1)$ with equality on Support(ξ). Now solving the linear system corresponding to the conditions $Q(-2) = Q(-1) = Q(1) = 1$ and $Q(0) = Q'(0) = 0$ yields $Q(x) = x^2(5-x^2)/4$. But then $Q(x) < 1$ for $1 < x \leq 5$. Since the required polynomial $Q(x)$ does not exist, ξ must be inadmissible.

This example leads to the consideration of problem 2.2 in the present context. As in section 2.2, it may happen that the unrestricted solution $\bar{\xi}_0$ satisfies the restriction $\bar{\xi}_0(G_1) \leq p_1$. Otherwise, further analysis is needed. Methods to solve problem 2.2 when $\bar{\xi}_0(G_1) > p_1$ will now be indicated for the cases $n = 1$ and 2.

For linear regression, Support($\bar{\xi}_0$) = $\{a,b\}$. Thus if $G_1 = (\alpha_1, \beta_1) \subset (a,b)$, then $\bar{\xi}_0 \in \Xi$ and problem 2.2 is trivially solved. Assume now that a or $b \in G_1$; in particular, assume that $G_1 = (\alpha_1, b]$. If $\bar{\xi}_0(G_1) > p_1$, then it must be true that $I(\bar{\xi}) \leq 2 < I_0(\text{Support}(\bar{\xi}))$ and hence Support($\bar{\xi}$) = $\{a, b, \alpha_1\}$. Also, $\bar{\xi}(\{b\}) = p_1$, $\bar{\xi}(\{\alpha_1\}) = [\mu_1 - a - p_1(b-a)]/(\alpha_1 - a)$, and

$\bar{\xi}(\{a\}) = 1 - p_1 - \bar{\xi}(\{\alpha_1\})$ must hold in order to satisfy $\int_a^b u(x)d\bar{\xi}(x) = \mu$.

For quadratic regression, Support$(\bar{\xi}_0) = \{a, b, x_1\}$ where $a < x_1 < b$ and $\bar{\xi}_0(G_1) > p_1$. Consider first the case that a or $b \in G_1$. In particular, $G_1 = (\alpha_1, b]$ without loss of generality. The first possibility is that $\bar{\xi} = \rho_0\delta_a + \rho_1\delta_y + \rho_2\delta_{\alpha_1} + p_1\delta_b$, where $a < y < \alpha_1$. Then

$$\mu = \int_a^b u(x)d\bar{\xi}(x) = \rho_0 u(a) + \rho_1 u(y) + \rho_2 u(\alpha_1) + p_1 u(b) \qquad (2.4.4)$$

implies that $0 = |u(a), u(\alpha_1), \mu - p_1 u(b), u(y)| = P(y)$. Thus y must be a root of the polynomial P and the mass distribution must solve (2.4.4). The second possibility is that $\bar{\xi} = \rho_0\delta_a + \rho_1\delta_y + \rho_2\delta_{\alpha_1} + \rho_3\delta_b$, where $\alpha_1 < y < b$. Then

$\mu = \int_a^b u(x)d\bar{\xi}(x) = \rho_0 u(a) + \rho_1 u(y) + \rho_2 u(\alpha_1) + \rho_3 u(b)$ and $\rho_1 + \rho_3 = p_1$

imply that the mass distribution must solve

$$\begin{bmatrix} u(a) & u(y) & u(\alpha_1) & u(b) \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_0 \\ \vdots \\ \rho_3 \end{bmatrix} = \begin{bmatrix} \mu \\ p_1 \end{bmatrix} \qquad (2.4.5)$$

and hence $0 = \begin{vmatrix} u(a) & u(\alpha_1) & u(b) & \mu & u(y) \\ 0 & 0 & 1 & p_1 & 1 \end{vmatrix} = P(y)$ must hold.

Thus y must be a root of the polynomial P and the mass distribution must solve (2.4.5).

Example 2.4.1 will serve to illustrate the first of these two possibilities. Recall that $[a,b] = [-2,1]$, that $G_1 = (0,1]$, that $p_1 = \frac{1}{5}$, and that $\xi = \frac{1}{6}\delta_{-1} + \frac{2}{3}\delta_0 + \frac{1}{6}\delta_1$. The standard calculations for $\bar{\xi}_0$ yield $\bar{\xi}_0 = \frac{4}{162}\delta_{-2} + \frac{125}{162}\delta_{-1/5} + \frac{33}{162}\delta_1$. Thus $\bar{\xi}_0(G_1) = 11/54 > p_1$ and $\bar{\xi}_0 \notin \Xi$. Now the form of $\bar{\xi}_0$ suggests that Support$(\bar{\xi}) = \{-2, y, 0, 1\}$

where $-2 < y < 0$ and $\bar{\xi}_0(\{1\}) = p_1$. Then the roots of $P(y) = |u(-2), u(0), u - \frac{1}{5}u(1), u(y)|$ are 0, -2, and $y = -1/4$. Solution of (2.4.4) yields the design $\bar{\xi}$ given in example 2.4.1. Recall that $\bar{\xi}$ is admissible and hence solves problem 2.2 for this example. It may be worth noting the result if the possibility that $0 < y < 1$ is considered for this example. It turns out that the only real root of $P(y) = \begin{vmatrix} u(-2) & u(0) & u(1) & \mu & u(y) \\ 0 & 0 & 1 & 1/5 & 1 \end{vmatrix}$ is $y = 1$. Thus the method fails when this incorrect form of Support($\bar{\xi}$) is proposed.

Now consider the case that $G_1 = (\alpha_1, \beta_1) \subset (a,b)$. The first possibility is that $\bar{\xi} = \rho_0\delta_a + p_1\delta_y + \rho_1\delta_{\alpha_1} + \rho_2\delta_b$, where $\alpha_1 < y < \beta_1$. Then

$$\mu = \int_a^b u(x)d\bar{\xi}(x) = \rho_0 u(a) + p_1 u(y) + \rho_1 u(\alpha_1) + \rho_2 u(b) \quad (2.4.6)$$

implies that $0 = |u(a), u(\alpha_1), u(b), \mu - p_1 u(y)| = P(y)$. Thus y must be a root of the polynomial P and the mass distribution must solve (2.4.6). The second possibility is that α_1 is replaced by β_1 in the preceeding. The final possibility is that $\bar{\xi} = \rho_0\delta_a + \rho_1\delta_{\alpha_1} + p_1\delta_y + \rho_2\delta_{\beta_1} + \rho_3\delta_b$. If $\bar{\xi}$ is to have this form, there must exist a quartic polynomial $Q(x)$ such that $Q(x) \geq 1$ for $x \in [a,\alpha_1] \cup [\beta_1,b]$ and $Q(x) \geq 0$ for $x \in (\alpha_1,\beta_1)$ with equality at the support points. This requirement implies that $0 = \frac{1}{y-a} + \frac{1}{y-\alpha_1} + \frac{1}{y-\beta_1} + \frac{1}{y-b} = g(y)$ which has one zero in the interval (α_1,β_1). Once y is determined, the mass distribution may be obtained by solving the linear system

$$\mu = \int_a^b u(x)d\bar{\xi}(x) = \rho_0 u(a) + \rho_1 u(\alpha_1) + p_1 u(y) + \rho_2 u(\beta_1) + \rho_3 u(b).$$

Example 2.4.2 will now be used to illustrate this approach. Recall that $[a,b] = [-2,5]$, that $G_1 = (-1,1)$, that $p_1 = 1/5$, and that $\xi = \frac{1}{5}(\delta_{-2} + \delta_{-1} + \delta_0 + \delta_1 + \delta_5)$ is inadmissible. Note that $\mu = (1, 3/5, 31/5, 117/5)'$. The standard calculations for $\bar{\xi}_0$ yield approximately $\bar{\xi}_0 = .2953\delta_{-2} + .4875\delta_{-3/14} + .2172\delta_5$. Thus $\bar{\xi}_0(G_1) > p_1$ and $\bar{\xi}_0 \notin \Xi$. To consider the first possibility, note that $P(y) = |u(-2), u(-1), u(5), \mu - \frac{1}{5}u(y)|$ has no roots $y \in (-1,1)$. Thus the possibility that $\bar{\xi} = \rho_0\delta_{-2} + p_1\delta_y + \rho_1\delta_{-1} + \rho_2\delta_5$ is ruled out. Likewise $\bar{\xi} = \rho_0\delta_{-2} + p_1\delta_y + \rho_1\delta_1 + \rho_2\delta_5$ may be ruled out. That leaves only the possibility that $\bar{\xi} = \rho_0\delta_{-2} + \rho_1\delta_{-1} + p_1\delta_y + \rho_2\delta_1 + \rho_3\delta_5$. Recall that y must be a zero of $g(y) = \frac{1}{y+2} + \frac{1}{y+1} + \frac{1}{y-1} + \frac{1}{y-5}$. Approximately, $y = .1299$. Then solving for the mass distribution yields $\rho_0 \approx .1981$, $\rho_1 \approx .2161$, $\rho_2 \approx .1857$, and $\rho_3 \approx .2001$. The fact that the required polynomial $Q(x)$ exists also implies that $\bar{\xi}$ is admissible. Note finally that $\mu_4 = 128.6 < 128.6340 \approx \bar{\mu}_4$. (Thus $\bar{\xi}$ is only a slight improvement over ξ in this example.)

The preceding discussion provides means to solve problem 2.2 for linear and quadratic regression when $s = 1$. For $n \geq 3$, the solution is more difficult to obtain.

Sufficient conditions for the solution of problem 2.1 will now be sketched for the case that $s = 2$, that $G_1 = [a,\beta_1)$, and that $G_2 = (\alpha_2,b]$, where $\beta_1 < \alpha_2$. The results obtained will closely parallel those for the case $s = 1$ when a or $b \in G_1$.

The following lemma characterizes the dual cone $\mathcal{P}_{m+1}$ when $p_1 + p_2 \leq 1$. It holds regardless of the forms of G_1 and G_2. This lemma is analogous to lemma 2.4.3.

Lemma 2.4.5: Let s = 2 and $p_1 + p_2 \leq 1$. Also let $\pi \in \mathbf{R}^{m+1}$ and let $\lambda_i = \inf_{G_i} \pi' u(x_i)$ for i = 0, 1, 2. Then $\pi \in \mathcal{P}_{m+1}$ if and only if the inequalities

i. $\lambda_0 \geq 0$,

ii. $p_i\lambda_i + (1-p_i)\lambda_0 \geq 0$ for i = 1,2, and

iii. $p_1\lambda_1 + p_2\lambda_2 + (1-p_1-p_2)\lambda_0 \geq 0$

all hold.

Proof: The proof closely resembles that of lemma 2.4.3. Necessity is easily proved by considering measures δ_x, measures $p_i\delta_{x_i} + (1-p_i)\delta_{x_0}$ for i = 1, 2, and measures $p_1\delta_{x_1} + p_2\delta_{x_2} + (1 - p_1 - p_2)\delta_{x_0}$, where the points $x_i \in G_i$ for i = 0, 1, 2.

To prove sufficiency, first note that if $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, then $\pi'\mu = \int_a^b \pi' u(x) d\sigma(x) \geq 0$ for any $\sigma \in \Sigma$. Consider next the case that $\lambda_1 < 0$ and $\lambda_2 \geq 0$. If $\lambda_2 \leq \lambda_0$ also holds, then $\pi'\mu \geq \gamma[p_1\lambda_1 + p_2\lambda_2 + (1-p_1-p_2)\lambda_0] \geq 0$ for any $\sigma = \gamma\xi \in \Sigma$. If $\lambda_0 < \lambda_2$ also holds, then $\pi'\mu \geq \gamma[p_1\lambda_1 + (1-p_1)\lambda_0] \geq 0$. For the case that $\lambda_1 \geq 0$ and $\lambda_2 < 0$, a similar argument shows that $\pi'\mu \geq 0$. Finally, the case $\lambda_1 < 0$ and $\lambda_2 < 0$ yields $\pi'\mu \geq \gamma[p_1\lambda_1 + p_2\lambda_2 + (1-p_1-p_2)\lambda_0] \geq 0$. Thus $\pi'\mu \geq 0$ for any $\mu = \int_a^b u(x) d\sigma(x) \in \mathcal{M}_{m+1}$. That is, $\pi \in \mathcal{P}_{m+1}$ and the proof is complete.

The following theorem provides a converse to theorem 2.4.1 in the case that $s = 1$, that $G_1 = [a,\beta_1)$, that $G_2 = (\alpha_2,b]$, that $\xi(G_1) \leq p_1$, and that $\xi(G_2) \leq p_2$. Thus it is analogous to theorem 2.4.5.

Theorem 2.4.9: Let $s = 2$, let $G_1 = [a,\beta_1)$, and let $G_2 = (\alpha_2,b]$. If $I(\xi) \leq m$, then $\mu = \int_a^b u(x)d\xi(x) \in \partial M_{m+1}$.

Proof: As argued for theorem 2.4.5, the proof need only be carried out in the case that $I_0(\text{Support}(\xi)) > m = I(\xi) = m$. If $I_0(\text{Support}(\xi) - \partial G_1) \leq m$ also holds, then theorem 2.4.5 implies that $\mu \in \partial M_{m+1}$. The same is true if $I_0(\text{Support}(\xi) - \partial G_2) \leq m$. Thus the theorem need only be proven for the case that $I(\xi) = m$, that $I_0(\text{Support}(\xi) - \partial G_1) > m$, and that $I_0(\text{Support}(\xi) - \partial G_2) > m$.

By inspection, there are precisely three possible ways for this case to arise. All of them require $\xi(G_i) = p_i$ for $i = 1, 2$ and $p_1 + p_2 < 1$. The first possibility, $m = 2k$ and Support(ξ) consists of k points from $(a,b) - \{\beta_1,\alpha_2\}$ as well as the points β_1 and α_2. For the second, $m = 2k$ and Support(ξ) consists of $k - 1$ points from $(a,b) - \{\beta_1,\alpha_2\}$, the points β_1 and α_2, and both endpoints. For the final possibility, $m = 2k + 1$ and Support(ξ) consists of k points from $(a,b) - \{\beta_1,\alpha_2\}$, the points β_1 and α_2, and exactly one endpoint.

For each of these cases, it will be argued that there exists a polynomial $P(x) = \pi'u(x)$ with the following four properties:

i. $P(x) \geq \lambda_i$ on G_i for $i = 1,2$,

ii. $P(x) \geq 1$ on $[a,b] - \bigcup_{i=1}^{2} G_i$ with equality precisely on Support(ξ),

iii. $\lambda_i \leq 1$ for $i = 1,2$, and

iv. $p_1\lambda_1 + p_2\lambda_2 + (1 - p_1 - p_2) = 0$.

This will imply that $\pi \in \mathcal{P}_{m+1}$ according to lemma 2.4.5 and $\pi'\mu = p_1\lambda_1 + p_2\lambda_2 + (1 - p_1 - p_2) = 0$ so that $\mu \in \partial M_{m+1}$.

The existence of P(x) will be demonstrated only for the second possible case: $m = 2k$ and Support(ξ) = $\{a, b, \beta_1, \alpha_2, x_1, \ldots, x_{k-1}\}$. The other two cases are readily proven with only minor modifications to the proof of this case.

First let $\{x_1, \ldots, x_r\} \subset (a, \beta_1)$, let $\{x_{r+1}, \ldots, x_t\} \subset (\beta_1, \alpha_2)$, and let $\{x_{t+1}, \ldots, x_{k-1}\} \subset (\alpha_2, b)$. Then the polynomial $P(x) = \pi' u(x)$ must be such that π, λ_1, and λ_2 solve the linear system

$$\begin{bmatrix} u_0(a) & \cdots & u_m(a) & -1 & 0 \\ u_0(x_1) & \cdots & u_m(x_1) & -1 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ u_0(x_r) & \cdots & u_m(x_r) & -1 & 0 \\ u_0(\beta_1) & \cdots & u_m(\beta_1) & 0 & 0 \\ u_0(x_{r+1}) & \cdots & u_m(x_{r+1}) & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ u_0(x_t) & \cdots & u_m(x_t) & 0 & 0 \\ u_0(\alpha_2) & \cdots & u_m(\alpha_2) & 0 & 0 \\ u_0(x_{t+1}) & \cdots & u_m(x_{t+1}) & 0 & -1 \\ \vdots & & \vdots & \vdots & \vdots \\ u_0(x_{k-1}) & \cdots & u_m(x_{k-1}) & 0 & -1 \\ u_0(b) & \cdots & u_m(b) & 0 & -1 \\ u_0'(x_1) & \cdots & u_m'(x_1) & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ u_0'(x_{k-1}) & \cdots & u_m'(x_{k-1}) & 0 & 0 \\ 0 & \cdots & 0 & p_1 & p_2 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \vdots \\ \pi_m \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ p_1 + p_2 - 1 \end{bmatrix}. \quad (2.4.7)$$

It will be shown that the coefficient matrix U in this linear system is non-singular. In fact, elementary manipulations on the determinant $|U|$ yield

$$|U| = \begin{vmatrix} u(a), & u(x_1), & \ldots, & u(x_r) & u(\beta_1), & u(x_{r+1}), & \ldots, & u(x_t), & u(\alpha_2), & u(x_{t+1}), & \ldots, & u(x_{k-1}), & u(b), & u'(x_1), & \ldots, & u'(x_{k-1}) \\ -p_2, & -p_2, & \ldots, & -p_2, & 0, & 0, & \ldots, & 0, & 0, & p_1, & \ldots, & p_1, & p_1, & 0, & \ldots, & 0 \end{vmatrix} .$$

If this determinant were to vanish, then there would exist $w \in \mathbb{R}^{m+1}$ such that

$$\begin{bmatrix} u_0(a) & \ldots & u_m(a) \\ u_0(x_1) & \ldots & u_m(x_1) \\ \vdots & & \vdots \\ u_0(x_r) & \ldots & u_m(x_r) \\ u_0(\beta_1) & \ldots & u_m(\beta_1) \\ u_0(x_{r+1}) & \ldots & u_m(x_{r+1}) \\ \vdots & & \vdots \\ u_0(x_t) & \ldots & u_m(x_t) \\ u_0(\alpha_2) & \ldots & u_m(\alpha_2) \\ u_0(x_{t+1}) & \ldots & u_m(x_{t+1}) \\ \vdots & & \vdots \\ u_0(x_{k-1}) & \ldots & u_m(x_{k-1}) \\ u_0(b) & \ldots & u_m(b) \\ u_0'(x_1) & \ldots & u_m'(x_1) \\ \vdots & & \vdots \\ u_0'(x_{k-1}) & \ldots & u_m'(x_{k-1}) \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} -P_2 \\ -P_2 \\ \vdots \\ -P_2 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ P_1 \\ \vdots \\ P_1 \\ P_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

It will now be argued that no such w can exist. If it did, then consider the polynomial $Q(x) = w'u(x)$. It has the properties that $Q(a) = Q(x_1) = \ldots = Q(x_r) = -P_2$, that $Q(\beta_1) = Q(x_{r+1}) = \ldots = Q(x_t) = Q(\alpha_2) = 0$, that $Q(x_{t+1}) = \ldots = Q(x_{k-1}) = Q(b) = p_1$, and that $Q'(x_1) = \ldots = Q'(x_{k-1}) = 0$. Hence $Q'(x)$ must vanish at least once between every adjacent pair of support points with the possible exception of the pairs (x_r, β_1) and (α_2, x_{t+1}). In the case that $Q'(\beta_1) < 0$, the facts that $Q(x_r) = -p_1$ and $Q(\beta_1) = 0$ imply that $Q'(x)$ must also vanish somewhere in (x_r, β_1). If $Q'(\beta_1) > 0$, then $Q'(x)$ must have an additional root somewhere in (β_1, x_{t+1}) or else $Q(x_{t+1}) = p_1 > 0$ is impossible. Thus, in any case, $Q'(x)$ has at least $(k-1) + k + 1 = m$ roots. Since $Q'(x)$ is a non-trivial polynomial of degree $m - 1$, this contradiction implies that no such $w \in \mathbb{R}^{m+1}$ can exist. Hence $|U| \neq 0$ and the polynomial $P(x) = \pi'u(x)$ does exist.

The fact that a solution to (2.4.7) exists yields $p_1\lambda_1 + p_2\lambda_2 + (1 - p_1 - p_2) = 0$. Since $p_1 + p_2 < 1$, it must be true that at least one of the λ_i's is negative. Without loss of generality, it may be assumed that $\lambda_1 < 0$. Suppose that $\lambda_2 > 1$. Then $P(x)$ would have to have the same properties as $Q(x)$ which has been shown to be impossible. Thus $\lambda_i \leq 1$ for $i = 1, 2$.

Now it must be true that $P(x) \geq \lambda_i$ on G_i for $i = 1, 2$ or else $P'(x)$ has at least m roots. Likewise, $P(x) \geq 1$ on $[a,b] - \bigcup_{i=1}^{2} G_i$ or else $P'(x)$ has at least $m + 1$ roots. Furthermore, $P(x)$ cannot attain equality at a non-support point or else $P'(x)$ has at least $m + 1$ roots. Thus $P(x)$ satisfies all the required properties and the proof of the theorem is complete.

The following theorem establishes the uniqueness of any boundary moment point. It is exactly analogous to theorem 2.4.6.

Theorem 2.4.10: Let $s = 1$, let $G_1 = [a,\beta_1)$, and let $G_2 = (\alpha_2,b]$. Then $\mu \in \partial M_{m+1}$ if and only if it has a unique representation $\xi \in \Xi$.

Proof: The proof of the sufficiency proceeds exactly as in theorem 2.4.6.

To prove the necessity, first note that theorem 2.4.6 implies that ξ is unique unless $\xi(G_i) = p_i$ for $i = 1, 2$ and $p_1 + p_2 < 1$. Now consider this case. There exists $\pi \in \mathcal{P}_{m+1} - \{0\}$ such that $\pi'\mu = 0$. Let λ_0, λ_1, and λ_2 be as in lemma 2.4.1(ii). Then this lemma implies that

$$\text{Support}(\xi) \subset \bigcup_{i=0}^{2} \{x_i \in G_i \mid \pi'u(x_i) = \lambda_i\} \qquad (2.4.8)$$

for any measure which represents μ. Let $A = \{x_0,\ldots,x_k\}$ denote the right hand side of (2.4.8). By now familiar derivative arguments, k is maximized if A consists of the points β_1, α_2, a, b, and $[(m + 1 - 4)/2]$ points from $(a,b) - \{\beta_1,\alpha_2\}$. Thus $k \leq 3 + [(m-3)/2] \leq m$ unless $m = 1$. In the special case that $m = 1$, the number k is seen to be 1 or less. The remainder of the proof follows that of theorem 2.4.6 exactly. Thus the theorem is proved.

Note that at this point theorems 2.4.9 and 2.4.10 yield the conclusion that if $I(\xi) \leq m$, then ξ solves problem 2.1 for the special case at hand. The following theorem provides a sufficient condition for the solution of problem 2.1 when $I(\xi) > m$. It is analogous to theorem 2.4.7.

<u>Theorem 2.4.11</u>: Let $s = 2$, let $G_1 = [a,\beta_1)$, and let $G_2 = (\alpha_2,b]$. Also let $\mu \in \text{Interior}(M_{m+1})$. If ξ is upper (lower) principal, then it attains the maximum (minimum) value of μ_{m+1} among all measures from Ξ which represent μ.

<u>Proof</u>: The proof follows that of theorem 2.4.7 verbatim.

Note that the existence and uniqueness of the upper and lower principal representations have now been established. Also, problem 2.1 has been solved for $s = 2$ when $G_1 = [a,\beta_1)$ and $G_2 = (\alpha_2,b]$. The following theorem applies these results to the admissibility problem. It serves as a converse to theorem 2.4.4 and an analogue to theorem 2.4.8.

<u>Theorem 2.4.12</u>: Let $s = 2$, let $G_1 = [a,\beta_1)$, and let $G_2 = (\alpha_2,b]$. If either a. $I(\xi) < 2n$ or b. $I(\xi) = 2n$ and $b \in \text{Support}(\xi)$, then $\xi \in \Xi$ is admissible.

<u>Proof</u>: The proof is an immediate consequence of theorems 2.1.1, 2.4.9, 2.4.10, and 2.4.11.

Note that all admissible designs are unique according to theorem 2.4.10.

Problem 2.2 will now be briefly considered in the present context. As in section 2.2, it may happen that the unrestricted solution $\bar{\xi}_0 \in \Xi$. The next level would be to solve problem 2.2 with only one of the two restrictions imposed. Let $\bar{\xi}_i \in \Xi(\{i\})$ be the solution under only the restriction that $\xi(G_i) \leq p_i$. If $\bar{\xi}_i \in \Xi$, then it is the solution to problem 2.2. Otherwise,

$\text{Support}(\xi) = \{a, b, \beta_1, \alpha_2, x_1,\ldots,x_{n-1}\}$ and further analysis is needed. Methods to solve problem 2.2 will now be indicated for $n = 1$ and 2.

For linear regression, the designs $\bar{\xi}_0$, $\bar{\xi}_1$, and $\bar{\xi}_2$ have already been exhibited. If all of them do not belong to Ξ, then $\text{Support}(\bar{\xi}) = \{a, b, \beta_1, \alpha_2\}$. Also, $\bar{\xi}(\{a\}) = p_1$, $\bar{\xi}(\{b\}) = p_2$, $\bar{\xi}(\{\beta_1\}) = (\mu_1 - \alpha_2 - p_1 a - p_2 b)/(\beta_1 - \alpha_2)$, and $\bar{\xi}(\{\alpha_2\}) = 1 - p_1 - p_2 - \bar{\xi}(\{\beta_1\})$ must hold in order to satisfy $\int_a^b u(x)d\bar{\xi}(x) = \mu$.

For quadratic regression, methods to determine $\bar{\xi}_0$, $\bar{\xi}_1$, and $\bar{\xi}_2$ have already been indicated. If all of them do not belong to Ξ, then $\text{Support}(\bar{\xi}) = \{a, b, \beta_1, \alpha_2, y\}$. If $a < y < \beta_1$ is to hold, then y must be a root of $P(y) = \begin{vmatrix} u(a), & u(\beta_1), & u(\alpha_2), & \mu - p_2 u(b), & u(y) \\ 1, & 0, & 0, & 0, & 1 \end{vmatrix}$.

If $\beta_1 < y < \alpha_2$ is to hold, then y must be a root of $P(y) = |u(\beta_1), u(\alpha_2), \mu - p_1 u(a) - p_2 u(b), u(y)|$. If $\alpha_2 < y < b$ is to hold, then y must be a root of

$P(y) = \begin{vmatrix} u(\beta_1), & u(\alpha_2), & u(b), & \mu - p_1 u(a), & u(y) \\ 0, & 0, & 1, & 0, & 1 \end{vmatrix}$. In any of these cases, the mass distribution of $\bar{\xi}$ is readily determined from the linear system $\mu = \int_a^b u(x)d\bar{\xi}(x)$ along with the constraints $\bar{\xi}(G_i) = p_i$ for $i = 1, 2$.

These methods may be illustrated by a modified version of example 2.4.1 with $s = 2$, with $G_1 = [-2,-1)$, with $G_2 = (0,1]$, with $p_1 = .02$, and with $p_2 = 1/5$. Recall that

$\bar{\xi}_0 = \frac{4}{162}\delta_{-2} + \frac{125}{162}\delta_{-1/5} + \frac{33}{162}\delta_1 \notin \Xi$ and that

$\bar{\xi}_2 = \frac{5}{210}\delta_{-2} + \frac{128}{210}\delta_{-1/4} + \frac{35}{210}\delta_0 + \frac{1}{5}\delta_1 \notin \Xi$ since $\bar{\xi}_2(G_1) > p_1$. The same methods yield approximately

$\bar{\xi}_1 = .02\delta_{-2} + .1581\delta_{-1} + .7471\delta_{-18/109} + .0748\delta_1$. Thus $\bar{\xi}_1(G_2) = .0748 < p_2$ implies that $\bar{\xi} = \bar{\xi}_1 \in \Xi$. Furthermore $I(\bar{\xi}_1) = 4$ and $\bar{\xi}_1$ is upper so theorem 2.4.12 implies that $\bar{\xi}_1$ solves problem 2.2 for this example. Note also that $\int_a^b x^4 d\bar{\xi}_1(x) \approx .5535 > 1/3 = \mu_4$.

CHAPTER III

D-OPTIMALITY

3.1 General Results

While the concept of admissibility induces a partial ordering on the set of information matrices, a specific optimal design criterion yields an essentially complete ordering on this set. Such a criterion is often specified in terms of a real-valued function Φ which is defined on the set of non-negative definite matrices. Thus a Φ-optimal design ξ would achieve a minimum value for $\Phi(M^{-1}(\xi))$.

An often-used criterion is that of "D-optimality". A D-optimal design should attain the minimum value of the determinant $|M^{-1}(\xi)|$. Thus a rationale for D-optimality might be to minimize the generalized volume of a classical confidence region for $\theta \in \mathbb{R}^{n+1}$. The following definition makes the criterion explicit.

Definition 3.1.1: A design $\xi \in \Xi$ is D-optimal if and only if it attains the minimum possible value of $|M^{-1}(\xi)|$. Equivalently, it maximizes $|M(\xi)|$.

Of course for non-triviality it must be assumed that there exists at least one design $\xi \in \Xi$ such that $M(\xi)$ is non-singular.

The question of D-optimality has been addressed by many authors in the unrestricted design setting where $\Xi = \Xi_0$, the set of all probability measures on $\mathcal{X}$. Guest ('58) and Hoel ('58) have considered the problem for polynomial regression on an interval. Kiefer and Wolfowitz ('60) have provided a general equivalence theorem for D-optimality. More recently, Cook and Thibodeau ('80) have treated D-optimality in the case that Ξ arises from the marginal restriction problem. Also, Wynn ('77) has considered D-optimality in the (φ,ψ)-problem setting with $\varphi = 0$.

It should be noted that the relationship between admissibility and D-optimality for polynomial regression of degree n is not as helpful in the restricted design setting as it was in the unrestricted. For $\Xi = \Xi_0$, an admissible design can have no more than n + 1 support points and a D-optimal design must have at least that many. As a consequence, $|M(\xi)|$ factors into the square of a Vandermonde determinant involving n + 1 points and the product of the corresponding masses. This result is key to the works of Guest ('58) and Hoel ('58). The developments of sections 2.2, 2.3, and 2.4 make it clear that an admissible restricted design for polynomial regression may have more than n + 1 support points. Subsequent examples will include situations where a D-optimal design's support does include more than n + 1 points.

The main intent of the present section is to adapt the Kiefer and Wolfowitz equivalence result to the problem of determining a D-optimal restricted design. The later sections will provide examples of D-optimal designs under various restrictions. These examples should serve to indicate how D-optimality depends on the particular form of restriction.

The equivalence result of theorem 3.1.1 will involve the "variance function" of a design.

Definition 3.1.2: The variance function of a design $\xi \in \Xi$ with non-singular information matrix is $d(x,\xi) = f(x)'M^{-1}(\xi)f(x)$.

The following lemmas will be used to prove theorem 3.1.1. Both they and the theorem follow Federov ('72).

Lemma 3.1.1: i. $\int_{\mathcal{X}} d(x,\xi)d\xi(x) = n + 1$ for any $\xi \in \Xi$.

ii. $\min_{\tilde{\xi}} \max_{\xi} \int d(x,\tilde{\xi})d\xi(x) \geq n + 1$.

Proof: i. The proof of the first part involves only an elementary calculation and may be found in Federov ('72).

ii. For any $\tilde{\xi} \in \Xi$, $\max_{\xi} \int_{\mathcal{X}} d(x,\tilde{\xi})d\xi(x) \geq \int_{\mathcal{X}} d(x,\tilde{\xi})d\tilde{\xi}(x) = n + 1$. Hence $\min_{\tilde{\xi}} \max_{\xi} \int d(x,\tilde{\xi})d\xi(x) \geq n + 1$.

Lemma 3.1.2: i. $\log|M(\xi)|$ is strictly concave in ξ.

ii. $\frac{d}{d\gamma}\{\log|(1-\gamma)M(\xi_0) + \gamma M(\xi)|\}_{\gamma=0} = \text{tr } M^{-1}(\xi_0)M(\xi) - (n+1)$ for any designs ξ_0, $\xi \in \Xi$.

Proof: A proof of both parts is given in Federov ('72).

Theorem 3.1.1: $|M(\xi_0)| = \max_{\xi \in \Xi} |M(\xi)|$ if and only if

$$\max_{\xi \in \Xi} \int_{\mathcal{X}} d(x,\xi_0)d\xi(x) = \min_{\tilde{\xi} \in \Xi} \max_{\xi \in \Xi} \int_{\mathcal{X}} d(x,\tilde{\xi})d\xi(x) = n + 1.$$

Proof: First note that $|M(\xi_0)| = \max_{\xi} |M(\xi)|$ if and only if $\log|M(\xi_0)| = \max_{\xi} \log|M(\xi)|$.

If ξ_0 is D-optimal, then $0 \geq \frac{d}{d\gamma}\{\log|(1-\gamma)M(\xi_0) + \gamma M(\xi)|\}_{\gamma=0}$ for any $\xi \in \Xi$. According to lemma 3.1.2(ii) this inequality becomes

$$n + 1 \geq \text{tr } M^{-1}(\xi_0)M(\xi) = \text{tr } M^{-1}(\xi_0) \int_{\mathcal{X}} f(x)f(x)'d\xi(x) = \int_{\mathcal{X}} d(x,\xi_0)d\xi(x)$$

for any $\xi \in \Xi$. Then application of lémma 3.1.1(ii) to this inequality yields

$$\min_{\tilde{\xi}} \max_{\xi} \int_{\mathcal{X}} d(x,\tilde{\xi})d\xi(x) \geq n + 1 \geq \max_{\xi} \int_{\mathcal{X}} d(x,\xi_0)d\xi(x)$$

$$\geq \min_{\tilde{\xi}} \max_{\xi} \int_{\mathcal{X}} d(x,\tilde{\xi})d\xi(x).$$

Thus necessity is proved.

To prove sufficiency, suppose that ξ_0 is not D-optimal. Then there must exist $\xi' \in \Xi$ such that $|M(\xi')| > |M(\xi)|$. Then lemma 3.1.2(i) implies that $0 < \frac{d}{d\gamma}\{\log|(1-\gamma)M(\xi_0) + \gamma M(\xi')|\}_{\gamma=0}$. That is, $n + 1 < \int_{\mathcal{X}} d(x,\xi_0)d\xi'(x) \leq \max_{\xi} \int_{\mathcal{X}} d(x,\xi_0)d\xi(x)$. Thus the conclusion of the theorem cannot hold if ξ_0 is not D-optimal so that the theorem is proved.

It may be noted that an analogous theorem is readily proven for other optimality criteria which require that $\Phi(M^{-1}(\xi))$ be minimized. In the more general case, the variance function should be replaced by $d(x,\xi) = -f(x)'\text{grad}\Phi(M(\xi))f(x)$. Note also that if $\Xi = \Xi_0$, then $\max_{\xi \in \Xi} \int_{\mathcal{X}} d(x,\tilde{\xi})d\xi(x) = \max_{\mathcal{X}} d(x,\tilde{\xi})$ and theorem 3.1.1 is precisely the Kiefer and Wolfowitz result.

3.2 The E-Problem

As already remarked, the E-problem may be considered as an unrestricted design setting with $\mathcal{X}' = E$. Thus the Kiefer and Wolfowitz equivalence result is obtained.

Corollary 3.2.1: $|M(\xi_0)| = \max_{\xi \in \Xi} |M(\xi)|$ if and only if

$\max_{E} d(x,\xi_0) = \min_{\tilde{\xi} \in \Xi} \max_{E} d(x,\tilde{\xi}) = n + 1$.

Proof: For the E-problem, $\max_{\xi \in \Xi} \int_E d(x,\tilde{\xi})d\xi(x) = \max_E d(x,\tilde{\xi})$ for any $\tilde{\xi} \in \Xi$. Thus the corollary follows immediately from theorem 3.1.1.

Examples will now be given to show how the E-problem restriction can affect known D-optimality results.

Consider first the setting of linear regression on $\mathcal{X} = [-1,1]$. The unrestricted D-optimal design is well known to be $\xi_0 = \frac{1}{2}(\delta_{-1} + \delta_1)$. By definition of the E-problem, $\{-1,1\} \subset E$ so that $\xi_0 \in \Xi$ and is D-optimal no matter what restrictions govern the allocation of observations within $(-1,1)$.

The problem of quadratic regression on $\mathcal{X} = [-1,1]$ gives more interesting results. The following derivation of the D-optimal design will be summarized by lemma 3.2.1.

The solution for the D-optimal design will first be obtained in the case that $E = [-1,\alpha] \cup [\beta,1]$, where $-\beta \le \alpha \le 0$. The solution for general E will follow readily. Now if $|M(\xi_0)| > 0$ is to hold, then Support(ξ_0) must include at least three points. In order for ξ_0 to be admissible, theorem 2.2.6(ii) requires that its support consists of no more than three points unless Support(ξ_0) $= \{-1, \alpha, \beta, 1\}$.

In the case that $\xi_0 = \rho_0\delta_{x_0} + \rho_1\delta_{x_1} + \rho_2\delta_{x_2}$, where $-1 \le x_0 < x_1 < x_2 \le 1$, the determinant

$$|M(\xi_0)| = \rho_0\rho_1\rho_2 \begin{vmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ x_0^2 & x_1^2 & x_2^2 \end{vmatrix}^2 = \rho_0\rho_1\rho_2(x_1-x_0)^2(x_2-x_0)^2(x_2-x_1)^2.$$

Thus $\rho_0 = \rho_1 = \rho_2 = 1/3$ and $x_0 = -x_2 = -1$ are required for $|M(\xi_0)|$ to be maximized. It remains to maximize $(1-x_1^2)$ by choice of $x_1 \in (-1,1) \cap E = (-1,\alpha] \cup [\beta,1)$, where $-\beta \leq \alpha \leq 0$. Thus $x_1 = \alpha$ if ξ_0 is to be D-optimal.

It will now be verified that $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_\alpha + \delta_1)$ is D-optimal under a certain condition on α and β. It will be seen that the resultant condition quantifies the intuitive idea that β should be much further from zero than α. As derived by Guest ('58),

$$\frac{1}{3}\, d(x,\xi_0) = L_0^2(x) + L_1^2(x) + L_2^2(x) \qquad (3.2.1)$$

where $L_0(x) = (x-1)(x-\alpha)/2(1+\alpha)$, where $L_1(x) = (1-x^2)/(1-\alpha^2)$, and where $L_2(x) = (x+1)(x-\alpha)/2(1-\alpha)$ are the Lagrange polynomials corresponding to -1, α, 1. Now corollary 3.1.1 implies that $d(-1,\xi_0) = d(\alpha,\xi_0) = d(1,\xi_0) = 3$ and that $d(x,\xi_0) \leq 3$ for all $x \in [-1,\alpha] \cup [\beta,1]$ are the conditions for ξ_0 to be D-optimal. Hence the variance function is a quartic such that

$$\frac{1}{3}\, d(x,\xi_0) = 1 + C(x+1)(x-\alpha)(x-w)(x-1), \qquad (3.2.2)$$

where $C > 0$ and $\alpha < w \leq \beta$. Equating the leading and trailing coefficients from (3.2.1) and (3.2.2) yields $C = (3 + \alpha^2)^{-2}$ and $w = \alpha(\alpha^2 - 5)/(3 + \alpha^2)$. Thus ξ_0 is D-optimal if and only if $\beta \geq w = \alpha(\alpha^2 - 5)/(3 + \alpha^2)$. It may be noted that if $\beta = \alpha(\alpha^2 - 5)/(3 + \alpha^2)$, then $\xi_0 = \frac{1}{3}\{\delta_{-1} + [(1-\gamma)\delta_\alpha + \gamma\delta_\beta] + \delta_1\}$ is

D-optimal for any $\gamma \in [0,1]$. It may also be verified by elementary calculations that $-\alpha \leq \alpha(\alpha^2 - 5)/(3 + \alpha^2) \leq 1$ holds for all $\alpha \in [-1,0]$.

If $-\alpha \leq \beta \leq \alpha(\alpha^2 - 5)/(3 + \alpha^2)$, then it must be true that Support(ξ_0) = $\{-1, \alpha, \beta, 1\}$ as already remarked. Thus it only remains to determine the optimal mass distribution among these points. Let ρ_0, ρ_1, ρ_2, ρ_3 denote the masses ξ_0 assigns to the points -1, α, β, 1 (respectively). Here $\rho_0 + \rho_1 + \rho_2 + \rho_3 = 1$. Also, let each V_i denote the Vandermonde determinant relative to the three points obtained by deleting the point corresponding to ρ_i from Support(ξ_0). According to the Binet-Cauchy formula, $|M(\xi_0)| = \sum_{i=0}^{3} \prod_{\substack{j=0 \\ j\neq i}}^{3} \rho_i V_i^2$. This function of ρ_0, ρ_1, ρ_2, ρ_3 may be maximized subject to $\rho_0 + \rho_1 + \rho_2 + \rho_3 = 1$ by Lagrange's method.

The resultant system of equations may be repeatedly manipulated to yield the solution

$$\rho_i = \frac{1}{3} - V_i^2 \left\{ \sum_{\substack{j=0 \\ j\neq i}}^{3} V_j^2/\rho_j \right\}^{-1} \quad \text{for } i = 0, 1, 2, 3.$$

Implementation of this implicit solution may be accomplished by an interative procedure. If $\rho_j^{(k)}$ for $j = 0, 1, 2, 3$ denote the values at stage k, the values at stage $k+1$ may be determined from

$$\rho_i^{(k+1)} = \frac{1}{3} - V_i^2 \left\{ \sum_{\substack{j=0 \\ j\neq i}}^{3} V_j^2/\rho_j^{(k)} \right\}^{-1} \quad \text{for } i = 0, 1, 2, 3.$$

Such a procedure has been programmed. Results of several runs appear in table 3.2.1. The stopping criterion used was

$$\sum_{i=0}^{3} [\rho_i^{(k)} - \rho_i^{(k-1)}]^2 \leq 10^{-8}.$$

Table 3.2.1: D-Optimal Design Masses on {-1, α, β, 1}

for Quadratic Regression

α	β	k	$\rho_0^{(k)}$	$\rho_1^{(k)}$	$\rho_2^{(k)}$	$\rho_3^{(k)}$
-.1	.1	0	.2	.3	.2	.3
		387	.3325	.1675	.1674	.3325
		0	.33	.16	.16	.35
		268	.3325	.1675	.1675	.3325
		0	.7	.05	.2	.05
		359	.3325	.1675	.1676	.3325
-.1	.165	0	.25	.25	.25	.25
		220	.3333	.3318	.0016	.3333
-.1	.2	0	.33	.16	.16	.35
		30(1)			<0	
-.5	.5	0	.6	.1	.2	.1
		16	.3114	.1885	.1886	.3114
		0	.3	.4	.2	.2(2)
		16	.3114	.1886	.1886	.3114
-.5	.65	0	.25	.25	.2	.3
		13	.3242	.2967	.0695	.3096
		0	.5	.1	.2	.2
		14	.3242	.2966	.0695	.3096
-.5	.9	0	.25	.25	.4	.1
		3(1)			<0	
-.9	.9	0	.33	.16	.16	.35
		8	.2630	.2370	.2370	.2630
		0	.45	.05	.1	.4
		9	.2630	.2370	.2370	.2630

(1) The procedure was truncated whenever a $\rho_i^{(k)} < 0$.

(2) Note that this distribution is improper (by intention).

Some features of these tabled results deserve note. The procedure seems to be insensitive to the starting distribution. In fact, even the improper initial distribution corresponding to footnote (2) led to a solution. Also, the further α and β are from zero, the faster the convergence. Finally, the procedure seems to have an interesting diagnostic feature. Recall that if $\beta > \alpha(\alpha^2 - 5)/(3 + \alpha^2)$, then $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_\alpha + \delta_1)$ is D-optimal. For the two table entries corresponding to footnote (1), this is the case. Notice that in each of these instances, the procedure truncated because a $\rho_i^{(k)} < 0$. The preceding arguments are summarized by the following lemma.

Lemma 3.2.1: Let $\{-1,1\} \subset E \subset [-1,1]$, let $\alpha = \max((-1,0] \cap E)$, and let $\beta = \min([0,1) \cap E)$.

i. If $\beta \geq \alpha(\alpha^2 - 5)/(3 + \alpha^2)$, then the D-optimal design for quadratic regression is $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_\alpha + \delta_1)$.

ii. If $\alpha \leq \beta(\beta^2 - 5)/3 + \beta^2)$, then $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_\beta + \delta_1)$.

iii. If $\beta(\beta^2 - 5)/(3 + \beta^2) < \alpha < \beta < \alpha(\alpha^2 - 5)/(3 + \alpha^2)$, then $\xi_0 = \rho_0\delta_{-1} + \rho_1\delta_\alpha + \rho_2\delta_\beta + \rho_3\delta_1$. The mass distribution must satisfy

$$\rho_i = \frac{1}{3} - V_i^2 \left\{ \sum_{\substack{j=0 \\ j \neq i}}^{3} V_j^2/\rho_j \right\}^{-1}$$

for i = 0, 1, 2,3, where each V_i is the Vandermonde determinant relative to the three support points not corresponding to ρ_i.

Proof: The preceding arguments have proven the lemma for $E = [-1,\alpha] \cup [\beta,1]$. The resultant D-optimal designs satisfy $\xi_0 \in \Xi$ for any E such that $\alpha = \max((-1,0] \cap E)$ and $\beta = \min([0,1) \cap E)$. Thus the designs ξ_0 are D-optimal for the (possibly) more restrictive case and the lemma is proved.

Note that both parts i. and ii. yield $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$ if $0 \in E$. Note also that the case $\alpha = -\beta < 0$ furnishes an example where the D-optimal design for polynomial regression of degree n is supported by more than n + 1 points.

For polynomial regression of degree n > 2, similar methods may be used to determine ξ_0. The following simple example will show how corollary 3.2.1 may be implemented for regression on two variables. Modified versions of this example will follow in sections 3.3 and 3.4.

Example 3.2.1: Let $\mathcal{X} = \{(x_1,x_2) | -1 \leq x_1, x_2 \leq 1\}$ and $f(x_1,x_2) = (1,x_1,x_2)'$. If $E = \mathcal{X}$, it is well known that the D-optimal design assigns equal mass to the four corners of $\mathcal{X}$. Now consider the restriction that $E = \{(x_1,x_2) | \ |x_1 + x_2| \leq 1\} \cap \mathcal{X}$. (Such a constraint on the control variables might be the result of experimental limitation or governmental regulation.) A natural conjecture is that the D-optimal design still assigns equal mass to the corners of E. Let ξ_0 denote the measure assigning equal mass to these six corners. Then

$$M(\xi_0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & -1/3 \\ 0 & -1/3 & 2/3 \end{bmatrix} \text{ and}$$

$$d(x_1,x_2;\xi_0) = f(x_1,x_2)'M^{-1}(\xi_0)f(x_1,x_2) = 1 + 2(x_1^2 + x_1x_2 + x_2^2).$$

Now the contour $d(x_1,x_2;\xi_0) = 3$ is an ellipse which contains E and touches precisely at the six corners. Furthermore, $d(x_1,x_2;\xi_0) \leq 3$ for all $(x_1,x_2) \in E$. Hence corollary 3.1.1 implies that ξ_0 is D-optimal.

The following example will show how the D-optimal design may be determined for an additive k × k ANOVA.

Example 3.2.2: Let $\mathcal{X} = \{1,\ldots,k\} \times \{1,\ldots,k\}$, let $f_0(i,j) = 1$, let $f_r(i,j) = \chi_{\{r+1\}}(i)$ for $r=1,\ldots,k-1$, and let $f_r(i,j)=f_{r-(k-1)}(j,i)$ for $r = k,\ldots,2k-2$. Here $\chi_{\{r+1\}}$ denotes the characteristic function of the singleton $\{r+1\}$. For $E = \mathcal{X}$, the concavity of $\log|M(\xi)|$ and the symmetry of the problem imply that the D-optimal design allocates an equal proportion of the observations to each of the k^2 points of $\mathcal{X}$.

Now consider the restriction that $E = \mathcal{X} - \{(k,k)\}$. That is, observations may not be obtained when both factors are set at their highest levels. (Such a constraint might arise if the particular combination of factor levels would be lethal to subjects in a study.) In this case, the problem is invariant with respect to permutations of the first k-1 levels of either factor and with respect to interchange of the two factors. Hence the D-optimal design ξ_0 should have these properties. This implies that $d\xi_0(i,k) = d\xi_0(k,j) = \tau$ for $1 \leq i, j \leq k - 1$ and that $d\xi_0(i,j) = [1 - 2(k-1)\tau]/(k-1)^2$ for $1 \leq i, j \leq k - 1$, where $0 \leq \tau \leq 1/2(k-1)$. Thus the problem is reduced to that of determining the value of τ which maximizes $|M(\xi_0)|$, which is just a polynomial in τ.

For k = 2, there are only three available treatment combinations. It is readily shown that the optimal value of τ is 1/3. That is,. an equal proportion of observations is allocated to each available treatment combination. However, this does not hold true for k > 2.

For k = 3, elementary manipulations yield $|M(\xi_0)| = C\,\tau^2(\frac{1}{2} - \tau)(\frac{1}{4} - \tau)^2$, where $C > 0$. This polynomial has one maxima for $0 \leq \tau \leq \frac{1}{4}$ which satisfies $0 = 160\tau^3 - 32\tau^2 - 6\tau + 1$. Approximately, $\tau = .1351$. Note that $\tau > 1/8$. Thus ξ_0 compensates for the missing cell by allocating a slightly larger proportion of observations to the highest levels of the two factors.

3.3 The (φ,ψ)-Problem

Recall that for the (φ,ψ)-problem, φ and ψ are measures on $\mathcal{X}$ with the properties that $\varphi(\mathcal{X}) < 1 < \psi(\mathcal{X})$ and $d\varphi \leq d\psi$. Recall also that $\Xi = \{\xi | d\varphi \leq d\xi \leq d\psi\}$. The following theorem is an application of theorem 3.1.1 to the present setting. It follows a theorem of Wynn ('77). See also Wynn ('82).

Theorem 3.3.1: Let $\xi_0 \in \Xi$, let $\bar{B}$ = Support $(\xi_0-\varphi)$, and let $\underline{B}$ = Support $(\psi-\xi_0)$. Then ξ_0 is D-optimal if and only if $\min_{\bar{B}} d(x,\xi_0) \geq \max_{\underline{B}} d(x,\xi_0)$.

Proof: According to remark 1.3.1, there is no loss of generality in assuming that $\varphi = 0$. For purposes of notation, let $\kappa_1 = \min_{\bar{B}} d(x,\xi_0)$ and $\kappa_2 = \max_{\underline{B}} d(x,\xi_0)$.

Assume first that $\kappa_1 \geq \kappa_2$. Now for any $\xi \in \Xi$,

$$\int_{\mathcal{X}} d(x,\xi_0)d\xi_0(x)-\int_{\mathcal{X}} d(x,\xi_0)d\xi(x)=\int_{\mathcal{X}}[d(x,\xi_0)-\kappa_2]\left[\frac{d\xi_0}{d\psi}(x)-\frac{d\xi}{d\psi}(x)\right]d\psi(x). \quad (3.3.1)$$

For $\frac{d\xi_0}{d\psi}(x) = 1$, the difference $\frac{d\xi_0}{d\psi}(x) - \frac{d\xi}{d\psi}(x) \geq 0$. Furthermore, $\frac{d\xi_0}{d\psi}(x) = 1$ implies that $x \in \bar{B}$ and hence that $d(x,\xi_0)-\kappa_2 = [d(x,\xi_0)-\kappa_1] + [\kappa_1-\kappa_2] \geq 0 + 0 = 0$. For $\frac{d\xi_0}{d\psi}(x) = 0$, the same type of reasoning implies that $\frac{d\xi_0}{d\psi}(x) - \frac{d\xi}{d\psi}(x) \leq 0$ and that

$d(x,\xi_0)-\kappa_2 \le 0$. Finally, $0 < \frac{d\xi_0}{d\psi}(x) < 1$ implies that $x \in \bar{B} \cap \underline{B}$ so that $d(x,\xi_0) = \kappa_2 = \kappa_1$ must hold. Thus the integrand on the right hand side of (3.3.1) is non-negative so that

$\max_{\xi} \int_{\mathcal{X}} d(x,\xi_0)d\xi(x) \le \int_{\mathcal{X}} d(x,\xi_0)d\xi_0(x)$. Therefore, theorem 3.1.1 implies that ξ_0 is D-optimal.

Assume next that $\kappa_1 < \kappa_2$. Then there exist $B_1 \subset \bar{B}$ and $B_2 \subset \underline{B}$ such that $d(x_1,\xi_0) < d(x_2,\xi_0)$ for all $x_1 \in B_1$ and $x_2 \in B_2$. Without loss of generality, $\frac{d\xi_0}{d\psi}(x) \le 1-\varepsilon$ on B_2 for some $\varepsilon > 0$. Now set $d\xi_1 = d\xi_0$ on B_1 and 0 elsewhere, set $d\xi_2 = d\psi$ on B_2 and 0 elsewhere, and set $\tilde{\xi} = \xi_0 + \gamma[\xi_2/\psi(B_2) - \xi_1/\xi_0(B_1)]$. Here $0 < \gamma \le \min[\xi_0(B_1), \varepsilon\psi(B_2)]$. It will first be shown that $\tilde{\xi} \in \Xi$. For any $A \subset \mathcal{X}$,

$$\tilde{\xi}(A) \ge \xi_0(A-B_1) + \xi_0(A \cap B_1) - \gamma\, \xi_0(A \cap B_1)/\xi_0(B_1)$$

$$\ge \xi_0(A \cap B_1)\,[1-\gamma/\xi_0(B_1)] \ge 0.$$

For $x \notin B_2$, $\frac{d\tilde{\xi}}{d\psi}(x) \le \frac{d\xi_0}{d\psi}(x) \le 1$. For $x \in B_2$, $\frac{d\tilde{\xi}}{d\psi}(x) \le \frac{d\xi_0}{d\psi}(x) + \frac{\gamma}{\psi(B_2)} \le (1-\varepsilon) + \varepsilon = 1$. Thus $0 \le d\tilde{\xi} \le d\psi$. Also, $\tilde{\xi}(\mathcal{X}) = \xi_0(\mathcal{X})$. Therefore, $\tilde{\xi} \in \Xi$. Finally, note that

$\int_{\mathcal{X}} d(x,\xi_0)d\tilde{\xi}(x) > \int_{\mathcal{X}} d(x,\xi_0)d\xi_0(x) + \gamma[0]$ by definition of B_1 and B_2. Hence theorem 3.1.1 implies that ξ_0 is not D-optimal in this case and the proof is complete.

Examples will now be given to show how theorem 3.3.1 may be implemented to obtain a D-optimal design.

Example 3.3.1: Consider the situation of polynomial regression of degree n on $\mathcal{X} = [-1,1]$ where $d\varphi = 0$ and $d\psi(x) = \gamma dx$ for some $\gamma > \frac{1}{2}$. This problem is symmetric with respect to reflection about the origin, so the D-optimal design ξ_0 should have the same symmetry. Then by analogy with the unrestricted solution and in view of theorem 2.3.5(ii), ξ_0 should have the form

$$d\xi_0(x) = \begin{cases} \gamma dx & x \in [-1,y_1] \cup [y_1,y_2] \cup \cdots \cup [y_{2n-2},y_{2n-1}] \cup [y_{2n},1] \\ 0 & \text{otherwise,} \end{cases}$$

where $y_i = -y_{2n-i+1}$ for $1 \leq i \leq n$. This symmetry implies that the elements of $M(\xi_0)$ are

$$M_{ij}(\xi_0) = \begin{cases} \mu_{i+j} & i+j = \text{even} \\ 0 & i+j = \text{odd} \end{cases}$$

and hence that $d(x,\xi_0) = P_n(x^2;\, y_1,\ldots,y_n)$, where

$P_n(x^2;\, y_1,\ldots,y_n) = \sum_{i=0}^{n} a_i(y_1,\ldots,y_n)x^{2i}$ is a polynomial of degree n in x^2 whose coefficients are functions of $y_1,\ldots,y_n$. Now if ξ_0 is to be D-optimal, theorem 3.3.1 implies that the variance function must have the form $d(x,\xi_0) = D+C\prod_{i=1}^{n}(x^2-y_i^2) = D+C\sum_{i=0}^{n} s_i(y_1,\ldots,y_n)x^{2i}$.

Equating coefficients of the two forms of the variance function yields the n-1 equations $\frac{a_1(y_1,\ldots,y_n)}{s_1(y_1,\ldots,y_n)} = \frac{a_i(y_1,\ldots,y_n)}{s_i(y_1,\ldots,y_n)}$ for $i = 2,\ldots,n$.

In addition, the requirement that $\xi_0(\mathcal{X}) = 1$ yields a linear constraint of the form $\sum_{i=1}^{n} b_i y_i = \kappa$. Solution of all n equations in $y_1,\ldots,y_n$

should yield the D-optimal design ξ_0.

For linear regression, it is easy to obtain

$$d\xi_0(x) = \begin{cases} \gamma dx & x \in [-1, \frac{1}{2\gamma} -1] \cup [1 - \frac{1}{2\gamma}, 1] \\ 0 & \text{otherwise} \end{cases}$$

and to verify that this design satsifies the condition of theorem 3.3.1. Note that as $\gamma \to \infty$, ξ_0 converges to $\frac{1}{2}(\delta_{-1}+\delta_1)$. This example may also be compared to one of Wynn ('77).

For quadratic regression, ξ_0 should have the form

$$d\xi_0(x) = \begin{cases} \gamma dx & x \in [-1,y_1] \cup [y_2,-y_2] \cup [-y_1,1] \\ 0 & \text{otherwise,} \end{cases}$$

where $-1 < y_1 < y_2 < 0$ and where $\xi_0(\mathcal{X}) = 1$ implies that $y_1 - y_2 = \kappa = (1-2\gamma)/2\gamma$. Now

$$M(\xi_0) = \begin{bmatrix} 1 & 0 & \mu_2 \\ 0 & \mu_2 & 0 \\ \mu_2 & 0 & \mu_4 \end{bmatrix} ,$$

where $\mu_2 = 2\gamma(1 + y_1^3 - y_2^3)/3$ and $\mu_4 = 2\gamma(1 + y_1^5 - y_2^5)/5$. Thus

$$d(x,\xi_0) = [\mu_4 + (\mu_4 - 3\mu_2^2)x^2/\mu_2 + x^4]/(\mu_4 - \mu_2^2)$$
$$= D + C(x^2-y_1^2)(x^2-y_2^2).$$

Equating coefficients of x^2 and x^4 yields

$$\frac{\mu_4-3\mu_2^2}{\mu_2(\mu_4-\mu_2^2)} \Big/ (-y_1^2 - y_2^2) = 1 \Big/ (\mu_4 - \mu_2^2).$$

Then substituting for μ_2 and μ_4, employing the constraint $y_1 - y_2 = \kappa$, and repeatedly simplifying yields that y_1 must solve

$$0 = 45\kappa\, y_1^4 - 90\kappa^2 y_1^3 + 5(2\kappa^4 + 17\kappa^3 - 4\kappa + 2)y_1^2$$
$$- 10\kappa(\kappa^4 + 4\kappa^3 - 2\kappa + 1)y_1 + (3\kappa^6 + 8\kappa^5 - 5\kappa^3 + 5\kappa^2 + 3\kappa - 2).$$

Table 3.3.1 summarizes the results of this method for various values of γ.

Table 3.3.1: D-Optimal Design for Quadratic Regression on [-1,1] under $d\psi(x) = \gamma dx$

γ	$-y_2$	$-y_1$	$\xi_0([y_2,-y_2])$	$\xi_0([-1,y_1])=\xi_0([-y_1,1])$
.5	.4472	.4472	.4472	.2764
.6	.3724	.5391	.4469	.2766
.7	.3146	.6004	.4404	.2798
.8	.2707	.6457	.4331	.2835
.9	.2367	.6811	.4260	.2870
1.0	.2098	.7098	.4196	.2902
1.5	.1322	.7988	.3966	.3017
2.0	.0957	.8457	.3828	.3086
3.0	.0613	.8946	.3678	.3161
5.0	.0355	.9355	.3550	.3225
10	.0172	.9672	.3440	.3280
100	.0017	.9967	.3348	.3326
∞	0	1	1/3	1/3

Observe that the limiting cases of $\gamma = 1/2$ and $\gamma \to \infty$ yield (respectively) $\xi_0 = \psi$ and $\xi_0 = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$ as they should.

<u>Example 3.3.2</u>: Consider a modification of example 3.2.1 with $\mathcal{X} = \{(x_1,x_2) | -1 \le x_1,x_2 \le 1\}$, with $f(x_1,x_2) = (1,x_1,x_2)'$, with $d\,\varphi = 0$, and with

$$d\psi(x_1,x_2) = \begin{cases} 1 & (x_1,x_2) \in \{(-1,1),\ (1,-1),\ (1,0),(0,1),(0,-1),(-1,0)\} \\ \gamma & (x_1,x_2) \in \{(1,1),\ (-1,-1)\} \\ 0 & (x_1,x_2) \in \{(x_1,x_2) \mid |x_1+x_2| > 1\} - \{(1,1),\ (-1,-1)\} \\ \text{arbitrary} & \text{otherwise,} \end{cases}$$

where $0 \leq \gamma \leq 1/4$. That is, ψ has atoms of mass 1 at the six corners from example 3.2.1, atoms of mass γ at (1,1) and (-1,-1), no mass outside the region E from example 3.2.1 except at (1,1) and (-1,-1), and arbitrary mass distribution within the region E from example 3.2.1. Theorem 3.1.1 will now be used to establish D-optimality.

For $\frac{1}{8} \leq \gamma \leq \frac{1}{4}$, let

$$d\xi_0(x_1,x_2) = \begin{cases} \gamma & (x_1,x_2) \in \{(1,1),\ (-1,-1)\} \\ \frac{1}{2}-\gamma & (x_1,x_2) \in \{(1,-1),\ (-1,1)\} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$M(\xi_0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4\gamma-1 \\ 0 & 4\gamma-1 & 1 \end{bmatrix}$$

and $d(x_1,x_2;\ \xi_0) = 1 + (x_1^2-2\kappa x_1x_2 + x_2^2)/(1-\kappa^2)$, where $\kappa = 4\gamma-1$. Now $\bar{B} = \{(1,1),\ (-1,-1),\ (1,-1),\ (-1,1)\}$ and $\underline{B} = \mathcal{X} - \{(1,1),\ (-1,-1)\}$. Thus $\min_{\bar{B}} d(x_1,x_2;\ \xi_0) = d(1,-1;\ \xi_0) = d(-1,1;\ \xi_0)$. Also, $\max_{\underline{B}} d(x_1,x_2;\ \xi_0) = d(1,-1;\ \xi_0) = d(-1,1;\ \xi_0)$. Thus theorem 3.3.1 implies that ξ_0 is D-optimal. Note that when $\gamma = 1/4$, then ξ_0 reduces to the known unrestricted result.

For $0 \leq \gamma \leq 1/8$, let

$$d\xi_0(x_1,x_2)= \begin{cases} \gamma & (x_1,x_2) \in \{(1,1),\ (-1,-1)\} \\ \frac{1}{6} + \frac{5}{3}\gamma & (x_1,x_2) \in \{(1,-1),\ (-1,1)\} \\ \frac{1}{6} - \frac{4}{3}\gamma & (x_1,x_2) \in \{(1,0),\ (0,1),\ (-1,0),\ (0,-1)\} \\ 0 & \text{otherwise.} \end{cases}$$

Then $$M(\xi_0)= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2(1+4\gamma)/3 & -(1+4\gamma)/3 \\ 0 & -(1+4\gamma)/3 & 2(1+4\gamma)/3 \end{bmatrix}$$

and $d(x_1,x_2;\ \xi_0) = 1+2(x_1^2 + x_1x_2 + x_2^2)/3\kappa$, where $\kappa = (1+4\gamma)/3$. Now $\bar{B} = \{(1,1),\ (-1,-1)\}$ and $\underline{B} = \mathcal{X} - \bar{B}$. Also, it is readily shown that

$$\min_{\bar{B}} d(x_1,x_2;\ \xi_0) = 1+2/\kappa > 1 + 2/3\kappa = \max_{\underline{B}} d(x_1,x_2;\ \xi_0).$$

Thus theorem 3.3.1 again implies that ξ_0 is D-optimal. Note that when $\gamma = 0$, then ξ_0 reduces to the solution of example 3.2.1.

For an application of these methods to a one-way ANOVA, see Wynn ('77). The following example illustrates the two-way case.

<u>Example 3.3.3</u>: Consider a modification of example 3.2.2 with $d\,\varphi = 0$ and with

$$d\psi(i,j) = \begin{cases} \gamma & (i,j) = (k,k) \\ 1 & \text{otherwise,} \end{cases}$$

where $0 \leq \gamma \leq 1/k^2$. That is, ψ limits only the proportion of observations that may be allocated to the highest levels of both factors. This problem has precisely the same invariance as example 3.2.2, so the D-optimal design ξ_0 must satisfy $d\xi_0(i,k) = d\xi_0(k,j) = \tau$ and $d\xi_0(i,j) = [1 - \rho - 2(k-1)\tau]/(k-1)^2$ for $1 \leq i, j \leq k - 1$, where $\rho = d\xi_0(k,k)$. Here $0 \leq \rho \leq \gamma$ and $0 \leq \tau \leq \rho + 2(k-1)\tau \leq 1$.

Now let $\tilde{\xi}_0 = [\xi_0 - \rho\delta_{(k,k)}]/(1-\rho)$. Then

$$|M(\xi_0)| = |(1-\rho)M(\tilde{\xi}_0) + \rho f(k,k)f(k,k)'|$$

$$= (1-\rho)|M(\tilde{\xi}_0)|\left\{1 + \frac{\rho f(k,k)'M^{-1}(\tilde{\xi}_0)f(k,k)}{1-\rho}\right\}$$

$$= |M(\tilde{\xi}_0)|\{1 + \rho[f(k,k)'M^{-1}(\tilde{\xi}_0)f(k,k) - 1]\}$$

$$= |M(\tilde{\xi}_0)|\{1 + \rho[d(k,k;\tilde{\xi}_0) - 1]\}.$$

According to example 3.2.2, $d(k,k;\tilde{\xi}_0) \geq \max\limits_{i+j<2k} d(i,j;\tilde{\xi}_0) \geq 2k-1$. The first inequality must hold since $\tilde{\xi}_0$ is not D-optimal for the unrestricted problem. The second inequality follows from lemma 3.1.1(ii) with $n = 2(k-1)$. Hence $d(k,k;\tilde{\xi}_0) - 1 \geq 2(k-1) \geq 0$. Thus $|M(\xi_0)|$ is maximized for any $\tilde{\xi}_0$ by choosing $\rho = \gamma$. It now remains only to determine the value of τ which maximizes $|M(\xi_0)|$.

For $k = 2$, it is readily shown that $\tau = (1-\gamma)/3$. That is, an equal proportion of observations is allocated to each of the "unrestricted" treatment combinations. Thus $\tilde{\xi}_0$ coincides with the solution of example 3.2.2 for $k = 2$. That such a coincidence need not be true in general is apparent by considering the limiting case $\gamma = 1/k^2$. In this case, ξ_0 must allocate equal mass to all cells but the design $\tilde{\xi}_0$ from example 3.2.2 does not.

3.4 The p-Problem

Recall that for the p-problem, $\{G_\omega | \omega \in \Omega\}$ is a collection of open sets, $p_\omega \in (0,1)$ for each $\omega \in \Omega$, and $\Xi = \{\xi | \xi(G_\omega) \leq p_\omega;\ \omega \in \Omega\}$. Now consider the case that $\Omega = \{1,\ldots,s\}$. Then let $G_0 = \mathcal{X} - \bigcup_{i=1}^{s} G_i$ and $p_0 = 1$. The following corollary provides a more easily implemented version of theorem 3.1.1.

Corollary 3.4.1: Let $\Omega = \{1,\ldots,s\}$, let $\xi_0 \in \Xi$, and let $\Delta_i = \sup_{G_i} d(x,\xi_0)$ for each $i = 0,\ldots,s$. Also, let $\Delta_{(0)} \geq \cdots \geq \Delta_{(s)}$ be the ordered values of Δ_i, and let $r \in \{0,\ldots,s\}$ be the largest integer for which $\sum_{i=0}^{r} p_{(i)} \leq 1$, where each $p_{(i)}$ corresponds to $\Delta_{(i)}$. Then ξ_0 is D-optimal if and only if

$$\sum_{i=0}^{r} p_{(i)}\Delta_{(i)} + \left(1 - \sum_{i=0}^{r} p_{(i)}\right)\Delta_{(r+1)} = n + 1.$$

Proof: Under the assumptions made,

$$\max_{\xi\in\Xi} \int d(x,\xi_0)d\xi(x) = \sum_{i=0}^{r} p_{(i)}\Delta_{(i)} + \left(1 - \sum_{i=0}^{r} p_{(i)}\right)\Delta_{(r+1)}.$$ Thus the corollary follows immediately from theorem 3.1.1.

Examples will now be given to show how a D-optimal design may be obtained for the p-problem.

Example 3.4.1: Consider the setting of linear regression on $\mathcal{X} = [-1,1]$ with $s = 1$, with $G_1 = (\alpha_1, 1]$, and with $p_1 < 1/2$. The D-optimal design turns out to be $\xi_0 = p_1\delta_1 + \gamma\delta_{\alpha_1} + (1 - p_1 - \gamma)\delta_{-1}$, where $\gamma = \frac{1}{2}(\alpha_1 + 1 - 4p_1)_+$. Thus if $\alpha_1 \leq 4p_1 - 1$, then ξ_0 allocates observations as equally as possible to the endpoints. Otherwise, α_1 is "close enough" to 1 to be included in Support(ξ_0). In the case that $\alpha_1 \leq 4p_1 - 1$, elementary calculations yield

$\Delta_1 = \sup_{G_1} d(x,\xi_0) = d(1,\xi_0) = 1/p_1$ and

$\Delta_0 = \sup_{\mathcal{X}-G_1} d(x,\xi_0) = d(-1,\xi_0) = 1/(1-p_1) < 1/p_1$. Then $p_1\Delta_1 + (1-p_1)\Delta_0 = 2$ and so corollary 3.4.1 implies that ξ_0 is in fact D-optimal. In the case that $\alpha_1 \geq 4p_1 - 1$, it can be shown that $\Delta_1 = d(1,\xi_0)$, that $\Delta_0 = d(\alpha_1,\xi_0) < \Delta_1$, and that $p_1\Delta_1 + (1-p_1)\Delta_0 = 2$. Thus ξ_0 must be D-optimal in any case.

Example 3.4.2: Consider the setting of quadratic regression on $\mathcal{X} = [-1,1]$ with $s = 1$, with $G_1 = (-\alpha,\alpha)$, and with $p_1 < 1/3$. This problem is symmetric about the origin and so the D-optimal design ξ_0 must also be symmetric. The only way to satisfy the symmetry requirement, to satisfy theorem 2.4.4, and to have $|M(\xi_0)| > 0$ is that either Support(ξ_0) = $\{-1, 0, 1\}$ or Support(ξ_0) = $\{-1, -\alpha, 0, \alpha, 1\}$ and $\xi_0(\{0\}) = p_1$. It turns out that the latter case yields the D-optimal design $\xi_0 = p_1\delta_0 + \gamma(\delta_{-\alpha} + \delta_\alpha)/2 + (1 - p_1 - \gamma)(\delta_{-1} + \delta_1)/2$, where $\gamma = (2 - \alpha^2 - \sqrt{\alpha^4 - \alpha^2 + 1})/(3 - 3\alpha^2)$. For this design, $\Delta_1 = d(0,\xi_0)$ and $\Delta_0 = d(-1,\xi_0) = d(-\alpha,\xi_0) = d(\alpha,\xi_0) = d(1,\xi_0) < \Delta_0$. Also, $p_1\Delta_1 + (1 - p_1)\Delta_0 = 3$ so that application of corollary 3.4.1 establishes that ξ_0 is D-optimal.

The following two examples will show how the p-problem may be related to the (φ,ψ)-problem.

Example 3.4.3: Consider a modification of example 3.3.2 with $\mathcal{X} = \{(x_1,x_2) \mid -1 \leq x_1,\ x_2 \leq 1\}$, with $f(x_1,x_2) = (1,x_1,x_2)'$, with $G_1 = \{(x_1,x_2) \mid x_1 + x_2 > 1\} \cap \mathcal{X}$, with $G_2 = \{(x_1,x_2) \mid x_1 + x_2 < -1\} \cap \mathcal{X}$, and with $p_1 = p_2 = \gamma \leq 1/4$. That is, observations may be allocated at will within the set E of example 3.2.1 but no more than a proportion γ may be allocated to either G_1 or G_2. It will be argued that the design ξ_0 of example 3.3.2 is still D-optimal in the present context.

For $\frac{1}{8} \leq \gamma \leq \frac{1}{4}$, recall from example 3.3.2 that $\Delta_1 = \Delta_2 = 1 + 2(1+\kappa)^{-1}$ and that $\Delta_0 = 1 + 2(1-\kappa)^{-1} < \Delta_1$, where $\kappa = 4\gamma - 1$. Thus $p_1\Delta_1 + p_2\Delta_2 + (1 - p_1 - p_2) = 3$ implies that ξ_0 is D-optimal according to corollary 3.4.1. For $0 \leq \gamma \leq \frac{1}{8}$, recall from example 3.3.2 that $\Delta_1 = \Delta_2 = 1 + 2/\kappa$ and that $\Delta_0 = 1 + 2/3\kappa < \Delta_1$, where $\kappa = (1 + 4\gamma)/3$. Thus $p_1\Delta_1 + p_2\Delta_2 + (1 - p_1 - p_2)\Delta_0 = 3$ implies that ξ_0 is D-optimal in this case as well.

Example 3.4.4: Consider a modified version of example 3.3.3 with $G_1 = \{(k,k)\}$ and $p_1 = \gamma \leq 1/k^2$. In this particular case of a finite set $\mathcal{X}$, the set Ξ of allowable measures for the p-problem coincides with that of the (φ,ψ)-problem. Thus the developments of example 3.3.3 apply to the present example in full.

It should be noted that corollary 3.4.1 lends itself to an iterative procedure for the construction of a D-optimal design when $\Omega = \{1,\ldots,s\}$. Let $\xi^{(k)}$ denote the design at stage k and $\Delta_i^{(k)} = \sup_{G_i} d(x,\xi^{(k)}) = d(x_i,\xi^{(k)})$. Also, let $\Delta_{(0)}^{(k)} \geq \ldots \geq \Delta_{(s)}^{(k)}$ and let $p_{(i)}^{(k)}$ and $x_{(i)}^{(k)}$ correspond to $\Delta_{(i)}^{(k)}$ for $0 \leq i \leq s$. Finally, let $r^{(k)}$ be the largest integer for which $\sum_{i=0}^{r^{(k)}} p_{(i)}^{(k)} \leq 1$ and let

$\tilde{\xi}^{(k)} = \sum_{i=0}^{r^{(k)}} p_{(i)}^{(k)} \delta_{x_{(i)}^{(k)}} + (1 - \sum_{i=0}^{r^{(k)}} p_{(i)}^{(k)}) \delta_{x_{(r+1)}^{(k)}}$. Then for the next stage of the iteration, set $\xi^{(k+1)} = (1 - \tau^{(k)})\xi^{(k)} + \tau^{(k)}\tilde{\xi}^{(k)}$, where $0 \leq \tau^{(k)} \leq 1$. To complete the specification of the procedure, it remains only to prescribe a choice of $\tau^{(k)}$ at each stage. It is suggested that $\tau^{(k)}$ be chosen according to Federov's ('72) scheme: use $\tau^{(k)} = \tau$ until $|M(\xi^{(k)})|$ starts to decrease; then use $\tau^{(k)} = \tau/2$ until $|M(\xi^{(k)}|$ starts to decrease again; continue reducing the mixture proportion by half until a desired stability of $|M(\xi^{(k)})|$ is achieved.

In the special case of polynomial regression on an interval, the admissibility methods of section 2.4 may be used at any stage of such a procedure to improve upon $\xi^{(k)}$. Such an improvement would typically simplify the support of $\xi^{(k)}$ and speed the convergence of the procedure.

Detailed investigation of the properties of such a procedure is beyond the scope of the present study. A good reference may be Welch ('81, to appear).

3.5 The Marginal Restriction Problem

Recall that for the marginal restriction problem $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, a measure ξ_1^* on $\mathcal{X}_1$ is prescribed, and $\Xi = \{\xi | \xi(A \times \mathcal{X}_2) = \xi_1^*(A)$ for all Borel sets $A \subset \mathcal{X}_1\}$. The following result of Cook and Thibodeau ('80) is an easy consequence of the more general theorem 3.1.1.

Corollary 3.5.1: For the marginal restriction problem, ξ_0 is D-optimal if and only if

$$\int_{\mathcal{X}_1} \max_{\mathcal{X}_2} d(x_1,x_2;\xi_0)d\xi_1^*(x_1) = \min_{\tilde{\xi}\in\Xi} \int_{\mathcal{X}_1} \max_{\mathcal{X}_2} d(x_1,x_2;\tilde{\xi})d\xi_1^*(x_1) = n + 1.$$

Proof: For the marginal restriction problem,

$\max_{\xi\in\Xi} \int_{\mathcal{X}_1}\int_{\mathcal{X}_2} d(x_1,x_2;\tilde{\xi})d\xi(x_1,x_2) = \int_{\mathcal{X}_1} \max_{\mathcal{X}_2} d(x_1,x_2;\tilde{\xi})d\xi_1^*(x_1)$. Thus the corollary follows immediately from theorem 3.1.1.

The following consequence of corollary 3.5.1 may add insight to its criterion for D-optimality.

Corollary 3.5.2: ξ_0 is D-optimal if and only if

$$\max_{\mathcal{X}_2} d(x_1,x_2;\xi_0) = \int_{\mathcal{X}_2} d(x_1,x_2;\xi_0)d\xi_0(x_2|x_1) \text{ for all } x_1 \in \text{Support}(\xi_1^*).$$

Here $d\xi_0(\cdot|x_1)$ denotes the conditional measure of ξ_0 on $\mathcal{X}_2$ given each $x_1 \in \text{Support}(\xi_1^*)$.

Proof: This result has been proven by Cook and Thibodeau ('80).

Two additional results of Cook and Thibodeau will now be summarized by the following lemma. They concern a product and an additive model.

Lemma 3.5.1: Let g_1 and g_2 be vector-valued functions defined on $\mathcal{X}_1$ and $\mathcal{X}_2$. Also, let ξ_2 be a D-optimal design for g_2 on $\mathcal{X}_2$ and $\tilde{\xi}_2$ be D-optimal for $(1,g_2')'$ on $\mathcal{X}_2$.

i. If $f(x_1,x_2) = g_1(x_1) \otimes g_2(x_2)$, the Kronecker product, then the product measure $d\xi_0(x_1,x_2) = d\xi_2(x_2)d\xi_1^*(x_1)$ is a marginally restricted D-optimal design.

ii. If $f(x_1,x_2) = (1,g_1(x_1)', g_2(x_2)')'$, then $d\xi_0(x_1,x_2) = d\tilde{\xi}_2(x_2)d\xi_1^*(x_1)$ is D-optimal.

Proof: See Cook and Thibodeau ('80).

Example 3.5.1: Consider again the situation of example 3.2.2 with a marginal restriction imposed. This problem fits exactly into the mold of lemma 3.5.1(ii) with $g_1 = (f_1,\dots,f_{k-1})'$ and $g_2 = (f_k,\dots,f_{2k-2})'$. Thus the D-optimal design is $d\xi_0(x_1,x_2) = d\tilde{\xi}_2(x_2)d\xi_1^*(x_1)$, where $\tilde{\xi}_2$ is D-optimal for $(1,g_2')'$ on $\mathcal{X}_2$. To complete the solution, it is only necessary to note that the balanced one-way design $\tilde{\xi}_2 = (\delta_1 +\dots+\delta_k)/k$ is D-optimal for $(1,g_2')'$ on $\mathcal{X}_2$.

The following example, with one minor correction, is due to Cook and Thibodeau ('80).

Example 3.5.2: Let $\mathcal{X} = [-1,1] \times [-1,1]$ and $f(x_1,x_2) = (1,x_2,x_1x_2)'$. Also, let $\xi_2 = \frac{1}{2}(\delta_{-1} + \delta_1)$ and $d\xi_0(x_1,x_2) = d\xi_2(x_2)d\xi_1^*(x_1)$. Then it is readily demonstrated that $d(x_1,x_2;\xi_0) = 1 + x_2^2 + x_2^2(x_1-\mu_1)^2/(\mu_2-\mu_1^2)$, where $\mu_1 = \int_{\mathcal{X}_1} x_1 d\xi_1^*(x_1)$ and $\mu_2 = \int_{\mathcal{X}_2} x_1^2 d\xi_1^*(x_1)$. Thus

$$\int_{\mathcal{X}_1} \max_{\mathcal{X}_2} d(x_1,x_2;\xi_0)d\xi_1^*(x_1) = \int_{-1}^{1} [2 + (x_1-\mu_1)^2/(\mu_2-\mu_1^2)]d\xi_1^*(x_1) = 3$$

so that corollary 3.5.1 implies that ξ_0 is D-optimal for this problem.

For a detailed discussion of procedures to construct marginally restricted D-optimal designs, see Cook and Thibodeau ('80).

CHAPTER IV

c-OPTIMALITY

4.1 General Results

Now that D-optimality has been treated, the criterion of "c-optimality" will be considered. The setting for c-optimality is that a particular linear combination $c'\theta$ is to be estimated by $c'\hat{\theta}$. Thus it is natural to attempt to minimize the variance of $c'\hat{\theta}$. If $|M(\xi)| > 0$, this variance is proportional to $\Delta(c,\xi) = c'M^{-1}(\xi)c$. For $|M(\xi)| = 0$, the inverse operation should be replaced by a generalized inverse. In any case, c should be estimable with respect to $M(\xi)$. The following definition, which is developed in Karlin and Studden ('66b), incorporates these versions of the criterion.

Definition 4.1.1: Let $V = \{v|M(\xi)v = 0\}$ and let $V^{\perp}$ be its orthogonal complement. The variance function of the design ξ is

$$\Delta(c,\xi) = \max_{\substack{v\in V^{\perp} \\ v\neq 0}} v'cc'v/v'M(\xi)v$$

if c is estimable with respect to ξ and ∞ otherwise.

The following lemma provides a more convenient form of this variance function for present purposes.

Lemma 4.1.1: The variance function is $\Delta(c,\xi) = 1/\min_{v'c=1} v'M(\xi)v$ if c is estimable with respect to ξ and ∞ otherwise.

Proof: Recall first that c is estimable if and only if $c \in V^{\perp}$. Hence $\max_{\substack{v \in V^{\perp} \\ v \neq 0}} v'cc'v/v'M(\xi)v \geq (c'c)^2/c'M(\xi)c > 0$. Thus the maximization may be carried out for $v'c \neq 0$ so that $\Delta(c,\xi) = \max_{\substack{v \in V^{\perp} \\ v'c=1}} 1/v'M(\xi)v$ when c is estimable. For arbitrary v, let $v_{\perp}$ denote its projection onto $V^{\perp}$. Then $v'c = v_{\perp}'c$ for $c \in V^{\perp}$ and $v'M(\xi)v = v_{\perp}'(\xi)v_{\perp}$ by defnition of $V^{\perp}$. Thus $\Delta(c,\xi) = \max_{v'c=1} 1/v'M(\xi)v = 1/\min_{v'c=1} v'M(\xi)v$ for c estimable so that the lemma is proved.

The following definition makes explicit the design criterion of c-optimality.

Definition 4.1.2: A design $\xi_0 \in \Xi$ is c-optimal if and only if

$$\Delta(c,\xi_0) = \min_{\xi \in \Xi} \Delta(c,\xi) = 1/\max_{\xi \in \Xi} \min_{v'c=1} v'M(\xi)v.$$

Of course for non-triviality it must be assumed that there exists at least one design $\xi \in \Xi$ such that $\Delta(c,\xi) < \infty$.

The question of c-optimality has been addressed by many authors in the unrestricted design setting where $\Xi = \Xi_0$, the set of all probability measures on $\mathcal{X}$. Elfving ('52) has provided a very appealing geometric characterization of c-optimality. An equivalent approximation theory problem has been derived by Kiefer and Wolfowitz ('59). They then obtain the optimal design for estimating the highest coefficient of a polynomial regression function. The problem of polynomial extrapolation has been solved by Hoel and Levine ('64). Studden ('68) has obtained the optimal design for estimating any one of the coefficients of a polynomial regression function.

The main intent of the present section is to adapt the approximation theory equivalence to the problem of determining a c-optimal restricted design. The later sections will provide examples of c-optimal designs under various restrictions. These examples should serve to indicate how c-optimality can depend on the particular form of restriction.

The following theorem provides a characterization of c-optimality. It is essentially an adaptation of a result of Kiefer and Wolfowitz ('59) to the restricted design setting.

Theorem 4.1.1: ξ_0 is c-optimal if and only if there exists $\tilde{v}$ such that $\tilde{v}'c=1$ and

$$\max_{\xi\in\Xi} \int_{\mathcal{X}}[\tilde{v}'f(x)]^2 d\xi(x) = \int_{\mathcal{X}}[\tilde{v}'f(x)]^2 d\xi_0(x) = \min_{v'c=1} \int_{\mathcal{X}}[v'f(x)]^2 d\xi_0(x). \quad (4.1.1)$$

Proof: Recall from definition 4.1.2 that $\xi_0 \in \Xi$ is c-optimal if and only if $\min_{v'c=1} v'M(\xi_0)v = \max_{\xi\in\Xi} \min_{v'c=1} v'M(\xi_0)v$. The theorem will now be proven by formulating it as a game theory problem. To do so, let $K(\xi,v) = v'M(\xi)v$ be the payoff function of the game. Then $K(\xi,v)$ is continuous in both arguments, linear in ξ, and convex in v. Also, Ξ is convex and (weakly) compact and $\{v|v'c = 1\}$ is convex. Thus, according to Karlin ('59), the game is determined and there exist. $\xi_0 \in \Xi$ and $\tilde{v}$ such that $\tilde{v}'c=1$ and such that

$\max_{\xi\in\Xi} K(\xi,\tilde{v}) = K(\xi_0,\tilde{v}) = \min_{v'c=1} K(\xi_0,v) = \max_{\xi\in\Xi} \min_{v'c=1} K(\xi,v)$. To complete the proof of the theorem, it is only necessary to observe that $K(\xi,v) = v'M(\xi)v = \int_{\mathcal{X}}[v'f(x)]^2 d\xi(x)$.

Note that in the unrestricted case that $\Xi = \Xi_0$, the theorem yields the Kiefer and Wolfowitz ('59) result that

$$\min_{v'c=1} \max_{\mathcal{X}} [v'f(x)]^2 = \int_{\mathcal{X}} [\tilde{v}'f(x)]^2 d\xi_0(x) = \min_{v'c=1} \int_{\mathcal{X}} [v'f(x)]^2 d\xi_0(x).$$

In order to accent the approximation theory flavor of this theorem, let the linear space $C(\mathcal{X})$ of continuous real-valued functions on $\mathcal{X}$ be endowed with the norm whose square is $||g||^2 = \max_{\xi \in \Xi} \int g^2 d\xi$. To verify that $||\cdot||$ is in fact a norm, note first that it may be assumed without loss of generality that $\mathcal{X} = \bigcup_{\xi \in \Xi} \text{Support}(\xi)$. Then $||g||^2 = 0$ if and only if $g(x) = 0$ on $\mathcal{X}$. Also, $||\gamma g||^2 = \gamma^2 ||g||^2$ is immediate. Thus it remains only to verify the triangle inequality. For arbitrary $g_1, g_2 \in C(\mathcal{X})$,

$$\begin{aligned} ||g_1 + g_2||^2 &= \max_{\xi \in \Xi} \int [g_1^2 + 2g_1 g_2 + g_2^2] d\xi \\ &\leq \max_{\xi \in \Xi} \int_{\mathcal{X}} g_1^2 d\xi + \max_{\xi \in \Xi} \int_{\mathcal{X}} g_2^2 d\xi + 2\max_{\xi \in \Xi} \int_{\mathcal{X}} g_1 g_2 d\xi \\ &= ||g_1||^2 + ||g_2||^2 + 2\max_{\xi \in \Xi} \int_{\mathcal{X}} g_1 g_2 d\xi . \end{aligned}$$

By the Cauchy-Schwartz inequality,

$$\begin{aligned} \max_{\xi \in \Xi} \left[\int_{\mathcal{X}} g_1 g_2 d\xi \right]^2 &\leq \max_{\xi \in \Xi} \left[\int_{\mathcal{X}} g_1^2 d\xi \int_{\mathcal{X}} g_2^2 d\xi \right] \\ &\leq \max_{\xi \in \Xi} \int_{\mathcal{X}} g_1^2 d\xi \max_{\xi \in \Xi} \int_{\mathcal{X}} g_2^2 d\xi \\ &= ||g_1||^2 ||g_2||^2. \end{aligned}$$

Thus $\max_{\xi \in \Xi} \int_{\mathcal{X}} g_1 g_2 d\xi \leq \max_{\xi \in \Xi} \left| \int_{\mathcal{X}} g_1 g_2 d\xi \right| \leq ||g_1|| \, ||g_2||$ so that $||g_1 + g_2||^2 \leq ||g_1||^2 + ||g_2||^2 + 2||g_1|| \, ||g_2||$, verifying the triangle inequality. It may now be noted that, in terms of this norm $||\cdot||$, the quantity in (4.1.1) is

$$\min_{v'c=1} \max_{\xi \in \Xi} \int_{\mathcal{X}} [v'f(x)]^2 d\xi(x) = \min_{v'c=1} ||v'f||^2 = ||\tilde{v}'f||^2.$$

It turns out that there is a dual version of this approximation theory problem which leads to an analogue of Elfving's ('52) theorem. For purposes of notation, let the vector $\underset{\sim}{F}(f) = [F(f_0), F(f_1), \ldots, F(f_n)]'$ for any bounded linear functional F on $C(\mathcal{X})$. The dual problem is to determine $B^{-1} = \min_{\underset{\sim}{F}(f)=c} ||F||$, where the standard notation $||F|| = \max_{g \in C(\mathcal{X})} |F(g)|/||g||$ is used for the norm of the functional F. It is proved in Krein and Nudelman ('77) that $B = \min_{v'c=1} ||v'f||$. Furthermore, if $R = \{\underset{\sim}{F}(f) | \; ||F|| \leq 1\}$, then $Bc = \underset{\sim}{F}_0(f) \in \partial R$, where F_0 is some linear functional of norm 1.

Now according to the Riesz representation theorem, any bounded linear functional on $C(\mathcal{X})$ admits an integral representation of the form $F(g) = \int_{\mathcal{X}} g(x) d\sigma(x)$, where σ is a finite signed measure on $\mathcal{X}$. In the unrestricted case that $\Xi = \Xi_0$, $||F|| = 1$ implies that the signed measure $d\sigma(x) = s(x)d\xi(x)$, where $|s(x)|$ and $\xi \in \Xi_0$. Thus

$$Bc = \int_{\mathcal{X}} f(x)s(x)d\xi_0(x) \quad \partial R \tag{4.1.2}$$

for some $\xi_0 \in \Xi_0$. Hence $B = B\tilde{v}'c = \int \tilde{v}'f(x)s(x)d\xi_0(s)$ so that $B^2 \leq \int_{\mathcal{X}} [v'f(x)]^2 d\xi_0(x) \leq ||\tilde{v}'f||^2 = B^2$. That is, ξ_0 is c-optimal so that (4.1.2) is merely Elfving's ('52) theorem for $\Xi = \Xi_0$. Furthermore, in this unrestricted case, R is the convex hull of $\{\pm f(x) | x \in \mathcal{X}\}$. It will be shown by example 4.3.1 in section 4.3 that (4.1.2) need not hold for general Ξ.

4.2 The E-Problem

As already remarked, the E-problem may be considered as an unrestricted design setting with $\mathcal{X}' = E$. Thus the following corollary summarizes the comments of the previous section for $\Xi = \Xi_0$.

Corollary 4.2.1: i. $\xi_0 \in \Xi$ is c-optimal if and only if

$$\min_{v'c=1} \max_{E} [v'f(x)]^2 = \int_E [\tilde{v}'f(x)]^2 d\xi_0(x) = \min_{v'c=1} \int_E [v'f(x)]^2 d\xi_0(x).$$

ii. Let R denote the convex hull of $\{\pm f(x) | x \in E\}$. Then $\xi_0 \in \Xi$ is c-optimal if and only if there exist $B > 0$ and $|s(x)| = 1$ such that $\int_E f(x)s(x)d\xi_0(x) = Bc \in \partial R$.

Proof: This corollary merely summarizes results of Kiefer and Wolfowitz ('59) and Elfving ('52).

It should be noted that, according to Caratheodory's theorem, any point on the boundary of R may be represented as a convex combination of n + 1 or fewer extreme points of R. Thus Support(ξ_0) need include at most n + 1 points of E. It will be seen that for some special cases of the E-problem, the c-optimal design may have more than n + 1 support points.

Examples will now be given to show how the E-problem restriction can affect known c-optimality results.

Known c-optimality results for unrestricted polynomial regression on [-1,1] will first be recalled. For extrapolation to $\tilde{x}$, where $|\tilde{x}| > 1$, the appropriate $c = f(\tilde{x})$. For estimation of θ_n, the highest regression coefficient, $c = (0,\ldots,0,1)'$. For the two types of c, it has been shown by Hoel and Levine ('64) and Kiefer and Wolfowitz ('59) that Support(ξ_0) = $\{x_0,\ldots,x_n\}$, where the $x_i = \cos[(1-i/n)\pi]$ are the Tchebycheff points of order n. For $c = f(\tilde{x})$, the corresponding design weights ρ_i are proportional to $|L_i(\tilde{x})| = |\prod_{\substack{j=0 \\ j\neq i}}^{n} (\tilde{x}-x_j)/(x_i-x_j)|$. For

For $c = (0,\ldots,0,1)'$, the weights ρ_i are proportional to $2 - \chi_{\{1\}}(x_i^2)$. Recall also the trivial result that for $c = f(\tilde{x})$ and $|\tilde{x}| \leq 1$, the c-optimal design is $\xi_0 = \delta_{\tilde{x}}$.

Example 4.2.1: Consider now the problem of quadratic regression on $E \subset [-1,1]$ for the purpose of extrapolation to $\tilde{x}$, where $|\tilde{x}| > 1$. Thus $f(x) = (1,x,x^2)'$, without loss of generality $\{-1,1\} \subset E$, and $c = f(\tilde{x})$. Let $\alpha = \max([-1,0] \cap E)$ and $\beta = \min([0,1] \cap E)$. In the case that $\alpha = \beta = 0$, the c-optimal design coincides with the unrestricted solution that $\xi_0 = \rho_0\delta_{-1} + \rho_1\delta_0 + \rho_2\delta_1$, where the design weights ρ_i are proportional to $|L_i(\tilde{x})| = |\prod_{\substack{j=0 \\ j\neq i}}^{2} (\tilde{x}-x_j)/(x_i-x_j)|$.

It remains to determine the c-optimal design ξ_0 when $\alpha < 0 < \beta$. It is claimed that $\text{Support}(\xi_0) = \{-1,\alpha,1\}$ if $|\alpha| \leq |\beta|$ and $\text{Support}(\xi_0) = \{-1,\beta,1\}$ if $|\alpha| \geq |\beta|$. Without loss of generality, only the case $|\alpha| \leq |\beta|$ need be considered. It will first be argued that $\min_{v'c=1} \max_{E} |v'f(x)| = \max_{E} |\tilde{v}'f(x)|$, where $\tilde{v} = (-\frac{1}{2}(1+\alpha^2), 0, 1)'$. Thus $\tilde{v}'f(x) = x^2 - (1+\alpha^2)/2$ so that $\tilde{v}'f(-1) = \tilde{v}'f(1) = -\tilde{v}'f(\alpha) = (1-\alpha^2)/2$ and $|\tilde{v}'f(x)| < (1-\alpha^2)/2$ for $x \in E - \{-1,\alpha,1\}$. There can be no better approximation $\tilde{\tilde{v}}'f(x)$ or else the quadratic $(\tilde{\tilde{v}}-\tilde{v})'f(x)$ has one root in $(-1,\alpha)$, one root in $(\alpha,1)$, and one root at $\tilde{x}$ which is impossible. Thus corollary 4.2.1(i) implies that $\text{Support}(\xi_0) = \{-1,\alpha,1\}$. That is, the effect of the design restriction is to replace the support point

0 by the point α. The design weights ρ_i are proportional to the $|L_i(x)|$ corresponding to $-1,\alpha,1$.

As pointed out by Kiefer and Wolfowitz ('65), the c-optimal design for $c = (0,\ldots,0,1)$ is the limit of the extrapolation designs as $|\tilde{x}| \to \infty$. Thus Support$(\xi_0) = \{-1,\alpha,1\}$ for $|\alpha| \le |\beta|$. The design weights obtained by solving the linear system of corollary 4.2.1(ii) are $\rho_0 = (1-\alpha)/4$, $\rho_1 = 1/2$, and $\rho_2 = (1+\alpha)/4$.

For $|\alpha| \ge |\beta|$, these solutions are merely reflected about the origin. For $|\alpha| = |\beta|$, any convex combination of the solutions for $|\alpha| \le |\beta|$ and for $|\alpha| \ge |\beta|$ is c-optimal.

For cubic regression on $E \subset [-1,1]$, the supports of the c-optimal designs for the highest coefficient and extrapolation problems will once again be the same. The following lemma will detail this support under the several relevant cases. The corresponding mass distributions are readily obtained by solving the linear system of corollary 4.2.1(ii).

Lemma 4.2.1: Let $\alpha_1 = \max([-1,-\frac{1}{2}] \cap E)$, let $\beta_1 = \min([-\frac{1}{2},1] \cap E)$, let $\alpha_2 = \max([-1,\frac{1}{2}] \cap E)$, and let $\beta_2 = \min([\frac{1}{2},1] \cap E)$. Also, let $w(y) = 2 + y - \sqrt{2(1+y)}$, let $r(y) = 2w(y) - y - 2$, and let $t(y,z) = [(\alpha_2-\alpha_1)(1-z)z - 2(1+y)]/[2(1+y) + (1-z)(\alpha_2-\alpha_1)]$.

i. If $\alpha_1 = \beta_1 = -\frac{1}{2}$ and $\alpha_2 = \beta_2 = \frac{1}{2}$, then the unrestricted solution that Support$(\xi_0) = \{-1, -\frac{1}{2}, \frac{1}{2}, 1\}$ is still c-optimal for the highest coefficient and extrapolation problems for cubic regression.

ii. If $\alpha_1 \le -\frac{1}{2} \le \beta_1 \le r(\alpha_1)$ and $w(\beta_1) \in E$, then Support$(\xi_0) = \{-1,\beta_1,w(\beta_1),1\}$.

iii. If $\alpha_1 \leq -\frac{1}{2} \leq r(\alpha_1) \leq \beta_1 \leq w(\alpha_1)$ and $w(\alpha_1) \in E$, then $\text{Support}(\xi_0) = \{-1, \alpha_1, w(\alpha_1), 1\}$.

iv. If $\alpha_1 \leq -\frac{1}{2} \leq w(\alpha_1) \leq \beta_1$, then $\text{Support}(\xi_0) = \{-1, \alpha_1, \beta_1, 1\}$.

v. If $\beta_2 \geq \frac{1}{2} \geq \alpha_2 \geq -r(-\beta_2)$ and $-w(-\alpha_2) \in E$, then $\text{Support}(\xi_0) = \{-1, -w(-\alpha_2), \alpha_2, 1\}$.

vi. If $\beta_2 \geq \frac{1}{2} \geq -r(\beta_2) \geq \alpha_2 \geq -w(-\beta_2)$ and $-w(-\beta_2) \in E$, then $\text{Support}(\xi_0) = \{-1, \beta_2, -w(-\beta_2), 1\}$.

vii. If $\beta_2 \geq \frac{1}{2} \geq -w(-\beta_2) \geq \alpha_2$, then $\text{Support}(\xi_0) = \{-1, \alpha_2, \beta_2, 1\}$.

viii. If $\alpha_1 \leq -\frac{1}{2} \leq t(\alpha_1, \alpha_2) \leq \beta_1$ and $\alpha_2 \leq \frac{1}{2} \leq -t(-\alpha_2, -\alpha_1) \leq \beta_2$, then $\text{Support}(\xi_0) = \{-1, \alpha_1, \alpha_2, 1\}$.

ix. If $\alpha_1 \leq -\frac{1}{2} \leq t(\alpha_1, \beta_2) \leq \beta_1$ and $\alpha_2 \leq -t(-\beta_2, -\alpha_1) \leq \frac{1}{2} \leq \beta_2$, then $\text{Support}(\xi_0) = \{-1, \alpha_1, \beta_2, 1\}$.

x. If $\alpha_1 \leq t(\beta_1, \alpha_2) \leq -\frac{1}{2} \leq \beta_1$ and $\alpha_2 \leq \frac{1}{2} \leq -t(-\alpha_2, -\beta_1) \leq \beta_2$, then $\text{Support}(\xi_0) = \{-1, \beta_1, \alpha_2, 1\}$.

xi. If $\alpha_1 \leq t(\beta_1, \beta_2) \leq -\frac{1}{2} \leq \beta_1$ and $\alpha_2 \leq -t(-\beta_2, -\beta_1) \leq \frac{1}{2} \leq \beta_2$, then $\text{Support}(\xi_0) = \{-1, \beta_1, \beta_2, 1\}$.

Proof: The proof will be sketched for cases iii. and viii. The others may be similarly verified.

iii. For this part, consider the possibility that $\alpha_1 \leq -\frac{1}{2} \leq \beta_1$ and that $\text{Support}(\xi_0) = \{-1, \alpha_1, w, 1\}$ for some $w \in (\beta_1, 1) \cap \text{Interior}(E)$. Since the supports of the c-optimal designs for the highest coefficient and extrapolation problems are the same, only the highest coefficient problem need be considered. Thus it is necessary to consider the cubic function $g(x) = x^3 + \tilde{v}_2 x^2 + \tilde{v}_1 x + \tilde{v}_0$ such that

$\max_E |g(x)| = \min_{v_0,v_1,v_2} \max_E |x^3 + v_2x^2 + v_1x + v_0| = B$. If Support($\xi_0$) = $\{-1,\alpha_1,w,1\}$ is to hold, then $|g(x)|$ must achieve its max at these support points and be smaller for all other $x \in E$. Thus $g(x) = -B + (x+1)(x-w)^2$. Manipulating the further requirements that $g(\alpha_1) = g(1) = B$ yields $w = 2 + \alpha_1 - \sqrt{2(1+\alpha_1)} = w(\alpha_1)$. It now remains only to determine the conditions under which $g(\beta_1) \leq B$. Manipulation of this inequality yields $\beta_1 \geq 2 + \alpha_1 - 2w(\alpha_1) = r(\alpha_1)$. Of course $\beta_1 \leq w(\alpha_1)$ if $w(\alpha_1) \in E$ is to hold.

This completes the proof of part iii. and indicates the validity of parts ii. and iv.

Parts v.-vii. may be obtained by reflecting the results of parts ii.-iv. about the origin, replacing α_1 by $-\beta_2$, and replacing β_1 by $-\alpha_2$.

viii. If Support(ξ_0) = $\{-1,\alpha_1,\alpha_2,1\}$ is to hold, then

$$\begin{aligned} g(x) &= B + (x-\alpha_1)(x-t_1)(x-1) \\ &= -B + (x+1)(x-\alpha_2)(x-t_2), \end{aligned}$$

where $\alpha_1 \leq t_1 \leq \beta_1$ and $\alpha_2 \leq t_2 \leq \beta_2$. Manipulation of these requirements yields $\beta_1 \geq t_1 = t(\alpha_1,\alpha_2)$ and $\beta_2 \geq t_2 = -t(-\alpha_2,-\alpha_1)$.

Parts ix.-xi. are proven similarly. Thus the lemma is proved.

The remainder of this section will be devoted to the consideration of a problem posed by Hoel ('65). The setting for this problem is that of polynomial regression ($f(x) = (1,x,\ldots,x^n)'$) on $E \subset [a,b]$, where $n \geq 2$ and $a, b \in E$. The goal is to extrapolate to an interior point $\tilde{x} \in [a,b] - E$; thus ξ_0 should be c-optimal with respect to $c = f(\tilde{x})$.

Now let $\alpha = \max([a,\tilde{x}) \cap E)$ and $\beta = \min((\tilde{x},b] \cap E)$. (Hoel assumes that $E = [-1,1] - (\alpha,\beta)$ but the problem is easily generalized.) The following developments correct an error of Hoel which can lead to the selection of a sub-optimal design; they should also serve to streamline the method of proof.

Lemma 4.2.2: There exists a unique polynomial $G(x) = \bar{v}'f(x)$ such that:

i. $|G(x)| \leq 1$ on E,

ii. $|G(x_i)| = 1$ for points $x_0 <\ldots< x_n$ from E,

iii. $x_r = \alpha$ and $x_{r+1} = \beta$ for some $r \in \{0,\ldots,n-1\}$,

iv. either $x_0 = a$ or $x_n = b$, and

v. $$\operatorname{sgn}[G(x_i)] = \begin{cases} (-1)^{r-i} & i \leq r \\ (-1)^{i-(r+1)} & i \geq r+1. \end{cases}$$

Proof: The proof of this lemma will be incorporated into the proof of theorem 4.2.1.

Theorem 4.2.1: Let the points $x_0 <\ldots< x_n$ be as in lemma 4.2.2 and the design weights ρ_i be proportional to $|L_i(x)| = |\prod_{\substack{j=0 \\ j\neq i}}^{n} (\tilde{x}-x_j)/(x_i-x_j)|$. Then the design $\xi_0 = \rho_0\delta_{x_0} +\ldots+ \rho_n\delta_{x_n}$ is c-optimal with respect to $c = f(\tilde{x})$. Furthermore, $\Delta(c,\xi_0) = d(\tilde{x},\xi_0) = G^2(\tilde{x})$.

Proof: The fact that the ρ_i are proportional to the $|L_i(\tilde{x})|$ will be immediate once the support of ξ_0 is confirmed. The proof of the theorem will now be set forward in several steps.

i. Let $g(x) = \tilde{v}'f(x)$ be the polynomial of corollary 4.2.1(i) such that $\tilde{v}'f(\tilde{x}) = 1$ and $B = \max_E |g(x)| = \max_E |\tilde{v}'f(x)| = \min_{v'c=1} \max_E |v'f(x)|$.

Then the c-optimal design ξ_0 must satisfy $\text{Support}(\xi_0) \subset \{x \mid |g(x)| = B\}$. Thus, if $x_0 < \ldots < x_k$ are the support points of ξ_0, corollary 4.2.1(ii) implies that

$$Bf(x) = \sum_{i=0}^{k} \varepsilon_i \rho_i f(x_i), \qquad (4.2.1)$$

where each $\varepsilon_i = \text{sgn}[g(x_i)]$ and each $\rho_i = \xi_0(\{x_i\})$. Equation (4.2.1) implies that $f(\tilde{x}), f(x_0), \ldots, f(x_k)$ are linearly dependent and hence that $k \geq n$. Also, as already remarked, a c-optimal design exists with $k \leq n$. Thus $k = n$ may be assumed. Further comment on the possibility that $k > n$ will follow the proof.

ii. It will now be shown that $g(x_i)$ alternates in sign as x_i moves away from $\tilde{x}$. Let $x_0 < \ldots < x_r \leq \alpha < \beta \leq x_{r+1} < \ldots < x_n$. (The next step of the proof will establish $0 \leq r \leq n - 1$.) Now (4.2.1) implies that

$$\varepsilon_i = \text{sgn}[\varepsilon_i \rho_i / B] = \text{sgn}\left[\frac{|f(x_0), \ldots, f(x_{i-1}), f(\tilde{x}), f(x_{i+1}), \ldots, f(x_n)|}{|f(x_0), \quad . \quad . \quad . \quad , f(x_n)|}\right]$$

$$= \begin{cases} (-1)^{r-i} & i \leq r \\ (-1)^{i-(r+1)} & i \geq r + 1, \end{cases}$$

according to Cramer's rule and the alternating property of the determinant which follows from definition 2.1.2.

iii. The next step of the proof is to establish that $x_0 \leq \alpha$ and $x_n \geq \beta$ (so that $0 \leq r \leq n - 1$). Suppose first that $x_n \leq \alpha$. Then $\text{sgn}[g(x_i)] = (-1)^{n-i}$ so that $g'(x)$ has at least $n - 1$ roots in (a, x_n).

Now note that the value of $C > 0$ may be chosen small enough so that the polynomial $g_0(x) = 1 - C(x-\tilde{x})^2$ has $\max_E |g_0(x)| < 1$. Hence $B < 1$. Then the inequalities $g(\alpha) \leq B$, $g(\tilde{x}) = 1 > B$, and $g(\beta) \leq B$ imply that $g'(x)$ has at least one root in (α,β). Thus $g'(x)$ is a non-trivial polynomial of degree $n - 1$ with at least $(n-1) + 1 = n$ roots. This contradiction implies that $x_n \geq \beta$. Similarly $x_0 \leq \alpha$.

iv. According to part iii., there exists $r \in \{0,\ldots,n-1\}$ such that $x_r \leq \alpha$ and $x_{r+1} \geq \beta$. It will now be shown that $x_r = \alpha$ and $x_{r+1} = \beta$ must hold. Suppose that $x_r < \alpha$. Then $g'(x)$ must have at least r roots in (a,x_r) and at least $n - r - 2$ roots in (β,b). Furthermore, $B < 1$ implies that $g'(x)$ has at least 2 roots in $(x_r,\tilde{x})$. Thus $g'(x)$ has at least n roots which is impossible. Therefore, $x_r = \alpha$ and, similarly, $x_{r+1} = \beta$.

v. In a similar fashion, either $x_0 = a$ or $x_n = b$ must hold. Otherwise, $g'(x)$ has at least $n - 1$ roots in $(a,\alpha) \cup (\beta,b)$ and at least one root in (α,β).

vi. At this point, it has been established that $G(x) = g(x)/B$ must satisfy the requirements of lemma 4.2.2. To show that the corresponding requirements on the support of ξ_0 are sufficient, suppose that there exists another polynomial $\tilde{g}(x)$ such that $\tilde{g}(\tilde{x}) = 1$ and $\max_E |\tilde{g}(x)| \leq B$. Then let $D(x) = g(x) - \tilde{g}(x)$. Note that $D(\tilde{x}) = 0$ and that the "sign" of $D(x_i)$ is ε_i (under the convention that the "sign" of 0 may be declared ± 1). This implies that $D(x)$ must have at least $n + 1$ roots so that $D(x) = 0$ on $[a,b]$. That is, $g(x)$ is the unique solution to the dual approximation problem.

vii. The previous part of the proof establishes that the polynomial $G(x)$ of lemma 4.2.2 does exist and that the conclusion of the theorem is necessary. To show that $G(x)$ is unique, suppose that there exists another polynomial $\tilde{G}(x)$ which satisfies the requirements of the lemma. Then $G(x) - \tilde{G}(x)$ must have at least $n + 1$ roots. This contradiction implies that $G(x)$ is unique and the conclusion of the theorem is sufficient.

viii. Note finally that $G^2(\tilde{x}) = 1/B^2 = d(\tilde{x},\xi_0)$. Thus the proof of the theorem and lemma is complete.

Several comments regarding this theorem are now in order. The error of Hoel ('65) is the conclusion that $x_0 = a$ and $x_n = b$. That this combined requirement can lead to selection of a sub-optimal design will be shown in examples 4.2.2 and 4.2.3. The former will show that the consequences of requiring both $x_0 = a$ and $x_n = b$ can be severe.

It should be noted that the proof of theorem 4.2.1 required only the assumption that $f_0(x) = 1, f_1,\ldots,f_n$ comprise a Tchebycheff system and that $f_1',\ldots,f_n'$ comprise a Tchebycheff system.

Note that the support of ξ_0 does not depend on the point $\tilde{x} \in (\alpha,\beta)$ of extrapolation. Of course the mass distribution of ξ_0 does depend on $\tilde{x}$. Also, the solution of theorem 4.2.1 is implicit; it requires the construction of the polynomial $G(x)$.

It may occur that $|G(x)| = 1$ for more than $n + 1$ points. In this case there will be a finite number of sequences $x_0 <\ldots< x_n$ satisfying lemma 4.2.2. Corresponding to each is a c-optimal design. Furthermore, any convex combination of these designs will be c-optimal. Such a situation will be illustrated by example 4.2.4.

Finally, the case that n = 1 has not been covered by theorem 4.2.1 and deserves a note. For linear regression, a trivial application of Elfving's theorem yields $\xi_0 = \rho_0\delta_a + \rho_1\delta_b$, where $\rho_0 = 1 - \rho_1 = (b-\tilde{x})/(b-a)$.

Example 4.2.2: Consider the setting of cubic regression on $E = [-1,0] \cup [.9,1]$ for the purpose of extrapolation to $\tilde{x} = .6$. Application of Hoel's result gives $\tilde{\xi} \approx .01\delta_{-1} + .05\delta_0 + .57\delta_{.9} + .37\delta_1$ and variance $d(.6,\tilde{\xi}) \approx 15.35$. In order to determine the optimal design, the polynomial G(x) of lemma 4.2.2 must be determined. In this case, it turns out that $G(x) \approx 1-5.54(x+1)x(x-.9)$. Furthermore, $\xi_0 \approx .10\delta_{-1} + .318\delta_{-.58} + .42\delta_0 + .17\delta_{.9}$ and $d(.6,\xi_0) \approx 6.75$. In this example, it is seen that requiring a, b, α, and $\beta \in \text{Support}(\xi)$ causes a severe inflation of the variance.

Example 4.2.3: The following example is treated by Hoel ('65). It turns out that the design proposed in the example is almost optimal. Let n = 5, let $E = [-1,0] \cup [.5,1]$, and let $\tilde{x} = .25$. If both points ± 1 are required to be in the support of ξ, then the best design has $\text{Support}(\tilde{\xi}) = \{-1, -.44, 0, .5, .82, 1\}$ and $d(.25,\tilde{\xi}) \approx 3.631$. It turns out that the optimal design has $\text{Support}(\xi_0) = \{-.93, -.44, 0, .5, .82, 1\}$ and $d(.25,\xi_0) \approx 3.627$. So in this example, $\tilde{\xi}$ only inflates the variance very slightly.

Example 4.2.4: Consider the setting of quadratic regression on $E = [-1,\alpha] \cup [-\alpha,1]$ for the purpose of extrapolation to $\tilde{x} = 0$. For this example, it is readily deduced that $G(x) = (1+\alpha^2-2x^2)/(1-\alpha^2)$.

Hence $G(-1) = -G(\alpha) = -G(-\alpha) = G(1) = -1$. Thus it is possible to choose Support$(\xi_0^{(1)}) = \{-1, \alpha, -\alpha\}$ or Support$(\xi_0^{(2)}) = \{\alpha, -\alpha, 1\}$. Furthermore,

$\xi_0^{(1)} = [2\alpha^2\delta_{-1} + (1-\alpha)\delta_\alpha + (1+\alpha)\delta_{-\alpha}]/2(1+\alpha^2)$ and

$\xi_0^{(2)} = [(1+\alpha)\delta_\alpha + (1-\alpha)\delta_{-\alpha} + 2\alpha^2\delta_1]/2(1+\alpha^2)$. Thus

$\xi_0 = [2\alpha^2(1-\gamma)\delta_{-1} + (1-\alpha+2\alpha\gamma)\delta_\alpha + (1+\alpha-2\alpha\gamma)\delta_{-\alpha} + 2\alpha^2\gamma\delta_1]/2(1+\alpha^2)$

is the general form of the c-optimal design for any $\gamma \in [0,1]$.

4.3 The (φ,ψ)-Problem

Recall that φ and ψ are finite measures on $\mathcal{X}$ such that $0 \le \varphi(x) < 1 < \psi(x)$, that $d\varphi \le d\psi$, and that $\Xi = \{\xi | d\varphi \le d\xi \le d\psi\}$. For this type of design restriction, examples of the application of theorem 4.1.1 will now be given. In order to apply the theorem, the following lemma will be of use.

Lemma 4.3.1: Let $\xi_0 \in \Xi$, let $\bar{B}$ = Support$(\xi_0 - \varphi)$, and let $\underline{B}$ = Support$(\psi - \xi_0)$. Then $\int_{\mathcal{X}} [\tilde{v}'f(x)]^2 d\xi_0(x) = \max_{\xi \in \Xi} \int_{\mathcal{X}} [\tilde{v}'f(x)]^2 d\xi(x)$ if and only if $\min_{\bar{B}} |\tilde{v}'f(x)| \ge \max_{\underline{B}} |\tilde{v}'f(x)|$.

Proof: The proof proceeds exactly as that of theorem 3.3.1 except that $[\tilde{v}'f(x)]^2$ now plays the role of $d(x,\xi_0)$.

In the setting of linear regression on $\mathcal{X} = [-1,1]$, it is readily seen that the D-optimal design will also be c-optimal for estimating $\theta_0(c = (1,0)')$ or $\theta_1(c = (0,1)')$. For $d\varphi = 0$ and $d\psi(x) = \gamma dx$, the design is specified in section 3.3. The following example of quadratic regression yields more interesting results.

Example 4.3.1: Let $\mathcal{X} = [-1,1]$, let $f(x) = (1,x,x^2)'$, let $d\varphi = 0$, and let $d\psi(x) = \gamma dx$ for $\gamma \geq 1/2$. For estimating any one of the regression coefficients, the problem is symmetrical with respect to reflection about the origin. Hence it is sufficient to consider only symmetrical designs. Thus the variance of $\hat{\theta}_1$ is proportional to $M_{11}^{-1} = 1/\mu_2$ and so the D-optimal design for linear regression (section 3.3) is optimal for estimation of θ_1.

Consider now the problem of estimating θ_0 in the same setting. By analogy with the unrestricted optimal design (a point mass at 0), it is tempting to guess that $d\tilde{\xi}(x) = \begin{cases} \gamma dx & -1/2\gamma \leq x \leq 1/2\gamma \\ 0 & \text{otherwise} \end{cases}$

is c-optimal with respect to $c = (1, 0, 0)'$. However, it may be confirmed that this design does not satisfy the condition of theorem 4.1.1. Suppose instead that the c-optimal design has the form

$$d\xi_0(x) = \begin{cases} 0 & w < |x| < y \\ \gamma dx & \text{otherwise.} \end{cases}$$

Here $y - w = 1 - 1/2\gamma = \kappa$ must hold in order that $\xi_0([-1,1]) = 1$. Now note that because of the symmetry of this problem,

$$\min_{v'c=1} \int_{\mathcal{X}} [v'f(x)]^2 d\xi_0(x) = \min_{v_2} \int_{-1}^{1} (1+v_2x^2)^2 d\xi_0(x) = \min_{v_2} [1+2v_2\mu_2^{(0)}+v_2^2\mu_4^{(0)}],$$

where each $\mu_i^{(0)} = \int_{-1}^{1} x^i d\xi_0(x)$. This expression is minimized by

$\tilde{v}_2 = -\mu_2^{(0)}/\mu_4^{(0)}$. Note also that the only way that lemma 4.3.1 can hold is that $(1+\tilde{\tilde{v}}_2w^2)^2 = (1+\tilde{v}_2y^2)^2$. This implies that $\tilde{v}_2 = -2/(w^2+y^2)$. Thus the condition of theorem 4.1.1 will be satisfied and ξ_0 will be c-optimal if $\frac{2}{w^2+y^2} = -\tilde{V}_2 = \frac{\mu_2^{(0)}}{\mu_4^{(0)}} = \frac{5(1+w^3-y^3)}{3(1+w^5-y^5)}$.

Applying the condition that $y = w + \kappa$ to this equation yields the requirement that

$$5(2+\kappa^3)(w^2+\kappa w) + (\kappa^5+5\kappa^2-6) = 0. \qquad (4.3.1)$$

Thus if w is a solution to (4.3.1), then ξ_0 is c-optimal for estimating θ_0. It is readily confirmed that (4.3.1) has one real root in $[0,1/2\gamma]$. Table 4.3.1 displays the properties of ξ_0 for various values of γ.

Table 4.3.1: θ_0-Optimal Design for Quadratic Regression on [-1,1] under $d\psi(x)=\gamma dx$

γ	w	y	$\xi_0([-w,w])$	$\xi_0([-1,-y) = \xi_0([y,1])$
.5	.7746	.7746	.7746	.1127
.6	.6859	.8525	.8230	.0885
.7	.6141	.8998	.8598	.0701
.8	.5545	.9295	.8872	.0564
.9	.5043	.9488	.9078	.0461
1.0	.4618	.9618	.9236	.0382
1.5	.3214	.9881	.9642	.0179
2.0	.2249	.9949	.9796	.0102
3.0	.1651	.9985	.9910	.0045
5.0	.0997	.9997	.9970	.0015
∞	0	1	1	0

Observe that the limiting cases of $\gamma = \frac{1}{2}$ and $\gamma \to \infty$ yield (respectively) $\xi_0 = \psi$ and $\xi_0 = \delta_0$ as they should.

It might be noted that this example shows that corollary 4.2.1(ii) need not hold for general design restrictions. For $\gamma = \frac{1}{2}$, $\kappa = 0$ holds and so $w = \sqrt{3/5}$ results from (4.3.1). If

$$s(x) = \text{sgn}[\tilde{v}'f(x)] = \begin{cases} +1 & |x| \leq w \\ -1 & |x| \geq y \end{cases}$$

then $\int_{-1}^{1} x^2 s(x) d\xi_0(x) = (w^3+y^3-1)/3 = (2w^3-1)/3 \neq 0$. Thus orthogonality to $f_2(x) = x^2$ does not hold.

Consider now the problem of estimating θ_2 in the same setting. The unrestricted optimal design for $c = (0, 0, 1)'$ is $d\bar{\xi}(x) = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$. This suggests that the restricted optimal design may have the form $d\xi_0(x) = \begin{cases} 0 & w < |x| < y \\ \gamma dx & \text{otherwise.} \end{cases}$

Here $y - w = 1 - 1/2\gamma = \kappa$ so that $\xi_0([-1,1]) = 1$. Due to the symmetry of the problem,

$$\min_{v'c=1} \int_{\mathcal{X}} [v'f(x)]^2 d\xi_0(x) = \min_{v_0} \int_{-1}^{1} (x^2+v_0)^2 d\xi_0(x) = \min_{v_0} [\mu_4^{(0)}+2v_0\mu_2^{(0)}+v_0^2],$$

which is minimized by $\tilde{v}_0 = -\mu_2^{(0)}$. Note also that the condition of lemma 4.3.1 will be fulfilled only if $(w^2+\tilde{v}_0)^2 = (y^2+\tilde{v}_0)^2$. Thus $\tilde{v}_0 = -(w^2+y^2)/2$. Hence theorem 4.1.1 implies that ξ_0 will be c-optimal if $\frac{w^2+y^2}{2} = -\tilde{v}_0 = \mu_2^{(0)} = \frac{2\gamma}{3}(1+w^3-y^3)$. Applying the condition $y = w + \kappa$ to this expression yields the requirement

$$6w^2 + 6\kappa w - (\kappa^3-3\kappa^2+2) = 0. \qquad (4.3.2)$$

Thus ξ_0 is c-optimal if w solves (4.3.2). Table 4.3.2 displays the properties of ξ_0 for various values of γ.

Table 4.3.2: θ_2-Optimal Design for Quadratic Regression on [-1,1] under $d\psi(x)=\gamma dx$

γ	w	y	$\xi_0([-w,w])$	$\xi_0([-1,-y]) = \xi_0([y,1])$
.5	.5774	.5774	.5774	.2113
.6	.4886	.6553	.5864	.2068
.7	.4200	.7057	.5880	.2060
.8	.3665	.7415	.5864	.2068
.9	.3242	.7686	.5834	.2083
1.0	.2901	.7901	.5802	.2099
1.5	.1878	.8545	.5634	.2183
2.0	.1379	.8879	.5516	.2242
3.0	.0895	.9228	.5368	.2316
5.0	.0523	.9523	.5230	.2385
10	.0256	.9756	.5120	.2440
100	.0025	.9975	.5012	.2494
∞	0	1	1/2	1/4

Observe that the limiting cases of $\gamma = \frac{1}{2}$ and $\gamma \to \infty$ yield (respectively) $\xi_0 = \psi$ and $\xi_0 = (\delta_{-1} + 2\delta_0 + \delta_1)/4$ as they should.

4.4 The p-Problem

Recall that for the p-problem, $\{G_\omega | \omega \in \Omega\}$ is a collection of disjoint sets which are open in $\mathcal{X}$ and each $p_\omega \in (0,1)$. Also, $\Xi = \{|\xi(G_\omega) \leq p_\omega;\ \omega \in \Omega\}$. The following corollary of theorem 4.1.1 is appropriate to the present restriction in the case that $\Omega = \{1,\ldots,s\}$. It employs the notation that $G_0 = \mathcal{X} - \bigcup_{i=1}^{s} G_i$ and $p_0 = 1$.

Corollary 4.4.1: Let ξ_0 and $\tilde{v}$ be as in theorem 4.4.1. Let $m_i = \sup_{G_i} |\tilde{v}'f(x_i)|$ for $i = 0,\ldots,s$. Also, let $m_{(0)} \geq m_{(1)} \geq \cdots \geq m_{(s)}$ denote their ordered values and let each $p_{(i)}$ and $G_{(i)}$ correspond to the ordered value $m_{(i)}$. Finally, let $r \in \{0,\ldots,s\}$ be the largest integer such that $\sum_{i=0}^{r} p_{(i)} \leq 1$. Then

$$\max_{\xi \in \Xi_{\mathcal{X}}} \int [\tilde{v}'f(x)]^2 d\xi(x) = \sum_{i=0}^{r} p_{(i)} m_{(i)}^2 + [1 - \sum_{i=0}^{r} p_{(i)}] m_{(r+1)}^2.$$

Furthermore, $\text{Support}(\xi_0) \subset \bigcup_{i=0}^{r+1} \{x \mid |\tilde{v}'f(x)| = m_{(i)}\}$.

Proof: The first conclusion is merely an application of the notation established by the corollary. The second conclusion is an immediate consequence of theorem 4.1.1.

One consequence of this theorem is that if $m_i < m_0$, then $\xi_0(G_i) = 0$. Also, if $\tilde{v}'f(x)$ does not attain m_i on G_i, then $\xi_0(G_i) = 0$. Examples of the application of theorem 4.1.1 will now be given.

Example 4.4.1: Consider the setting of quadratic regression on $[-1,1]$. Also, let $G_1 = (\alpha,-\alpha)$ and $0 < p_1 < 1$. For estimation of θ_1, the unrestricted optimal design $\xi_0 = \frac{1}{2}(\delta_{-1} + \delta_1) \in \Xi$ and hence is c-optimal under the design restriction.

Consider now the problem of estimating θ_0 in the same setting. By analogy with the unrestricted optimal design, it is tempting to guess that $\tilde{\xi} = p_1\delta_0 + (1-p_1)(\delta_\alpha + \delta_{-\alpha})/2$ is c-optimal for $c = (1, 0, 0)'$. However, it may be confirmed that this design does not satisfy the condition of theorem 4.1.1. Suppose instead that the c-optimal design has the form $\xi_0 = p_1\delta_0 + (1-p_1)[\gamma(\delta_\alpha + \delta_{-\alpha}) + (1-\gamma)(\delta_1 + \delta_{-1})]/2$. As in example 4.3.1, $\tilde{v} = (1, 0, \tilde{v}_2)'$ where $\tilde{v}_2 = -\mu_2^{(0)}/\mu_4^{(0)}$. Also, the only way that $\max_{\xi} \int_{-1}^{1}(1+\tilde{v}_2x^2)^2 d\xi(x) = \int_{-1}^{1}(1+\tilde{v}_2x^2)^2 d\xi_0(x)$ can hold is that $(1+\tilde{v}_2\alpha^2)^2 = (1+\tilde{v}_2 1^2)^2$. This implies that $\tilde{v}_2 = -2/(1+\alpha^2)$. Thus the condition of theorem 4.1.1 will be satisfied if

$$\frac{2}{1+\alpha^2} = -\tilde{v}_2 = \frac{\mu_2^{(0)}}{\mu_4^{(0)}} = \frac{1-\gamma(1-\alpha^2)}{1-\gamma(1-\alpha^4)} .$$

This implies that $\gamma = 1/(1+\alpha^2)$ in order for ξ_0 to be optimal for estimating θ_0.

Consider finally the problem of estimating θ_2 for the same example. If $p_1 \geq 1/2$, then the unrestricted optimal design $\xi_0 = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1 \in \Xi$ and hence is c-optimal under the design restriction. For $p_1 < \frac{1}{2}$, there are two relevant cases for estimating θ_2.

Suppose first that $\alpha^2 \geq 1 - 2p_1$. For this case, it is proposed that $\xi_0 = p_1\delta_0 + (1-p_1)(\delta_{-1}+\delta_1)/2$. As in example 4.3.1, $\tilde{v}_0 = -\mu_2^{(0)} = p_1 - 1$. Now $\sup_{G_1} |\tilde{v}'f(x)| = |\tilde{v}'f(0)| = 1 - p_1$. Also, because $\alpha^2 \geq 1 - 2p_1$, $\max_{G_0} |\tilde{v}'f(x)| = |\tilde{v}'f(\pm 1)| = p_1 < 1 - p_1$. Thus

$$\max_{\xi\in\Xi} \int_{\mathcal{X}} [\tilde{v}'f(x)]^2 d\xi(x) = p_1(1-p_1)^2 + (1-p_1)p_1^2 = p_1(1-p_1) = \int_{\mathcal{X}} [\tilde{v}'f(x)]^2 d\xi_0(x).$$

Hence theorem 4.1.1 implies that ξ_0 is c-optimal.

The remaining case is that $\alpha^2 \leq 1 - 2p_1$. For this case, it is proposed that $\xi_0 = p_1\delta_0 + (1-p_1)[\gamma(\delta_\alpha + \delta_{-\alpha}) + (1-\gamma)(\delta_{-1} + \delta_1)]/2$. Then $\tilde{v}_0 = -\mu_2^{(0)} = (p_1-1)[1-\gamma(1-\alpha^2)]$. Now

$$\max_{\xi} \int_{-1}^{1}(x^2+\tilde{v}_0)^2 d\xi(x) = \int_{-1}^{1}(x^2+\tilde{v}_0)^2 d\xi_0(x) \text{ only if } (\alpha^2+\tilde{v}_0)^2 = (1+\tilde{v}_0)^2.$$

Thus $\tilde{v}_0 = -(1+\alpha^2)/2$. Hence theorem 4.1.1 implies that ξ_0 will be c-optimal if $(1+\alpha^2)/2 = -\tilde{v}_0 = (1-p_1)[1-\gamma(1-\alpha^2)]$. That is, $\gamma = (1-2p_1-\alpha^2)/2(1-\alpha^2)(1-p_1)$. It is readily seen that $0 \leq \gamma \leq 1$ and hence that ξ_0 is optimal for estimating θ_2 when $\alpha^2 \leq 1 - 2p_1$.

4.5 The Marginal Restriction Problem

Recall that for the marginal restriction problem, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and a measure ξ_1^* on $\mathcal{X}_1$ is prescribed. Also, $\Xi = \{\xi | \xi(A\times\mathcal{X}_2) = \xi_1^*(A)$ for all Borel sets $A \subset \mathcal{X}_1\}$. The following corollary of theorem 4.1.1 is appropriate to the present restriction.

Corollary 4.5.1: ξ_0 is c-optimal for the marginal restriction problem if and only if there exists $\tilde{v}$ such that $\tilde{v}'c=1$ and

$$\int_{\mathcal{X}_1} \max_{\mathcal{X}_2}[\tilde{v}'f(x_1,x_2)]^2 d\xi_1^*(x_1) = \int_{\mathcal{X}_1}\int_{\mathcal{X}_2} [\tilde{v}'f(x_1,x_2)]^2 d\xi_0(x_2|x_1)d\xi_1^*(x_1)$$

$$= \min_{v'c=1} \int_{\mathcal{X}_1}\int_{\mathcal{X}_2} [v'f(x_1,x_2)]^2 d\xi_0(x_2|x_1)d\ {}_1^*(x_1).$$

Proof: The corollary is merely a translation of theorem 4.1.1 with $d\xi_0(x_1,x_2) = d\xi_0(x_2|x_1)d\xi_1^*(x_1)$ and

$$\max_{\xi\in\Xi} \int_{\mathcal{X}_1}\int_{\mathcal{X}_2} [\tilde{v}'f(x_1,x_2)]^2 d\xi(x_2|x_1)d\xi_1^*(x_1)$$

$$= \int_{\mathcal{X}_1} \max_{\mathcal{X}_2}[\tilde{v}'f(x_1,x_2)]^2 d\xi_0(x_2|x_1)d\xi_1^*(x_1).$$

The following example will illustrate the use of corollary 4.5.1.

Example 4.5.1: Let $\mathcal{X} = [-1,1] \times [-1,1]$ and $f(x_1,x_2) = (1,x_1,x_1^2,x_2^2,x_2,x_1x_2)'$. For purposes of estimating the interaction coefficient, $c = (0,0,0,0,0,1)'$. In the unrestricted setting that $\Xi = \Xi_0$, the c-optimal design obtained by Kiefer and Wolfowitz ('59) is $d\tilde{\xi}(x_1,x_2) = d\bar{\xi}(x_1)d\bar{\xi}(x_2)$, where $\bar{\xi} = \frac{1}{2}(\delta_{-1}+\delta_1)$. Under the marginal restriction, it will now be shown that $d\xi_0(x_1,x_2) = d\xi_1^*(x_1)d\bar{\xi}(x_2)$ is c-optimal.

It will be convenient to employ the notation that $f_0(x_1,x_2) = (1,x_1,x_1^2,x_2^2)'$ and $v_0 \in \mathbb{R}^4$. Thus $f(x_1,x_2) = [f_0(x_1,x_2)', x_2, x_1x_2]'$ and $v = (v_0',y,1)'$ for any v such that $v'c=1$. Now

$$\min_{v'c=1} \int_{\mathcal{X}_1}\int_{\mathcal{X}_2} [v'f(x_1,x_2)]^2 d\xi_0(x_2|x_1)d\xi_1^*(x_1)$$

$$= \min_{v_0,y} \int_{-1}^{1}\int_{-1}^{1}\{v_0'f_0(x_1,x_2)f_0(x_1,x_2)'v_0 + y^2x_2^2 + 2yx_1x_2^2 + x_1^2x_2^2$$

$$+ 2yv_0'f_0(x_1,x_2)x_2 + 2v_0'f_0(x_1,x_2)x_1x_2\}d\xi_0(x_2|x_1)d\xi_1^*(x_1)$$

$$\geq \min_{y} \int_{-1}^{1}\int_{-1}^{1}\{0 + x_2^2(y+x_1)^2 + 0 + 0\}d\xi_0(x_2|x_1)d\xi_1^*(x_1)$$

$$= \mu_2 - \mu_1^2,$$

where each $\mu_i = \int_{-1}^{1} x^i d\xi_1^*(x_1)$. The inequality results because $v_0'f_0f_0'v_0 \geq 0$. Also, the components of $f_0(x_1,x_2)$ are orthogonal to x_2 and to x_1x_2 with respect to ξ_0. It is also seen that

$\tilde{v} = (0,0,0,0,-\mu_1,1)'$. Thus

$$\int_{\mathcal{X}_1} \max_{\mathcal{X}_2}[\tilde{v}'f(x_1,x_2)]^2 d\xi_1^*(x_1) = \int_{-1}^{1} \max_{-1\leq x_2\leq 1}[-\mu_1 x_2+x_1 x_2]^2 d\xi_0(x_2|x_1)d\xi_1^*(x_1)$$

$$= \mu_2 - \mu_1^2$$

so that the condition of corollary 4.5.1 is satisfied and ξ_0 is c-optimal for this example.

BIBLIOGRAPHY

BIBLIOGRAPHY

Cook, R. D. and Thibodeau, L. A. (1980). Marginally restricted D-optimal designs. JASA, Vol. 75, 366-371.

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Statist., Vol. 23, 255-262.

Federov, V. V. (1972). Theory of Optimal Experiments. Translated by Studden, W. J. and Klimko, E. M., Academic Press, New York.

de la Garza, A. (1954). Spacing of information in polynomial regression. Ann. Math. Statist., Vol. 25, 123-130.

Guest, P. G. (1958). The spacing of observations in polynomial regression. Ann. Math. Statist., Vol 29, 294-299.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Statist., Vol. 29, 1134-1145.

Hoel, P. G. (1965). Optimum designs for polynomial extrapolation. Ann. Math. Statist., Vol. 36, 1483-1493.

Hoel, P. G. and Levine, A. (1964). Optimal spacing and weighing in polynomial regression. Ann. Math. Statist., Vol. 35, 1553-1560.

Karlin, S. (1959). Mathematical Methods and Theory in Games, Programming, and Economics, Vol. II. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

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

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

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser. B, Vol. 21, 273-319.

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

Kiefer, J. and Wolfowtiz, J. (1960). The equivalence to two extremum problems. _Can. J. Math._, Vol. 12, 363-366.

Kiefer, J. and Wolfowitz, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. _Ann. Math. Statist._, Vol 36, 1627-1655.

Krein, M. G. and Nudelman, A. A. (1977). _The Markov Moment Problem and Extremal Problems_. Translated by Louvish, D., American Mathematical Society, Providence, Rhode Island.

Studden, W. J. (1968). Optimal designs on Tchebycheff points. _Ann. Math. Statist._, Vol. 39, 1435-1447.

Welch, W. (1981). Branch and bound search for experimental designs based upon D-optimality and other criteria. _Technometrics_ (to appear).

Wynn, H. P. (1977). Optimum designs for finite population sampling. Gupta, S. S. and Moore, D. S. (Ed.). _Statistical Decision Theory and Related Topics II_. Academic Press, New York.

Wynn, H. P. (1982). Optimum submeasures with applications to finite population sampling. Gupta, S. S. and Berger, J. O. (Ed.). _Statistical Decision Theory and Related Topics III_. Academic Press, New York (to appear).

VITA

VITA

Kenneth Constantine was born on February 22, 1955 in Boston, Massachusetts to Gregory and Barbara Constantine. He received his Bachelor of Science degree in Mathematics and Physics from Eastern Nazarene College in May, 1977. He was married to Laura Ferreira in July, 1977. He received the Master of Science degree in Mathematical Statistics from Purdue University in August 1979.