# ROBUST DESIGNS FOR NEARLY LINEAR REGRESSION

by

W. Notz

Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-11

July 1979

Research sponsored in part by the National Science Foundation, NSF Grant MCS75-22481 A02.

# ROBUST DESIGNS FOR NEARLY LINEAR REGRESSION

by W. Notz[1)]

Purdue University

Summary. In this paper we seek designs and estimators which are optimal in some sense for multivariate linear regression on cubes and simplexes when the true regression function is unknown. More precisely, we assume the unknown true regression function is the sum of a linear part plus some contamination orthogonal to the set of all linear functions in the $L_2$ norm with respect to Lebesgue measure. The contamination is assumed bounded in absolute value and it is shown that the usual designs for multivariate linear regression on cubes and simplexes and the usual least squares estimators minimize the supremum over all possible contaminations of the expected mean square error.

Key Words: Multivariate linear regression, robustness, optimum designs, least squares estimates, $L_2$ norm, cubes, simplexes.

AMS Subject Classifications: Primary - 62J05, 62K05, Secondary - 62G35.

[1)] Research sponsored in part by the National Science Foundation, NSF Grant MCS75-22481 A02.

1. Introduction.

Consider the regression design problem given by

$$y(x_i) = f(x_i) + e_i \quad , i = 1, 2, \ldots, n$$

where the $\{e_i\}$ are uncorrelated random variables with mean 0 and variance $\sigma^2$. The $x_i$ are elements of a compact subset X of a Euclidean space, and f is a real-valued function on X from a class $F_0$. $F_0$ is typically composed of linear combinations of specified functions $f_0, f_1, \ldots, f_k$. The regression problem is concerned with making some inference about the unknown coefficients of these specified $f_j$ and the associated design problem is to choose the $x_i$ in an optimal manner for this inference. Many papers have been addressed to this problem. Box and Draper (1959) have discussed some of the dangers inherent in a strict formulation of $F_0$ which ignores the possibility that the true f may only be approximated by an element of $F_0$, e.g., in estimation there may result a large bias term. A careful description of some problems in this context is given by Kiefer (1973) in the case where the class of possible functions f, F, is a finite dimensional space containing $F_0$.

In a related direction Huber (1975) formulated a problem where $X = [-\frac{1}{2}, +\frac{1}{2}]$, $F_0$ = {linear functions on X}, and $F = \{f(x) = a + bx + g(x);$ $\inf_{\alpha,\beta} \int_{-\frac{1}{2}}^{+\frac{1}{2}} (g(x) - \alpha - \beta x)^2\, dx = \int_{-\frac{1}{2}}^{+\frac{1}{2}} g^2(x)\, dx \leq c\}$. $c > 0$ is a given constant. Notice that if $f \in F$ then $a + bx$ is the best linear approximation to f in the $L_2$ norm wih respect to Lebesgue measure on $[-\frac{1}{2}, +\frac{1}{2}]$. Huber confines himself to the use of the standard least squares estimates based on the model $F_0$ and finds the design which minimizes the maximum risk

$$\sup_{f \in F} E \int_{-\frac{1}{2}}^{+\frac{1}{2}} (\hat{a} + \hat{b}x - f(x))^2\, dx.$$

Unfortunately this formulation leads to the restriction that the designs must be absolutely continuous with respect to Lebesgue measure, otherwise the maximum risk above is infinite. This means no implementable design can have finite maximum risk.

In a similar spirit is some work by Marcus and Sacks (1976). They take $X = [-1, +1]$, $F_0$ = {linear functions on X}, and $F = \{f(x) = a + bx + g(x); |g(x)| \leq \phi(x)\}$. $\phi(x)$ is a given function on X with $\phi(0) = 0$. For $f \in F$ the contamination $g(x)$ may be thought of as the remainder term in a first order Taylor expansion of f. Marcus and Sacks restrict the estimators of a and b to be linear but not necessarily the standard least squares estimates based on the model $F_0$, and restrict designs to have finite support. They look for estimates and designs to minimize the mean square error

$$\sup_{f \in F} E(\hat{a} - a)^2 + \theta^2(\hat{b} - b)^2)$$

where $\hat{a}$ and $\hat{b}$ denote the estimates of a and b, and $\theta$ is a specified constant.

They are able to solve this problem for a number, but not all, choices of $\phi$. If $\phi(x) \geq mx$ then the unique optimal design is on the points $\{-1, 0, +1\}$. If $\phi$ is convex there is a wide range of cases for which a design can be found on two points $\{-z, +z\}$ where z depends on $\phi$ and $\theta$.

It should be noted that the condition $\phi(0) = 0$ in this formulation forces the contamination $g(x)$ to be zero at $x = 0$. This fact gives special value to the point 0 and is the reason that 0 is in the support of the unique optimal design in the case $\phi(x) \geq mx$.

In this paper some of the clever ideas of Marcus and Sacks and of Huber are modified and combined to get results in some multivariate settings. More specifically we take X to be the k-fold Cartesian product of [-1, +1] or to be the k-dimensional simplex. We take $F_0$ = {linear functions on X}, $F = \{f(\underline{x}) = \beta_0 + \underline{\beta}'x + g(\underline{x});\ \beta_0 \in R$ (0 when X is the simplex), $\underline{\beta} \in R^k$, $\underline{x} \in X$, $g: X \to R$, $|g(\underline{x})| \le c$, and $\inf \int_X (g(x) - b_0 - \underline{b}'x)^2\, dx = \int_X g^2(\underline{x})\, d\underline{x}$, where the inf is over all $b_0 \in R$ and $\underline{b} \in R^k$ (when X is the simplex, this is just over all $\underline{b} \in R^k$)}. Here $c \ge 0$ is some constant. Notice for $f \in F$, $\beta_0 + \underline{\beta}'\underline{x}$ is the best linear approximation to f in the $L_2$ norm with respect to Lebesgue measure on X. If estimates $\hat{\beta}_0$ and $\hat{\underline{\beta}}$ of $\beta_0$ and $\underline{\beta}$ are restricted to be linear but not necessarily the standard least squares estimates, and designs are restricted to have finite support, then the estimates and design which minimize the mean square error

$$\sup_{f \in F} E((\hat{\beta}_0 - \beta_0)^2 + \sum_{i=1}^{k} \theta_i^2\, (\hat{\beta}_i - \beta_i)^2)$$

where $\underline{\beta} = (\beta_1, \ldots, \beta_k)'$ and $0 \le \theta_i \le 1$ for $i = 1, \ldots, k$, are the usual least squares estimates and the usual optimal designs for multivariate linear regression on the cube or simplex (in the case of the simplex we take all $\theta_i = 1$).

## 2. Results for simplexes.

The following notation will be used in this paper. Lines underneath variables will denote column vectors and primes on vectors or matrices will denote transposes. The size of a given vector or matrix will be made clear from the context. We shall use the symbol R to denote the real line and $R^k$ to denote k-dimensional Euclidean space.

Consider the multivariate regression problem

(2.1) $$Y(\underline{x}_m) = \underline{\beta}'\underline{x}_m + g(\underline{x}_m) + e_m$$

where $m = 1, 2, \ldots, n$ (n is fixed), the $e_m$ are uncorrelated random variables with mean 0 and finite variance $\sigma^2 > 0$, $\underline{\beta} = (\beta_1, \ldots, \beta_k)' \varepsilon R^k$,

$S_k = \{(x_1, \ldots, x_k)' \varepsilon R^k; \sum_{i=1}^{k} x_i = 1, x_i \geq 0$ for all $i\} = k - 1$ dimensional simplex, $\underline{x}_m = (x_{1m}, \ldots, x_{km})' \varepsilon S_k$, and for some fixed constant $c > 0$,

$g \varepsilon G = \{g: S_k \to R; |g(\underline{x})| \leq c, \int g(\underline{x}) d\underline{x} = \int x_i\, g(\underline{x})\, d\underline{x} = 0$ for $i = 1, \ldots, k$.

Here $\underline{x} = (x_1, \ldots, x_k)'\}$.

In the definition of $G$ $d\underline{x}$ is Lebesgue measure on $S_k$ and all integrals are over $S_k$. In fact all integrals that appear in this section of the paper will be assumed to be over $S_k$ unless otherwise noted.

The conditions $\int g(\underline{x})\, d\underline{x} = \int x_i g\,(\underline{x})\, d\underline{x} = 0$ are equivalent to requiring $\int g^2(\underline{x}) d\underline{x} = \inf_{\underline{b} \varepsilon R^k} \int (g(\underline{x}) - \underline{b}'\underline{x})^2\, d\underline{x}$, which says that the best linear approximation of $g$ in the $L_2$ norm with respect to Lebesgue measure $d\underline{x}$ on $S_k$ is the function 0. This condition insures the uniqueness of the $\beta_i$ in our model (2.1).

A discrete porbability measure $\xi$ on $S_k$ will be called a p-exact design for p observations if $\xi(\underline{x}) = j(\underline{x})/p$ where $p > 0$ and $j(\underline{x})$ are integers, and $\underline{x} \varepsilon S_k$. We shall denote by $\Xi_p$ the class of all such designs.

We also define

$D_p = \{$probability measures $\xi$ on $S_k$; card (supp $\xi) \leq p\}$, $D = \bigcup_{p=1}^{\infty} D_p$.

For $\xi \varepsilon \Xi_n$, let the linear estimators of the $\beta_i$ be defined by

(2.2)
$$\hat{\beta}_i = \int Y(\underline{x})\; B_i(\underline{x})\; d\xi\; (\underline{x}) \quad , i = 1, \ldots, k$$

where the $B_i$ are real valued functions on $S_k$.

We shall consider the (expected) mean square error due to the design $\xi$ and the estimators $\hat{\beta}_i$ namely

(2.3)
$$\sum_{i=1}^{k} E(\hat{\beta}_i - \beta_i)^2 .$$

This mean square error can be rewritten as the sum of a variance term, $\sum_{i=1}^{k} E(\hat{\beta}_i - E\hat{\beta}_i)^2$, and a bias term, $\sum_{i=1}^{k} (\beta_i - E\hat{\beta}_i)^2$. Using (2.2) we can write the variance term as

(2.4) $$\sum_{i=1}^{k} E(\hat{\beta}_i - E\hat{\beta}_i)^2 = (\sigma^2/n) \sum_{i=1}^{k} \int B_i^2 (\underline{x})\, d\xi (\underline{x})$$

and the bias term is determined by the equations

(2.5) $$E\hat{\beta}_j - \beta_j = \sum_{i=1, i\neq j}^{k} \beta_i \int x_i B_j (\underline{x})\, d\xi (\underline{x}) + \beta_j [\int x_j B_j (\underline{x})\, d\xi (\underline{x}) - 1] + \int B_j(\underline{x})\, g(\underline{x})\, d\xi (\underline{x})$$

for j = 1, ..., k.

If the $\beta_i$ are unbounded in order for the error to be bounded we must have

(2.6) $$\int x_i B_j (\underline{x})\, d\xi (\underline{x}) = \delta_{ij} , \quad 1 \leq i, j \leq k$$

where $\delta_{ij}$ is the Kronecker delta. This is equivalent to saying that the linear estimators are unbiased if g = 0.

Also notice that since $\underline{x} = (x_1, \ldots, x_k)' \varepsilon S_k$ if and only if $\sum_{i=1}^{k} x_i = 1$ and $x_i \geq 0$ for all i, (2.6) implies

(2.7) $$\int B_j(\underline{x})\, d\xi (\underline{x}) = \int (\sum_{i=1}^{k} x_i) B_j (\underline{x})\, d\xi (\underline{x}) = 1$$

for j = 1, ..., k.

Let $\rho = \sigma^2/n$, $\underline{B} = (B_1, \ldots, B_k)'$, and define

(2.8) $$L(\underline{B}, \xi, g) = \sum_{i=1}^{k} (\int B_i(\underline{x})\, g(\underline{x})\, d\xi(\underline{x}))^2 + \rho \int (\sum_{i=1}^{k} B_i^2(\underline{x}))\, d\xi(\underline{x})$$

Notice that $L(\underline{B}, \xi, g)$ is equal to (2.3) with condition (2.6) imposed. Condition (2.6) and $L(\underline{B}, \xi, g)$ are well defined for $\xi \in D$ and from now on we shall not restrict $\xi$ to be an exact design for a particular p, but rather allow $\xi$ to be in D.

Our objective is to find $B_i$ and $\xi \in D$ satisfying (2.6) which minimize $\sup_{g \in G} L(\underline{B}, \xi, g)$. To do this we start with some lemmas.

LEMMA 2.1. If the $B_i$ and $\xi$ satisfy (2.6) and $\sum_{i=1}^{k} B_i^2$ is not constant on supp $\xi$, where $\xi \in D_p$, then there exists $B_i^*$, $i = 1, \ldots, k$ and $\xi^* \in D_p$ satisfying (2.6), with $\sum_{i=1}^{k} B_i^{*2}$ constant on supp $\xi^*$ and $\inf_{g \in G} [L(\underline{B}, \xi, g) - L(\underline{B}^*, \xi^*, g)] > 0$.

Proof. Let $d\xi^*(\underline{x}) = \alpha(\sum_{i=1}^{k} B_i^2(\underline{x}))^{1/2} d\xi(\underline{x})$, where $\alpha$ is the constant making $\xi^*$ a probability measure. Let $B_i^*(\underline{x}) = B_i(\underline{x})/\alpha(\sum_{j=1}^{k} B_j^2(\underline{x}))^{1/2}$ for $i = 1, \ldots, k$ and define $B_i^*(\underline{x})$ to be 0 if the denominator is 0. Notice $B_i^*\, d\xi^* = B_i\, d\xi$ for all i so that $B_i^*$ and $\xi^*$ satisfy (2.6). Also notice that $\sum_{i=1}^{k} B_i^{*2}(\underline{x}) = \frac{1}{\alpha^2}$ is constant on supp $\xi^*$. Since

$$\begin{aligned}\int (\sum_{i=1}^{k} B_i^{*2}(\underline{x}))\, d\xi^*(\underline{x}) &= (1/\alpha)^2 \int d\xi^*(\underline{x}) \\ &= [(1/\alpha) \int d\xi^*(\underline{x})]^2 \\ &= [\int (\sum_{i=1}^{k} B_i^2(\underline{x}))^{1/2} d\xi(\underline{x})]^2 \\ &\le \int (\sum_{i=1}^{k} B_i^2(\underline{x}))\, d\xi(\underline{x})\end{aligned}$$

with strict inequality unless $\sum_{i=1}^{k} B_i^2(\underline{x})$ is constant on supp $\xi$, the lemma follows. Q.E.D.

LEMMA 2.2. Suppose $1 \leq r < s \leq k$ where r,s are integers. Let

$\pi_{rs}(\underline{x}) = (x_1, \ldots, x_{r-1}, x_s, x_{r+1}, \ldots, x_{s-1}, x_r, x_{s+1}, \ldots, x_k)'$

if $r < s-1$ and let

$\pi_{rs}(\underline{x}) = (x_1, \ldots, x_{r-1}, x_s, s_r, x_{s+1}, \ldots, x_k)'$

if $r = s-1$. In other words, $\pi_{rs}$ interchanges the r-th and s-th coordinates of a point in $R^k$.

Define $\pi_{rs} \circ g = g \circ \pi_{rs}$. Let $L = \sup_{g\varepsilon G} L(\underline{B}, \xi, g)$ for specific functions $\underline{B} = (B_1, \ldots, B_k)'$ and design $\xi \varepsilon D_p$ all satisfying (2.6). Then there exists a design $\overline{\xi} \varepsilon D_{2p}$ and functions $\overline{B}_1, \ldots, \overline{B}_k$ with $\overline{\underline{B}} = (\overline{B}_1, \ldots, \overline{B}_k)'$ such that

$$\overline{B}_i(\underline{x}) = \overline{B}_i(\pi_{rs}(\underline{x})) \qquad \text{if } i \neq r \text{ and } i \neq s$$

$$\overline{B}_r(\underline{x}) = \overline{B}_s(\pi_{rs}(\underline{x}) \qquad \text{if } i = r$$

$$\overline{B}_s(\underline{x}) = \overline{B}_r(\pi_{rs}(\underline{x}) \qquad \text{if } i = s$$

$$\overline{\xi}(\underline{x}) = \overline{\xi}(\pi_{rs}(\underline{x}))$$

all satisfying (2.6) and

(2.9) $$\sup_{g\varepsilon G} L(\overline{\underline{B}}, \overline{\xi}, g) \leq \sup_{g\varepsilon G} L(\underline{B}, \xi, g) = L$$

Proof. By lemma 2.1 if $\sum_{i=1}^{k} B_i^2$ is not constant on supp $\xi$ we may replace the $B_i$ and $\xi$ by functions $B_i^*$ and $\xi^* \varepsilon D_p$ satisfying (2.6), having $L(\underline{B}^*, \xi^*, g) < L(\underline{B}, \xi, g)$ for all $g \varepsilon G$, and such that $\sum_{i=1}^{k} B_i^{*2}$ is constant on supp $\xi^*$. So we shall assume $\sum_{i=1}^{k} B_i^2$ is constant on supp $\xi$. Then

$$\int \left(\sum_{i=1}^{k} B_i^2(\underline{x})\right) d\xi(\underline{x}) = \left[\int \left(\sum_{i=1}^{k} B_i^2(\underline{x})\right)^{\frac{1}{2}} d\xi(\underline{x})\right]^2$$

Let $\xi^{o}(\underline{x}) = \xi(\pi_{rs}(\underline{x}))$

$B_i^{o}(\underline{x}) = B_i(\pi_{rs}(\underline{x}))$ if $i \neq r$ and $i \neq s$

$B_r^{o}(\underline{x}) = B_s(\pi_{rs}(\underline{x}))$ if $i = r$

$B_s^{o}(\underline{x}) = B_r(\pi_{rs}(\underline{x}))$ if $i = s$

Now for each $g \in G$ we have $L(\underline{B}, \xi, g) = L(\underline{B}^{o}, \xi^{o}, \pi_{rs} \circ g)$ and hence

(2.10)
$$\begin{aligned} L &= \sup_{g\in G} L(\underline{B}, \xi, g) \\ &= \sup_{g\in G} L(\underline{B}^{o}, \xi^{o}, \pi_{rs} \circ g) \\ &= \sup_{g\in G} L(\underline{B}^{o}, \xi^{o}, g) \end{aligned}$$

since $g \in G$ if and only if $\pi_{rs} \circ g \in G$.

Let $u_i(\underline{x}) = B_i(\underline{x})\ \xi(\underline{x})$, $u_i^{o}(\underline{x}) = B_i^{o}(\underline{x})\ \xi^{o}(\underline{x})$ for $i = 1, \ldots, k$. The $u_i$ and $u_i^{o}$ are defined on

$$T = \{\underline{x} \in S_k;\ \underline{x} \in \operatorname{supp} \xi \text{ or } \pi_{rs}(\underline{x}) \in \operatorname{supp} \xi\}.$$

Let $\underline{u} = (u_1, \ldots, u_k)'$, $\underline{u}^{o} = (u_1^{o}, \ldots, u_k^{o})'$. Using (2.8) we get

$$L(\underline{B}, \xi, g) = L^{o}(\underline{u}, g)$$
$$L(\underline{B}^{o}, \xi^{o}, g) = L^{o}(\underline{u}^{o}, g)$$

where

$$L^{o}(\underline{v}, g) = \sum_{i=1}^{k} \Big( \sum_{x\in T} g(\underline{x})\ v_i(\underline{x})\Big)^2 + \rho \Big( \sum_{x\in T} \Big( \sum_{i=1}^{k} v_i^2(\underline{x})\Big)^{1/2}\Big)^2.$$

$L^{o}(\underline{v}, g)$ is convex in $\underline{v}$ and clearly $\sup_{g\in G} L^{o}(\underline{v}, g)$ is convex in $\underline{v}$ also.

Furthermore (2.6) remains valid for convex combinations of $\underline{B}$ satisfying (2.6).

Let $\bar{u}_i = (u_i + u_i^o)/2$ for $i = 1, \ldots, k$. Notice

$$\bar{u}_i(\underline{x}) = \bar{u}_i\,(\pi_{rs}\,(\underline{x})) \text{ if } i \neq r \text{ and } i \neq s$$
$$\bar{u}_i(\underline{x}) = \bar{u}_s\,(\pi_{rs}\,(\underline{x})) \text{ if } i = r$$
$$\bar{u}_i(\underline{x}) = \bar{u}_r\,(\pi_{rs}\,(\underline{x})) \text{ if } i = s$$

We have $\sup_{g \varepsilon G} L^o(\bar{\underline{u}}, g) \leq L$ by convexity and (2.10). Define $\bar{\xi}(\underline{x}) = \bar{\alpha}(\sum_{i=1}^{k} \bar{u}_i^2\,(\underline{x}))^{1/2}$ where $\bar{\alpha}$ makes $\bar{\xi}$ a probability measure. Let $\bar{B}_i(\underline{x}) = \bar{u}_i(\underline{x})/\bar{\xi}(\underline{x})$ if $\bar{\xi}(\underline{x}) > 0$, and 0 otherwise, for $i = 1, \ldots, k$. These are the $\bar{B}_1, \ldots, \bar{B}_k$ and $\bar{\xi}$ stated in the lemma.

Q.E.D.

LEMMA 2.3. Suppose we are given functions $B_1, \ldots, B_k$, with $\underline{B} = (B_1, \ldots, B_k)'$, and design $\xi \varepsilon D$ all satisfying (1.6). For any $g \varepsilon G$ there exists $g^* \varepsilon G$ with $|g^*\,(\underline{x})| = c$ on supp $\xi$ and such that

$$L(\underline{B}, \xi, g^*) \geq L(\underline{B}, \xi, g).$$

Proof. $L(\underline{B}, \xi, g)$ is a convex quadratic function of $g(\underline{x})$ for fixed $\underline{x} \varepsilon$ supp $\xi$. Thus it can be maximized by assigning $g(\underline{x})$ its extreme values, namely $\pm$ c. Let $g^*$ be a function derived from g by redefining g at each point $\underline{x}$ in supp $\xi$ so as to maximize $L(\underline{B}, \xi, g)$ as a function of $g(\underline{x})$ and so that $|g^*| = c$ on supp $\xi$. Clearly the values of $g^*$ off supp $\xi$ can then be chosen so that $g^* \varepsilon G$. Q.E.D.

Applying lemma 2.2 for all $1 \leq r < s \leq k$ we find that we can restrict attention to functions $B_1, \ldots, B_k$ and designs $\xi \varepsilon D$ such that

(2.11)
$$\xi\,(\underline{x}) = \xi\,(\pi_{rs}\,(\underline{x})) \quad \text{for all } 1 \leq r < s \leq k$$
$$B_i(\underline{x}) = B_i(\pi_{rs}\,(\underline{x})) \quad \text{if } i \neq r \text{ and } i \neq s$$
$$B_r(\underline{x}) = B_s(\pi_{rs}\,(\underline{x}))$$
$$B_s(\underline{x}) = B_r(\pi_{rs}\,(\underline{x}))$$

for purposes of finding functions $B_1, \ldots, B_k$ and design $\xi \in D$ satisfying (2.6) to minimize $\sup_{g \in G} L(\underline{B}, \xi, g)$.

Let M be the set of ordered pairs $(\underline{B}, \xi)$ where $\underline{B} = (B_1, \ldots, B_k)'$, the $B_i$ are functions on $S_k$, $\xi \in D$, and $\underline{B}$ and $\xi$ satisfy (2.6) and (2.11). Notice the $B_i$ and $\xi$ also satisfy (2.7).

From lemma 2.2 it follows that the inf of $\sup_{g \in G} L(\underline{B}, \xi, g)$ over all $\underline{B}$ and $\xi \in D$ for which (2.6) holds is the same as $\inf_{(\underline{B},\xi) \in M} \sup_{g \in G} L(\underline{B}, \xi, g)$.

Next let $G(c) = \{g: S_k \to R;\ |g(\underline{x})| = c\}$. Since card (supp $\xi$) $< \infty$ for fixed $\xi \in D$, it follows that for all $h \in G(c)$ there is a $g \in G$ with $g = h$ on supp $\xi$. Hence by lemma 2.3, given $(\underline{B}, \xi) \in M$, we have $\sup_{g \in G} L(\underline{B}, \xi, g) = \sup_{g \in G(c)} L(\underline{B}, \xi, g)$.

We now proceed to find $(\underline{B}^*, \xi^*) \in M$ such that $\sup_{g \in G(c)} L(\underline{B}^*, \xi^*, g) = \inf_{(B,\xi) \in M} \sup_{g \in G(c)} L(\underline{B}, \underline{\xi}, g)$. Notice that for such a $(\underline{B}^*, \xi^*)$ we also have that $\underline{B}^*$ and $\xi^*$ minimize $\sup_{g \in G} L(\underline{B}, \xi, g)$ overall $\underline{B}$ and $\xi \in D$ satisfying (2.6).

For any $g \in G(c)$ and $(\underline{B}, \xi) \in M$,

(2.12)
$$L(\underline{B}, \xi, g) = \sum_{i=1}^{k} \left(\int B_i(\underline{x})\, g(\underline{x})\, d\xi\, (\underline{x})\right)^2 + \rho \sum_{i=1}^{k} \int B_i^2\, (\underline{x})\, d\xi\, (\underline{x})$$

$$= \sum_{i=1}^{k} \left(\int B_i(\pi_{1i}\, (\underline{x}))\, \pi_{1i} \circ g(\underline{x})\, d\xi\, (\pi_{1i}\, (\underline{x}))\right)^2$$

$$+ \rho \sum_{i=1}^{k} \int B_i^2(\pi_{1i}\, (\underline{x}))\, d\xi\, (\pi_{1i}\, (\underline{x}))$$

$$= \sum_{i=1}^{k} \left(\int B_1(\underline{x})\, \pi_{1i} \circ g(\underline{x})\, d\xi\, (\underline{x})\right)^2$$

$$+ \rho \sum_{i=1}^{k} \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})$$

$$\leq \sum_{i=1}^{k} (\int |B_1(\underline{x})| \; |\pi_{1i} \circ g(\underline{x})| \; d\xi \; (\underline{x}))^2$$

$$+ \rho k \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})$$

$$= kc^2 \; (\int |B_1(\underline{x})| \; d\xi \; (\underline{c}))^2 + \rho k \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})$$

This holds for any $g \in G(c)$, so

(2.13)
$$\inf_{(\underline{B},\xi)\in M} \; \sup_{g\in G(c)} \; L(\underline{B}, \xi, g)$$

$$\leq \inf_{(\underline{B},\xi)\in M} \; [kc^2(\int |B_1(\underline{x})| \; d\xi \; (\underline{x}))^2 + k\rho \int B_1^2 \, (\underline{x}) \; d\xi \; c(\underline{x})]$$

$$\leq \inf_{(\underline{B},\xi)\in M^+} \; [kc^2(\int |B_1(\underline{x})| \; d\xi \; (\underline{x}))^2 + k\rho \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})]$$

$$= \inf_{(\underline{B},\xi)\in M^+} \; [kc^2(\int B_1(\underline{x}) \; d\xi \; (\underline{x}))^2 + k\rho \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})]$$

$$= \inf_{(\underline{B},\xi)\in M^+} \; [kc^2 + k\rho \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})]$$

where $M^+$ is the set of all $(\underline{B},\xi)\in M$ such that $B_1 \geq 0$ on $S_k$. Notice that the last equality in (2.13) follows from the fact that $\int B_1(\underline{x}) \; d\xi \; (\underline{x}) = 1$ by (2.7).

Let $m_{21}(\xi) = \int x_1^2 \; d\xi \; (\underline{x})$. Notice that if $(\underline{B},\xi) \in M$ then $m_{21}(\xi) > 0$. For if $m_{21}(\xi) = 0$ then supp $\xi \subset \{\underline{x} \in S_k;\; x_1 = 0\}$ and one would have $\int x_1 \; B_1(\underline{x}) \; d\xi \; (\underline{x}) = 0$ contradicting (2.7). Since $m_{21}(\xi) > 0$ when $(\underline{B},\xi) \in M$ one has

(2.14)
$$\begin{aligned} \int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x}) &= (\int B_1^2 \, (\underline{x}) \; d\xi \; (\underline{x})) \; (\int (x_1^2/m_{21} \; (\xi)) \; d\xi \; (\underline{x})) \\ &\geq [\int (B_1 \; (\underline{x}) \; x_1/\sqrt{m_{21}(\xi)}) \; d\xi \; (\underline{x})]^2 \\ &= (\int x_1 \; B_1 \; (\underline{x}) \; d\xi \; (\underline{x}))^2/m_{21}(\xi) \\ &= 1/m_{21}(\xi) \end{aligned}$$

using the Cauchy-Schartz inequality and (2.6). Notice that there is equality throughout (2.14) if and only if $B_1(\underline{x}) = ax_1$ for some $a \in R$ on supp $\xi$. In fact for (2.6) to hold, one must have $a = 1/m_{21}(\xi)$. Thus given $\xi \in D$, $\int B_1^2(\underline{x})\, d\xi(\underline{x})$ is minimized by $B_1(\underline{x}) = x_1/m_{21}(\xi)$.

Next notice that among all $\xi \in D$ satisfying (2.11) one can show that $m_{21}(\xi)$ is maximized by the design $\xi^*$ which puts equal mass $1/k$ on the k points $(1, 0, \ldots, 0)'$, $(0, 1, 0, \ldots, 0)'$, $\ldots$, $(0, \ldots, 0, 1)'$ on $S_k$. In this case $m_{21}(\xi^*) = 1/k$.

Define $\underline{B}^* = (B_1^*, \ldots, B_k^*)'$ where $B_i^*(\underline{x}) = kx_i$. Notice that $(\underline{B}^*, \xi^*) \in M^+$. By the remarks in the preceeding two paragraphs one can see that $(\underline{B}^*, \xi^*)$ minimizes $\int B_1^2(\underline{x})\, d\xi(\underline{x})$ over all $(\underline{B}, \xi) \in M$. Thus using (2.13) we have

(2.15)
$$\begin{aligned}
\inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G(c)} L(\underline{B}, \xi, g)
&\le \inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G(c)} [kc^2 + k\rho \int B_1^2(\underline{x})\, d\xi\, )\underline{x})] \\
&= kc^2 + k\rho \int B_1^{*2}(\underline{x})\, d\xi^*(\underline{x}) \\
&= kc^2 + k^2\rho .
\end{aligned}$$

Now for any $(\underline{B},\xi)\in M$

(2.16)
$$\begin{aligned}
\sup_{g\in G(c)} L(\underline{B}, \xi, g) &\ge L(\underline{B}, \xi, c) \\
&= \sum_{i=1}^{k} \left(\int B_i(\underline{x})\, c\, d\xi(\underline{x})\right)^2 + \rho \sum_{i=1}^{k} \int B_i^2(\underline{x})\, d\xi(\underline{x}) \\
&= c^2 \sum_{i=1}^{k} \left(\int B_i(\underline{x})\, d\xi(\underline{x})\right)^2 + \rho \sum_{i=1}^{k} \int B_i^2(\pi_{1i}(\underline{x}))\, d\xi(\pi_{1i}(\underline{x})) \\
&= c^2 \sum_{i=1}^{k} 1 + \rho \sum_{i=1}^{k} \int B_1^2(\underline{x})\, d\xi(\underline{x}) \\
&= kc^2 + k\rho \int B_1^2(\underline{x})\, d\xi(\underline{x}).
\end{aligned}$$

Thus

(2.17) $$\inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G(c)} L(\underline{B}, \xi, g) \geq \inf_{(\underline{B},\xi)\in M} (kc^2 + k\rho \int B_1^2 (\underline{x})\, d\xi\, (\underline{x})).$$

By arguing as between equations (2.14) and (2.15) we can show that $(\underline{B}^*, \xi^*)$ minimizes $\int B_1^2 (\underline{x})\, d\xi\, (\underline{x})$ over M where $(\underline{B}^*, \xi^*)$ M are as before. Hence

(2.18) $$\inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G(c)} L(\underline{B}, \xi, g) \geq \inf_{(\underline{B},\xi)\in M} (kc^2 + k\rho \int B_1^2 (\underline{x})\, d\xi\, (\underline{x}))$$
$$= kc^2 + k\rho \int B_1^{*2} (\underline{x})\, d\xi^* (\underline{x})$$
$$= kc^2 + k^2\rho.$$

Combining (2.15) and (2.18) we get

(2.19) $$\inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G(c)} L(\underline{B}, \xi, g) = \sup_{g\in G(c)} L(\underline{B}^*, \xi^*, g)$$
$$= kc^2 + k^2\rho.$$

From all the above arguments it follows that $\underline{B}^*$ and $\xi^* \in D$ give a solution to our original minimax problem stated just before lemma 2.1. Summarizing we have the following theorem.

THEOREM 2.1. Let $\xi^* \in D$ be the probability measure putting mass 1/k on each of the k points $(1, 0, \ldots, 0)'$, $(0, 1, 0, \ldots, 0)'$, ..., $(0, \ldots, 0, 1)'$ in $S_k$. Suppose $\underline{B}^* = (B_1^*, \ldots, B_k^*)'$ where $B_i^* (\underline{x}) = kx_i$ for $i = 1, \ldots, k$. Let M be the set of all ordered pairs $(\underline{B}, \xi)$ where $\underline{B} = (B_1, \ldots, B_k)'$, the $B_i$ are functions on $S_k$, $\xi \in D$, and $\underline{B}$ and $\xi$ satisfy (2.6). Then

$$\sup_{g\in G} L(\underline{B}^*, \xi^*, g) = \inf_{(\underline{B},\xi)\in M}\ \sup_{g\in G} L(\underline{B}, \xi, g)$$

Notice that $\underline{B}^*$ gives rise to the usual best linear unbiased estimators with respect to $\xi^*$ for linear regression on $S_k$. Also notice $\xi^*$ is an optimal design for linear regression on $S_k$ with respect to a broad class of optimality criteria

## 3. Results for cubes.

In this section we shall examine a multivariate regression and design problem, analogous to that discussed in the preceeding section for simplexes, on the k-dimensional cube centered at the origin with sides of length 2. We shall again find that the usual designs and best linear unbiased estimators are optimal. The method used to prove this is similar to that used for simplexes but is somewhat more complicated due to the fact that the coordinates can take on negative values. This shall become clear as we proceed.

Consider the multivariate regression problem

(3.1) $$Y(\underline{x}(m)) = \beta_o + \underline{\beta}' \underline{x}_m + g(\underline{x}_m) + e_m$$

where $m = 1, 2, \ldots, n$, the $e_m$ are uncorrelated random variables with mean 0 and finite variance $\sigma^2 > 0$, $\beta_o \in R$, $\underline{\beta}' = (\beta_1, \ldots, \beta_k) \in R^k$, $I^k$ = k-fold Cartesian product of the closed interval $I = [-1, +1]$, $\underline{x}_m = (x_{1m}, \ldots, x_{km})'$ $\in I^k$, and for some fixed constant $c > 0$,

$$g \in G = \{g: \; I^k \to R; \; |g(\underline{x})| \leq c, \int g(\underline{x})\, d\underline{x} = \int x_i \, g(\underline{x})\, d\underline{x} = 0 \text{ for } i = 1, \ldots, k. \text{ Here } x = (x_1, \ldots, x_k)'\}.$$

In the definition of G $d\underline{x}$ is Lebesgue measure on $I^k$ and all integrals will be assumed to be over $I^k$ until otherwise noted.

The conditions $\int g(\underline{x})\, d\underline{x} = \int x_i \, g(\underline{x})\, d\underline{x} = 0$ are equivalent to requiring $\int g^2(\underline{x})\, d\underline{x} = \inf_{\underline{b} \in R^k} \int (g(\underline{x}) - b'\underline{x})^2 \, d\underline{x}$ which says that the best linear approximation to g in the $L_2$ norm with respect to Lebesgue measure on $I^k$ is the function 0. This condition insures the uniqueness of the $\beta_i$ in our model (3.1).

Analogous to what was done in the case of the simplex, we shall denote by $\Xi_p$ the class of all exact designs for p observations on $I^k$. We also define

$D_p$ = {probability measures $\xi$ on $I^k$; card (supp $\xi$ ) $\leq$ p}

$D = \bigcup_{p=1}^{\infty} D_p$

For $\xi \in \Xi_n$, let the linear estimators of the $\beta_i$ be defined by

(3.2) $$\hat{\beta}_i = \int Y(\underline{x})\ b_i(\underline{x})\ d\xi\ (\underline{x}) \quad , i = 0, 1, \ldots, k$$

where the $b_i$ are real valued functions on $I^k$.

We shall consider the (expected) mean square error due to the design $\xi$ and the estimators $\hat{\beta}_i$, namely

(3.3) $$\sum_{i=0}^{k} \theta_i^2\ E(\hat{\beta}_i - \beta_i)^2$$

where $\theta_o = 1$, $0 \leq \theta_i \leq 1$ for i = 1, ..., k. The $\theta_i$ are known constants. This mean square error can be rewritten as the sum of a variance term, $\sum_{i=0}^{k} \theta_i^2\ E(\hat{\beta}_i - E\hat{\beta}_i)^2$ and a bias term $\sum_{i=0}^{k} \theta_i^2\ (\beta_i - E\hat{\beta}_i)^2$. We can write the variance term as

(3.4) $$\sum_{i=0}^{k} \theta_i^2\ E(\hat{\beta}_i - E\hat{\beta}_i)^2 = (\sigma^2/n) \sum_{i=0}^{k} \theta_i^2 \int b_i^2\ (\underline{x})\ d\xi\ (\underline{x})$$

and the bias term is determined by the equations

(3.5) $$E\hat{\beta}_j - \beta_j = \sum_{i=0, i\neq j}^{k} \beta_i \int x_i\ b_j(\underline{x})\ d\xi\ (\underline{x})$$
$$+ \beta_j\ [\int x_j\ b_j\ (\underline{x})\ d\xi\ (\underline{x}) - 1]$$
$$+ \int b_j(\underline{x})\ g(\underline{x})\ d\xi(\underline{x})$$

for j = 0, 1, ..., k. We take $x_o = 1$ and $\underline{x} = (x_1, \ldots, x_k)'$ in (3.5).

If the $\beta_i$ are unbounded in order for the error to be bounded we must have $\int x_i \, b_j(\underline{x}) \, d\xi(\underline{x}) = \delta_{ij}$ for all $1 \leq i, j \leq k$, where $\delta_{ij}$ is the Kronecker delta. This is equivalent to saying the linear estimators are unbiased if $g = 0$. If we define $B_i(\underline{x}) = \theta_i \, b_i(\underline{x})$ then this condition becomes

(3.6) $$\int x_i \, B_j(\underline{x}) \, d\xi(\underline{x}) = \theta_j \, \delta_{ij} \quad , \; 0 \leq i, j \leq k.$$

If any $\theta_i = 0$ then $B_i(\underline{x}) = 0$ and we are really estimating fewer parameters and are in a lower dimensional case.

Let $\rho = \sigma^2/n$, $\underline{B} = (B_o, B_1, \ldots, B_k)'$, and define

(3.7) $$L(\underline{B}, \xi, g) = \sum_{i=0}^{k} \left(\int B_i(\underline{x}) \, g(\underline{x}) \, d\xi(\underline{x})\right)^2 + \rho \int \left(\sum_{i=0}^{k} B_i^2(\underline{x})\right) d\xi(\underline{x}).$$

Notice that $L(\underline{B}, \xi, g)$ is equal to (3.3) with condition (3.6) imposed. Condition (3.6) and $L(\underline{B}, \xi, g)$ are all well defined for $\xi \in D$ and from now on we shall not restrict $\xi$ to be an exact design for a particular p but rather allow $\xi$ to be in D.

Our objective is to find the $B_i$ and $\xi \in D$ satisfying (3.6) which minimize $\sup_{g \in G} L(\underline{B}, \xi, g)$. We begin with some lemmas similar to those used in the case of the simplex.

<u>LEMMA 3.1.</u> If the $B_i$ and $\xi$ satisfy (3.6) and $\sum_{i=0}^{k} B_i^2$ is not constant on supp $\xi$, where $\xi \in D_p$, then there exists $B_i^*$, $i = 0, 1, \ldots, k$, and $\xi^* \in D_p$ satisfying (3.6) with $\sum_{i=0}^{k} B_i^{*2}$ constant on supp $\xi^*$ and with

$$\inf_{g \in G} [L(\underline{B}, \xi, g) - L(B^*, \xi^*, g)] > 0.$$

<u>Proof</u>. The proof is similar to Lemma 2.1 and is therefore omitted. Q.E.D.

LEMMA 3.2. Suppose $1 \leq q \leq k$, q an integer, and let

(3.8) $$T_q(\underline{x}) = (x_1, \ldots, x_{q-1}, -x_q, x_{q+1}, \ldots, x_k)'.$$

Define $T_q \circ g = g \circ T_q$. Let $L = \sup_{g \varepsilon G} L(\underline{B}, \xi, g)$ for specific functions $\underline{B} = (B_o, B_1, \ldots, B_k)'$ and design $\xi \varepsilon D_p$ all satisfying (3.6). Then there exists a design $\overline{\xi} \varepsilon D_{2p}$ and $\underline{\overline{B}} = (\overline{B}_o, \overline{B}_1, \ldots, \overline{B}_k)'$ satisfying (3.6) with the property

(3.9) $$\begin{aligned} \overline{B}_i(\underline{x}) &= \overline{B}_i\ (T_q(\underline{x})) \quad , i \neq q \\ \overline{B}_q(\underline{x}) &= -\overline{B}_q(T_q(\underline{x})) \\ \overline{\xi}(\underline{x}) &= \overline{\xi}\ \ (T_q(\underline{x})) \end{aligned}$$

and such that

(3.10) $$\sup_{g \varepsilon G} L(\underline{\overline{B}}, \overline{\xi}, g) \leq \sup_{g \varepsilon G} L(\underline{B}, \xi, g) = L$$

Proof. The proof is similar to Lemma 2.2, with obvious modifications, and is therefore omitted. Q.E.D.

LEMMA 3.3. Suppose we are given $\underline{B} = (B_o, B_1, \ldots, B_k)'$ and design $\xi \varepsilon D$ satisfying (3.6). For any $g \varepsilon G$ there exists $g^* \varepsilon G$ with $|g^*(\underline{x})| = c$ on supp $\xi$ and such that $L(\underline{B}, \xi, g^*) \geq L(\underline{B}, \xi, g)$.

Proof. The proof is similar to Lemma 2.3 and is therefore omitted. Q.E.D.

Applying Lemma 3.2 for $q = 1, 2, \ldots, k$ we find that we can restrict attention to functions $B_o, B_1, \ldots, B_k$ and designs $\xi \varepsilon D$ such that

(3.11) $$\begin{aligned} \xi(\underline{x}) &= \xi(T_q(\underline{x})) && q = 1, \ldots, k \\ B_i(\underline{x}) &= B_i(T_q(\underline{x})) && i \neq q,\ i = 0, 1, \ldots, k,\ q = 1, \ldots, k \\ B_q(\underline{x}) &= -B_q(T_q(\underline{x})) && q = 1, \ldots, k \end{aligned}$$

for purposes of finding functions $B_o, B_1, \ldots, B_k$ and design $\xi \varepsilon D$ all satisfying (3.6) to minimize $\sup_{g \varepsilon G} L(\underline{B}, \xi, g)$. In fact, if we let M be the set of ordered pairs $(\underline{B},\xi)$ where $\underline{B} = (B_o, B_1, \ldots, B_k)'$, the $B_i$ are functions on $I^k$, $\xi \varepsilon D$, and $\underline{B}$ and $\xi$ satisfy (3.6) and (3.11), it follows from Lemma 3.2 that the inf of $\sup_{g \varepsilon G} L(\underline{B}, \xi, g)$ over all $\underline{B}$ and $\xi \varepsilon D$ for which (3.6) holds is the same as $\inf_{(\underline{B},\xi)\varepsilon M} \sup_{g\varepsilon G} L(\underline{B}, \xi, g)$.

Suppose $(\underline{B}, \xi) \varepsilon M$. Define $H^k$ to be the k-fold Cartesian product of the closed interval $H = [0,1]$. Let $n(\underline{x})$ be the number of non-zero coordinates of $\underline{x} = (x_1, \ldots, x_k)'$ and let $P(\underline{x}) = \{\underline{y} = (y_1, \ldots, y_k)' \varepsilon I^k; (|y_1|, \ldots, |y_k|)\} = (|x_1|, \ldots, |x_k|)\}$ for $\underline{x} = (x_1, \ldots, x_k)' \varepsilon I^k$. Notice card $P(\underline{x}) = 2^{n(\underline{x})}$ and $P(\underline{x})$ is the set of all points $(\underline{+}x_1, \ldots, \underline{+}x_k)'$.

Define

(3.12)
$$g_o(\underline{x}) = \sum_{y \varepsilon P(\underline{x})} g(\underline{y})/2^{n(\underline{x})}$$

$$g_i(\underline{x}) = \sum_{y \varepsilon P(\underline{x})} \operatorname{sgn}(y_i\ x_i)\ g(\underline{y})/2^{n(\underline{x})} \quad , i = 1, \ldots, k$$

where $\operatorname{sgn}(\underline{x})$ is +1 is $x > 0$, 0 if $x = 0$, and -1 if $x < 0$. We can then write

(3.13)
$$L(\underline{B}, \xi, g) = \sum_{i=0}^{k} [\int B_i(\underline{x})\ g_i(\underline{x})\ d\overline{\xi}\ (\underline{x})]^2 + \rho \int \sum_{i=0}^{k} B_i^2(\underline{x})\ d\overline{\xi}(\underline{x})$$

where for any $\xi \varepsilon D$ we define $\overline{\xi}$ to be the probability measure on $H^k$ with $\overline{\xi}(\underline{x}) = 2^{n(\underline{x})}\xi(\underline{x})$. The integrals in (3.13) are over $H^k$.

Let $G(c) = \{g\colon\ I^k \to R;\ |g(\underline{x})| = c\}$. Since card (supp $\xi$) $< \infty$ for fixed $\xi \varepsilon D$ it follows that for all $\overline{g} \varepsilon G(c)$ there is a $g \varepsilon G$ with $\overline{g} = g$ on supp $\xi$. Applying Lemma 3.3 we have for $(\underline{B},\xi)\varepsilon M$ that $\sup_{g\varepsilon G(c)} L(\underline{B}, \xi, g) = \sup_{g\varepsilon G} L(\underline{B}, \xi, g)$.

Notice that if $B_o$, $B_1$, ..., $B_k$ satisfy (3.11) then

(3.14) $$B_i(x_i, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) = 0 \quad , i = 1, \ldots, k.$$

Let us now examine (3.13) on more detail for the cases k = 1 and k = 2. This will help motivate Lemma 3.4. We assume $(\underline{B}, \xi) \in M$.

Suppose k = 1. Let $\overline{g}$ be in G(c). Fix $x \in H$. If $\overline{g}(x) = \overline{g}(-x) = c$ (or -c) then

(3.15) (i) $$\overline{g}_o(x) = [\overline{g}(x) + \overline{g}(-x)]/2 = c \quad \text{(or } -c)$$
$$\overline{g}_1(x) = [\overline{g}(x) - \overline{g}(-x)]/2 = 0$$

If $\overline{g}(x) = -\overline{g}(-x) = c$ (or -c) then

(3.15) (ii) $$\overline{g}_o(x) = 0$$
$$\overline{g}_1(x) = c \quad \text{(or } -c)$$

Notice case (ii) can only occur if $x \neq 0$.

Equation (3.15) covers all possible values of $\overline{g}_o$ and $\overline{g}_1$. We can then break H up into two disjoint sets. We get the partition $\{H_1(\overline{g}), H_2(\overline{g})\}$ of H where

$$H_1(\overline{g}) = \{x \in H;\ |\overline{g}_o(x)| = c,\ |\overline{g}_1(x)| = 0\}$$
$$H_2(\overline{g}) = \{x \in H;\ |\overline{g}_o(x)| = 0,\ |\overline{g}_1(x)| = c\}$$

Let $\Gamma(\overline{g}) = \{g \in G(c);\ \{H_1(g), H_2(g)\} = \{H_1(\overline{g}), H_2(\overline{g})\}\}$.
If we let $\{U_i\ V_i\}$ be any partition of $H_i(\overline{g})$, i = 1,2, then it is not hard to see that there is $g \in \Gamma(\overline{g})$ such that $g_o = c$ on $U_1$, $g_o = -c$ on $V_1$, $g_1 = c$ on $U_2$, and $g_1 = -c$ on $V_2$. In other words, any possible assignment of

signs to the $g_i$ in $H_1(\overline{g})$ and $H_2(\overline{g})$ is attainable for some $g \in \Gamma(\overline{g})$.

It is also not difficult to see that given any partition $\{H_1, H_2\}$ of $H$ with $0 \in H_1$ and for any partitions $\{U_i, V_i\}$ of the $H_i$ there is a $g \in G(c)$ such that $H_1(g) = H_1$, $H_2(g) = H_2$, $g_o = c$ on $U_1$, $g_o = -c$ on $V_1$, $g_1 = c$ on $U_2$, and $g_1 = -c$ on $V_2$. Equation (3.15) can be used to construct this g.

This allows us to rewrite (3.13) as

$$L(\underline{B}, \xi, g) = \sum_{i=0}^{1} \Big( \int_{i+1} B_i(x)\, g_i(x)\, d\overline{\xi}(x)\Big)^2 + \rho \int \Big(\sum_{i=0}^{1} B_i^2(x)\Big) d\overline{\xi}(x)$$

where $\int_{i+1}$ is over $H_{i+1}(g)$ and the last integral is over H. By letting

$$U_1 = \{x \in H_1(g);\ B_o(x) \geq 0\}$$

$$V_1 = \{x \in H_1(g);\ B_o(x) < 0\}$$

$$U_2 = \{x \in H_2(g);\ B_1(x) \geq 0\}$$

$$V_2 = \{x \in H_2(g);\ B_1(x) < 0\}$$

we can choose $g^*$ as in the preceeding paragraphs so that

$$L(\underline{B}, \xi, g^*) = \sum_{i=0}^{1} \Big( \int_{i+1} |B_i(x)|\ |g_i(x)|\ d\overline{\xi}(x)\Big)^2$$

$$+ \rho \int \Big(\sum_{i=0}^{1} B_i^2(x)\Big)\ d\overline{\xi}(x)$$

$$= c^2 \sum_{i=0}^{1} \Big( \int_{i+1} |B_i(x)|\ d\overline{\xi}(x)\Big)^2$$

$$+ \rho \int \Big(\sum_{i=0}^{1} B_i^2\ (x)\Big)\ d\overline{\xi}(x)$$

and we then get

$$\sup_{g \in G(c)} L(\underline{B}, \xi, g) = \sup_{\{H_1, H_2\}} c^2 \sum_{i=0}^{1} \left( \int_{i+1} |B_i(x)| \, d\bar{\xi}(x) \right)^2$$

$$+ \rho \int \left( \sum_{i=0}^{1} B_i^2(x) \right) d\bar{\xi}(x)$$

where the sup on the right hand side is over all partitions $\{H_1, H_2\}$ of $H$ such that $0 \in H_1$. The relationship between $\xi$ and $\bar{\xi}$ is as defined below (3.13). Our goal is to rewrite $\sup_{g \in G(c)} L(\underline{B}, \xi, g)$ in a manner similar to the above for general k.

To see that this can also be done for k = 2 and to get an idea of how one proceeds in general, we write out the case k = 2 in detail. Let $\bar{g} \in G(c)$ and let $\underline{x} = (x_1, x_2)' \in H^2$. for $\underline{x}$ such that $x_1 \neq 0$ and $x_2 \neq 0$ we find

(3.16) (i) if $\bar{g}(x_1, x_2) = \bar{g}(x_1, -x_2) = \bar{g}(-x_1, x_2) = \bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_0(x_1, x_2) = c \text{ (or } -c), \; \bar{g}_1(x_1, x_2) = 0, \; \bar{g}_2(x_1, x_2) = 0$$

(ii) If $-\bar{g}(x_1, x_2) = \bar{g}(x_1, -x_2) = \bar{g}(-x_1, x_2) = \bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_0(x_1, x_2) = c/2 \text{ (or } -c/2), \; \bar{g}_1(x_1, x_2) = -c/2 \text{ (or } c/2),$$

$$\bar{g}_2(x_1, x_2) = -c/2 \text{ (or } c/2).$$

(iii) If $\bar{g}(x_1, x_2) = -\bar{g}(x_1, -x_2) = \bar{g}(-x_1, x_2) = \bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_0(x_1, x_2) = c/2 \text{ (or } -c/2), \; \bar{g}_1(x_1, x_2) = -c/2 \text{ (or } c/2),$$

$$\bar{g}_2(x_1, x_2) = c/2 \text{ (or } -c/2).$$

(iv) If $\bar{g}(x_1, x_2) = \bar{g}(x_1, -x_2) = -\bar{g}(-x_1, x_2) = \bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_o(x_1, x_2) = c/2 \text{ (or } -c/2\text{)}, \; \bar{g}_1(x_1, x_2) = c/2 \text{ (or } -c/2\text{)},$$

$$\bar{g}_2(x_1, x_2) = -c/2 \text{ (or } c/2\text{)}.$$

(v) If $\bar{g}(x_1, x_2) = \bar{g}(x_1, -x_2) = \bar{g}(-x_1, x_2) = -\bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_o(x_1, x_2) = c/2 \text{ (or } -c/2\text{)}, \; \bar{g}_1(x_1, x_2) = c/2 \text{ (or } -c/2\text{)},$$

$$\bar{g}_2(x_1, x_2) = c/2 \text{ (or } -c/2\text{)}.$$

(vi) If $\bar{g}(x_1, x_2) = \bar{g}(x_1, -x_2) = -\bar{g}(-x_1, x_2) = -\bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_o(x_1, x_2) = 0, \; \bar{g}_1(x_1, x_2) = c \text{ (or } -c\text{)}, \; \bar{g}_2(x_1, x_2) = 0$$

(vii) If $\bar{g}(x_1, x_2) = -\bar{g}(x_1, -x_2) = \bar{g}(-x_1, x_2) = -\bar{g}(-x_1, -x_2) = c$ (or $-c$) then

$$\bar{g}_o(x_1, x_2) = 0, \; \bar{g}_1(x_1, x_2) = 0, \; \bar{g}_2(x_1, x_2) = c \text{ (or } -c\text{)}.$$

For $\underline{x} = (x_1, x_2)'$ such that $x_1 = 0$ and $x_2 \neq 0$ only cases (i) and (vii) can occur. For $\underline{x}$ such that $x_1 \neq 0$ and $x_2 = 0$ only cases (i) and (vi) can occur. For $\underline{x}$ such that $x_1 = x_2 = 0$ only case (i) can occur.

Equation (3.16) (i) - (vii) covers all possible values of $\bar{g}_o(\underline{x})$, $\bar{g}_1(\underline{x})$, and $\bar{g}_2(\underline{x})$ for $x \in H^2$. We can break $H^2$ into four disjoint sets, for fixed $\bar{g}$, by varying $\underline{x} \in H^2$. We get the partition $\{H_1(\bar{g}), H_2(\bar{g}), H_3(\bar{g}), H_4(\bar{g})\}$ of $H^2$ where

$$H_1(\bar{g}) = \{\underline{x} \in H^2;\ |\bar{g}_o(\underline{x})| = c,\ |\bar{g}_1(\underline{x})| = 0,\ |\bar{g}_2(\underline{x})| = 0\}$$

$$H_2(\bar{g}) = \{\underline{x} \in H^2;\ |\bar{g}_o(\underline{x})| = 0,\ |\bar{g}_1(\underline{x})| = c,\ |\bar{g}_2(\underline{x})| = 0\}$$

$$H_3(\bar{g}) = \{\underline{x} \in H^2;\ |\bar{g}_o(\underline{x})| = 0,\ |\bar{g}_1(\underline{x})| = 0,\ |\bar{g}_2(\underline{x})| = c\}$$

$$H_4(\bar{g}) = \{\underline{x} \in H^2;\ |\bar{g}_o(\underline{x})| = c/2,\ |\bar{g}_1(\underline{x})| = c/2,\ |\bar{g}_2(\underline{x})| = c/2\}$$

Notice $(0,0)' \in H_1(\bar{g})$, $(x_1, 0)' \in H_1(\bar{g}) \cup H_2(\bar{g})$, $(0, x_2)' \in H_1(\bar{g}) \cup H_3(\bar{g})$.

If we let

$\Gamma(\bar{g}) = \{g \in G(c);\ \{H_1(g), H_2(g), H_3(g), H_4(g)\} = \{H_1(\bar{g}), H_2(\bar{g}), H_3(\bar{g}), H_4(\bar{g})\}\}$ then an inspection of equation (3.16) (i) - (vii) shows that for any assignment of signs to the $|\bar{g}_i(\underline{x})|$ for each $\underline{x}$ in each $H_j(\bar{g})$, $i = 0, 1, 2$, $j = 1, 2, 3, 4$, there is a $g \in \Gamma(\bar{g})$ such that the $g_i(\underline{x})$ have precisely these signs. In other words the partition $\{H_1(\bar{g}), \ldots, H_4(\bar{g})\}$ is determined only by $|\bar{g}_i(\underline{x})|$ and among all $g \in G(c)$ such that $|g_i(\underline{x})| = |\bar{g}_i(\underline{x})|$ for all $\underline{x} \in H^2$, $i = 0, 1, 2$, all possible values for the sign of $|\bar{g}_i(\underline{x})|$ are attainable.

The result of all this is that we find for $(\underline{B}, \xi)' \in M$

$$L(\underline{B}, \xi, g) = \sum_{i=0}^{2} \left(\int_4 B_i(\underline{x})\, g_i(\underline{x})\, d\bar{\xi}(\underline{x}) + \int_{i+1} B_i(\underline{x})\, g_i(\underline{x})\, d\bar{\xi}(\underline{x})\right)^2$$

$$+ \rho \int \left(\sum_{i=0}^{2} B^2_{i}(\underline{x})\right) d\bar{\xi}(\underline{x})$$

where $\int_{i+1}$ is over $H_{i+1}(g)$ and the last integral is over $H^2$

Since any possible assignment of signs to the $|g_i(\underline{x})|$ are attainable for some $g^* \in \Gamma(g)$ we can choose $g^* \in \Gamma(g)$ so that $\operatorname{sgn} g_i^*(\underline{x}) = \operatorname{sgn} B_i(\underline{x})$ and $|g_i^*(\underline{x})| = |g_i(\underline{x})|$ for all $\underline{x} \in H^2$ and $i = 0, 1, 2$. Then

$$L(\underline{B}, \xi, g^*) = \sum_{i=0}^{2} [(c/2) \int_4 |B_i(\underline{x})| \, d\overline{\xi}(\underline{x}) + c \int_{i+1} |B_i(\underline{x})| \, d\overline{\xi}(\underline{x})]^2$$

$$+ \rho \int (\sum_{i=0}^{2} B_i^2(\underline{x})) \, d\overline{\xi}(\underline{x}).$$

One also finds that for any partition $\{H_1, H_2, H_3, H_4\}$ of $H^2$ such that $(x_1, 0)' \varepsilon H_i \cup H_2$ and $(x_2, 0)' \varepsilon H_1 \cup H_3$ there exists $g \varepsilon G(c)$ such that $\{H_1(g), \ldots, H_4(g)\} = \{H_1, \ldots, H_4\}$ and the signs of the $g_i(\underline{x})$ are as desired. To see this notice for any $\underline{x} \varepsilon H^2$ we have that $\underline{x} \varepsilon H_j$ for some j, $1 \leq j \leq 4$. This tells us what $|g_i(\underline{x})|$ should be for i = 0, 1, 2. We can then assign signs to the $|g_i(\underline{x})|$ as we desire and use equations (3.13) and (3.16) to determine $g(x_1, x_2)$, $g(x_1, -x_2)$, $g(-x_1, x_2)$, and $g(-x_1, -x_2)$. Repeating this for every $\underline{x} \varepsilon H^2$ gives us the desired $g \varepsilon G(c)$

All this gives us

$$\sup_{g \varepsilon G(c)} L(\underline{B}, \xi, g) = \sup \sum_{i=0}^{2} [(c/2) \int_4 |B_i(\underline{x})| \, d\overline{\xi}(\underline{x}) + c \int_{i+1} |B_i(\underline{x})| \, d\overline{\xi}(\underline{x})]^2$$

$$+ \rho \int (\sum_{i=0}^{2} B_i^2(\underline{x})) \, d\overline{\xi}(\underline{x})$$

where the sup on the right hand side of this last equation is over all partitions $\{H_1, H_2, H_3, H_4\}$ of $H^2$ into four disjoint sets such that $(x_1, 0)' \varepsilon H_1 \cup H_2$ and $(0, x_2)' \varepsilon H_1 \cup H_3$.

Our goal is to establish a result analogous to this last equation for general k. The next two lemmas are used to accomplish this.

<u>LEMMA 3.4.</u> For each k there exists an integer n(k) and numbers $a_{ijk}$ with $0 \leq a_{ijk} \leq c$, i = 0, 1, ..., k, j = 1, 2, ..., n(k) satisfying the following relations.

(i) $a_{01k} = c$, $a_{11k} = a_{21k} = \dots = a_{k1k} = 0$ and given any $g \in G(c)$ and $\underline{x} \in H^k$ the only possible values for $g_i(\underline{x})$ are $\pm a_{ijk}$, $j = 1, \dots, n(k)$.

(ii) Given $g \in G(c)$, if we define

$$H_j(g) = \{\underline{x} \in H^k;\ |g_i(\underline{x})| = a_{ijk},\ i = 0, 1, \dots, k\}$$

for $j = 1, 2, \dots, n(k)$ then $\{H_1(g), \dots, H_{n(k)}(g)\}$ is a partition of $H^k$ into $n(k)$ disjoint sets.

(iii) $\sum_{i=0}^{k} a_{ijk}^2 \le c^2$ for all $j$.

(iv) Let $J(i, k) = \{j;\ 1 \le j \le n_k \text{ and } a_{ijk} = 0\}$.

Then every point $\underline{x} \in H^k$ whose i-th coordinate $x_i$ is 0, $i = 1, \dots, k$, satisfies $\underline{x} \in \bigcup_{j\ J(i,k)} H_j(g)$. Notice $1 \in J(i, k)$ for all i.

(v) For each $g \in G(c)$, if we let

$$\Gamma(g,k) = \{h \in G(c);\ \{H_1(h), \dots, H_{n(k)}(h)\} = \{H_1(g), \dots, H_{n(k)}(g)\}\}$$

then for any assignment of plus or minus signs, depending on i and $\underline{x}$, to the values $|g_i(\underline{x})|$, there is a $g^* \in \Gamma(g, k)$ such that $g_i^*(\underline{x})$ has these signs and $|g_i^*(\underline{x})| = |g_i(\underline{x})|$.

(vi) Suppose $\{H_1, \dots, H_{n(k)}\}$ is a partition of $H^k$ into $n(k)$ disjoint sets such that if $\underline{x} \in H^k$ has its i-th coordinate 0 then $\underline{x} \in \bigcup_{j \in J(i,k)} H_j$.

There exists $g \in G(c)$ such that $H_j(g) = H_j$ for all j.

<u>Proof.</u> To prove this rather complicated lemma we use an induction argument. The cases $k = 1$ and $k = 2$ have been examined in detail above and one can verify that the lemma is true in those cases.

Assume that Lemma 3.4 has been proved for all k up to $t \geq 2$. We show that it holds for $k = t + 1$. For purposes of this proof we use the notation

$$G(c, k) = \{g: I^k \to R; |g| = c\}$$

instead of our usual G(c) so as to keep track of the dimensionality involved.

Suppose $g \in G(c, t+1)$. For any $z \in I$ if we define $g_z(x_1, \ldots, x_t) = g(x_1, \ldots, x_t, z)$, then for fixed $z \in I$ $g_z \in G(c, t)$. From the definition of $g_i$ in (3.12) we see that for any $y \in H$

(3.17)
$$g_i(x_1, \ldots, x_t, y) = [g_{yi}(x_1, \ldots, x_t) + g_{-yi}(x_1, \ldots, x_t]/2$$
$$g_{t+1}(x_1, \ldots, x_t, y) = [g_{yo}(x_1, \ldots, x_t) - g_{-yo}(x_1, \ldots, x_t)]/2$$

for $i = 0, 1, \ldots, t$.

Applying the induction assumption one has the numbers $a_{ijt}$ and n(t). Notice by (3.17) that each $g_i$ can only take on the values $(a_{ijt} \pm a_{i\ell t})/2$ or $(-a_{ijt} \pm a_{i\ell t})/2$ for $i = 0, 1, \ldots, t$ and $1 \leq j, \ell \leq n(t)$. Also, $g_{t+1}$ can only take on the values $(a_{ojt} \pm a_{o\ell t})/2$ or $(-a_{ojt} \pm a_{o\ell t})/2$ for $1 \leq j, \ell \leq n(t)$.

Upon taking absolute values these numbers give rise to a new set of numbers n(t+1) and $a_{ijt+1}$ with $0 \leq a_{ijt+1} \leq c$, for $i = 0, 1, \ldots, t+1$ and $j = 1, 2, \ldots, n(t+1)$, such that the following properties hold:

(a) For any j such that $1 \leq j \leq n(t+1)$ there exists $\ell$ and m with $1 \leq \ell, m \leq n(t)$ satisfying either $a_{ijt+1} = |a_{i\ell t} + a_{imt}|/2$ for $i = 0, 1, \ldots, t$ and $a_{t+1jt+1} = |a_{o\ell t} - a_{omt}|/2$ or $a_{ijt+1} = |a_{i\ell t} + a_{imt}|/2$ for $i = 0, 1, \ldots, t$ or $a_{t+1jt+1} = |a_{o\ell t} + a_{omt}|/2$.

(b) The $a_{ijt+1}$ may be constructed so that when $j = 1$ the appropriate $\ell$ and m guaranteed in (a) are $\ell = m = 1$ (this can be done because every $\ell$ and m yield an $a_{ijt+1}$ as described above) and $a_{i1t+1} = |a_{i1t} + a_{i1t}|/2 = |a_{i1t}|$ for $i = 0, 1, \ldots, t$. Also $a_{t+1,1,t+1} = |a_{o1t} - a_{o1t}|/2 = 0$. Notice that since $a_{o1t} = c$ and $a_{11t} = \cdots = a_{t1t} = 0$, this yields $a_{o1t+1} = c$ and $a_{11t+1} = \cdots\ a_{t+1,1,t+1} = 0$.

(c) The only possible values for $g_i(\underline{x})$, $i = 0, 1, \ldots, t+1$, $\underline{x} \varepsilon H^{t+1}$, are $\pm a_{ijt+1}$ for $j = 1, \ldots, n(t+1)$.

(d) If we define $H_j(g) = \{\underline{x} \varepsilon H^{t+1};\ |g_i(\underline{x})| = a_{ijt+1},\ i = 0, 1, \ldots, t+1\}$ for $j = 1, \ldots, n(t+1)$ it is not hard to see that $\{H_1(g), \ldots, H_{n(t+1)}(g)\}$ is a partition of $H^{t+1}$ into $n(t+1)$ disjoint sets.

(e) Suppose $\underline{x} = (x_1, \ldots, x_t, y)' \varepsilon H^{t+1}$. Let $\{H_1(g), \ldots, H_{n(t+1)}(g)\}$ be the partition of $H^{t+1}$ defined in (d). Then $\underline{x} \varepsilon H_j(g)$ if and only if there exist $\ell$, m with $1 \leq \ell, m \leq n(t)$ such that $\ell$ and m are associated with j as in (a) and such that $(x_1, \ldots, x_t)' \varepsilon H_\ell(g_y) \cap H_m(g_{-y})$. Here $\{H_1(g_y), \ldots, H_{n(t)}(g_y)\}$ and $\{H_1(g_{-y}), \ldots, H_{n(t)}(_{-y})\}$ are the partitions of $H^t$ associated with $g_y$ and $g_{-y}$ respectively, as guaranteed by the induction assumption.

That properties (i) - (vi) of the lemma hold for the $a_{ijt+1}$ follows from the fact that they hold for the $a_{ijt}$, from equation (3.17), and from (a) - (e) above.

To see, for example, that (iii) holds, suppose $1 \leq j \leq n(t+1)$. Let $\ell$, m be associated with j as in (a). Assume $a_{ijt+1} = |a_{i\ell t} + a_{imt}|/2$ for $i = 0, 1, \ldots, t$ and $a_{t+1,j,t+1} = |a_{o\ell t} - a_{omt}|/2$.
Then

that if $\underline{x} \in H^k$ has its i-th coordinate equal to 0 then $\underline{x} \in \bigcup_{j \in J(i,k)} \Pi_j$.

We then find that

(3.19)
$$\sup_{g \in G(c)} L(\underline{B}, \xi, g) = \sup_{\Pi \in P} \Lambda\, (\underline{B}, \overline{\xi}, \Pi)$$

where

(3.20)
$$\Lambda(B, \overline{\xi}, \Pi) = \sum_{i=0}^{k} \sum_{j=1}^{n(k)} [a_{ijk} \int_j |B_i(\underline{x})| \, d\overline{\xi}(\underline{x})]^2 + \rho \int \sum_{i=0}^{k} B_i^2(\underline{x}) \, d\overline{\xi}(\underline{x}).$$

In (3.20) $\int_j$ is over $H_j$, the last integral is over $H^k$, $(\underline{B}, \xi) \in M$, and $\overline{\xi}$ is related to $\xi$ as defined below (3.13).

Proof. By (3.13) we have

$$L(\underline{B}, \xi, g) = \sum_{i=0}^{k} [\int B_i(\underline{x}) \, g_i(\underline{x}) \, d\overline{\xi}(\underline{x})]^2 = \rho \int \sum_{i=0}^{k} B_i^2(\underline{x}) \, d\overline{\xi}(\underline{x})$$
$$= \sum_{i=0}^{k} [\sum_{j=1}^{n(k)} \int_j B_i(\underline{x}) \, g_i(\underline{x}) \, d\overline{\xi}(\underline{x})]^2 + \rho \int \sum_{i=0}^{k} B_i^2(\underline{x}) \, d\overline{\xi}(\underline{x})$$

where $\int_j$ is over $H_j$, the j-th set in the partition $\Pi$ of $H^k$.

By Lemma 3.4 (v) we can find $g^* \in \Gamma(g,k)$ such that $i = 0, 1, \ldots, k$, $g_i^*(\underline{x})$ is $\geq 0$ (respectively $\leq 0$) if $B_i(\underline{x})$ is $\geq 0$ (respectively $\leq 0$). Simply assign a plus sign to $|g_i(\underline{x})|$ if and only if $B_i(\underline{x})$ is $\geq 0$ and then choose $g^* \in \Gamma(g,k)$ as in (v) of Lemma 3.4. Then

$$L(\underline{B}, \xi, g^*) = \sum_{i=0}^{k} [\sum_{j=1}^{n(k)} \int_j |B_i(\underline{x})| \; |g_i^*(\underline{x})| \; \overline{\xi}(\underline{x})]^2$$

$$+ \rho \int \sum_{i=0}^{k} B_i^2(\underline{x}) \; d\overline{\xi}(\underline{x})$$

$$= \sum_{i=0}^{k} [\sum_{j=1}^{n(k)} a_{ijk} \int_j |B_i(\underline{x})| \; d\overline{\xi}(\underline{x})]^2$$

$$+ \rho \int \sum_{i=0}^{k} B_i^2(\underline{x}) \; d\overline{\xi}(\underline{x})$$

$$\geq L(\underline{B}, \xi, g).$$

Here $\int_j$ is over $H_j(g)$. Now use (vi) of Lemma 3.4 to get (3.19). Q.E.D.

Notice that if $(\underline{B}, \xi) \varepsilon M$ and $\overline{\xi}$ is related to $\xi$ as defined below (3.13) then (3.6) is equivalent to

(3.21)
$$\int B_o(\underline{x}) \; d\overline{\xi}(\underline{x}) = 1$$

$$\int B_i(\underline{x}) \; d\overline{\xi}(\underline{x}) = \theta_i \qquad i = 1, \ldots, k$$

where the $\theta_i$ are as in (3.3) and the integrals are over $H^k$.

Let $\overline{D}$ be the set of all probability measures on $H^k$ having finite support. Let N be the set of all pairs $(\underline{B},\overline{\xi})$ where $\overline{\xi} \varepsilon \overline{D}$, $\underline{B} = (B_o, B_1, \ldots, B_k)'$ is a k+1 fold column vector of real valued functions on $H^k$ and $\overline{\xi}$ and the $B_i$ satisfy (3.21). We seek $(\underline{B}^*, \overline{\xi}^*) \varepsilon$ N such that

(3.22)
$$\sup_{\Pi \varepsilon P} \Lambda(\underline{B}^*, \overline{\xi}^*, \Pi) = \inf_{(\underline{B},\overline{\xi}) \varepsilon N} \sup_{\Pi \varepsilon P} \Lambda(\underline{B}, \overline{\xi}, \Pi).$$

Suppose that $(\underline{B}^*, \overline{\xi}^*) \varepsilon$ N satisfy (3.22). Let $\xi^* \varepsilon$ D be related to $\overline{\xi}^*$ as below (3.13) and let $\xi^*$ satisfy (3.11). Extend $\underline{B}^*$ to all of $I^k$ so that

(3.11) is satisfied. It is easily verified that $(\underline{B}^*, \xi^*) \varepsilon M$. Furthermore, it follows from Lemma 3.5 and the remarks following Lemma 3.3 that $\underline{B}^*, \xi^*$ minimize $\sup_{g\varepsilon G} L(\underline{B}, \xi, g)$ over all $\underline{B}$ and $\xi \varepsilon D$ satisfying (3.6). We therefore now seek $(\underline{B}^*, \overline{\xi}^*) \varepsilon N$ satisfying (3.22). The following two lemmas will be used in the solution of this problem.

<u>LEMMA 3.6.</u> Suppose $1 \leq i \leq k$ and suppose $\theta_i > 0$ is fixed. For $\overline{\xi} \varepsilon \overline{D}$ define

$$F_i(\overline{\xi}) = \{f:\ H^k \to R;\ f(\underline{x}) \geq 0 \text{ and } \int x_i\ f(\underline{x})\ d\overline{\xi}(\underline{x}) = \theta_i\}$$

Here $\underline{x} = (x_1, \ldots, x_k)'$ and the integrals are over $H^k$. Notice that if $\overline{\xi}$ is such that $\int x_i^2\ d\overline{\xi}(\underline{x}) = 0$ then $F_i(\overline{\xi})$ is empty unless $\theta_i = 0$. Define

$$\phi(\overline{\xi}, \underline{x}) = \theta_i\ x_i/\int x_i^2\ d\overline{\xi}(\underline{x}) \qquad \text{if} \quad \int x_i^2\ d\overline{\xi}(\underline{x}) > 0,$$

otherwise take $\phi(\overline{\xi}, \underline{x}) = 0$. Notice that $\phi(\overline{\xi}, \cdot) \varepsilon F_i(\overline{\xi})$ if $F_i(\overline{\xi})$ is non-empty.

For any $\alpha, \beta \geq 0$, $f \varepsilon F_i(\overline{\xi})$ define $L_i(f, \overline{\xi}) = \alpha(\int f(\underline{x})\ d\overline{\xi}(\underline{x}))^2 + \beta \int f^2(\underline{x})\ d\overline{\xi}(\underline{x})$. Let $N_i$ be the set of ordered pairs $(f,\overline{\xi})$ such that $\overline{\xi} \varepsilon \overline{D}$, $F_i(\overline{\xi})$ is non-empty, and $f \varepsilon F_i(\overline{\xi})$. Then

$$\inf L_i(f, \overline{\xi}) = L_i(\phi(\overline{\xi}^*, \cdot), \overline{\xi}^*)$$

where $\overline{\xi}^*$ puts all its mass on the point $(1, 1, \ldots, 1)'$ and the inf is over all $(f, \overline{\xi}) \varepsilon N_i$.

<u>Proof</u>. For any $\overline{\xi} \varepsilon \overline{D}$ such that $F_i(\overline{\xi})$ is non-empty let $f \varepsilon F_i(\overline{\xi})$. Also let $m_{2i}(\overline{\xi}) = \int x_i^2\ d\overline{\xi}(\underline{x})$. If $m_{2i}(\overline{\xi}) > 0$ then

$$\begin{aligned} \int f^2(\underline{x})\ d\overline{\xi}(\underline{x}) &= [\int f^2(\underline{x})\ d\overline{\xi}(\underline{x})]\ [\int (x_i^2/m_{2i}(\overline{\xi}))\ d\overline{\xi}(\underline{x})] \\ &\geq [\int f(\underline{x})\ \left(x_i/\sqrt{m_{2i}(\overline{\xi})}\right)\ d\overline{\xi}(\underline{x})]^2 \\ &= \theta_i^2/m_{2i}(\overline{\xi}) \end{aligned}$$

where the inequality in the second line comes from the Cauchy-Schwartz inequality and the last line follows from the fact that $f \in F_i(\overline{\xi})$. The inequality above will be equality if and only if $f(\underline{x}) = ax_i$ on supp $\overline{\xi}$ for some $a \in R$. In fact we must have $a = \theta_i/m_{2i}(\overline{\xi})$ if there is to be equality above. Thus when $m_{2i}(\overline{\xi}) > 0$ $\int f^2(\underline{x})\, d\overline{\ }(\underline{x})$ is minimized over $F_i(\overline{\xi})$ by $\phi(\overline{\xi}, \underline{x}) = \theta_i\, x_i/m_{2i}(\overline{\xi})$.

If $m_{2i}(\overline{\xi}) = 0$ (and so it must be the case that $\theta_i = 0$) it is clear that $\phi(\overline{\xi}, \underline{x}) = 0$ minimizes $\int f^2(\underline{x})\, d\overline{\xi}(\underline{x})$ over $F_i(\overline{\xi})$.

Notice for $(f, \overline{\xi}) \in N_i$ that

$$[\int f(\underline{x})\, d\overline{\xi}(\underline{x})]^2 \leq \int f^2(\underline{x})\, d\overline{\xi}(\underline{x}).$$

We have when $m_{2i}(\overline{\xi}) > 0$ that

$$\begin{aligned} L_i(f, \overline{\xi}) &= \alpha[\int f(\underline{x})\, d\overline{\xi}(\underline{x})]^2 + \beta \int f^2(\underline{x})\, d\overline{\xi}(\underline{x}) \\ &\geq \alpha[\int x_i\, f(\underline{x})\, d\overline{\xi}(\underline{x})]^2 + \beta \int f^2(\underline{x})\, d\overline{\xi}(\underline{x}) \\ &= \alpha\, \theta_i^2 + \beta \int f^2(\underline{x})\, d\overline{\xi}(\underline{x}) \\ &\geq \alpha\, \theta_i^2 + \beta \int \phi^2(\overline{\xi}, \underline{x})\, d\overline{\xi}(\underline{x}) \\ &= \alpha\, \theta_i^2 + \beta\, \theta_i^2/m_{2i}(\overline{\xi}) \\ &\geq \theta_i^2\, [\alpha + \beta] \end{aligned}$$

for all $f \in F_i(\overline{\xi})$. Notice that the right hand lower bound is also valid when $m_{2i}(\overline{\xi}) = 0$ since this can only happen if $\theta_i = 0$. Clearly this lower bound is attained for $L_i(f, \overline{\xi})$ when $f = \phi(\overline{\xi}, \cdot)$ and $\overline{\xi} = \overline{\xi}^*$. Q.E.D.

LEMMA 3.7. Let $N^+$ be the set of all $(\underline{B}, \overline{\xi}) \in N$ such that $B_i \geq 0$ for all $i$ where $\underline{B} = (B_o, B_1, \ldots, B_k)'$. Let $\pi$ be any partition of $H^k$ into $n(k)$ disjoint sets. Then

$$\inf_{(\underline{B},\overline{\xi}) \in N^+} \Lambda(\underline{B}, \overline{\xi}, \pi) = \inf_{(\underline{B},\overline{\xi}) \in N} \Lambda(\underline{B}, \overline{\xi}, \pi)$$

Proof. Suppose $(\underline{B}, \overline{\xi}) \in N$ and $\underline{B} = (B_o, B_1, \ldots, B_k)'$. Let $U_i = \{\underline{x} \in H^k; B_i(\underline{x}) < 0\}$ for $i = 0, 1, \ldots, k$. Define

$$\overline{B}_o(\underline{x}) = B_o(\underline{x})/(1 - \int_{U_o} B_o(\underline{z})\, d\overline{\xi}(\underline{z})) \quad \text{if } \underline{x} \notin U_o$$

$$\overline{B}_o(\underline{x}) = 0 \qquad \text{if } \underline{x} \in U_o$$

$$\overline{B}_i(\underline{x}) = B_i(\underline{x})/(1 - \int_{U_i} z_i\, B_i(\underline{z})\, d\overline{\xi}(\underline{z})) \quad \text{if } \underline{x} \notin U_i$$

$$\overline{B}_i(\underline{x}) = 0 \text{ if } \underline{x} \in U_i$$

for $i = 1, \ldots, k$. We have that

$$|\overline{B}_i(\underline{x})| = \overline{B}_i(\underline{x}) \leq |B_i(\underline{x})| \text{ for all } \underline{x} \in H^k.$$

Clearly if $\overline{\underline{B}} = (\overline{B}_o, \overline{B}_1, \ldots, \overline{B}_k)'$ then $\Lambda(\overline{\underline{B}}, \overline{\xi}, \pi) \leq \Lambda(\underline{B}, \overline{\xi}, \pi)$ for any partition $\pi$ of $H^k$ into $n(k)$ disjoint sets. It is easy to verify that if $\underline{B}$ and $\overline{\xi}$ satisfy (3.21) then so do $\overline{\underline{B}}$ and $\overline{\xi}$. Hence $(\overline{\underline{B}}, \overline{\xi}) \in N^+$ if $(\underline{B}, \overline{\xi}) \in N$. The lemma follows. Q.E.D.

Now let $P^*$ be the set of all partitions $\pi = \{H_1, \ldots, H_{n(k)}\}$ of $H^k$ into $n(k)$ disjoint sets. Also let $\pi_j = \{H_{1j}, \ldots, H_{n(k)j}\}$ be that element of $P^*$ having $H_{jj} = H^k$ and all other $H_{ij}$ empty, where $1 \leq j \leq n(k)$.

Before solving the problem posed at equation (3.22) we first solve a different problem. We seek $(\underline{B}^*, \overline{\xi}^*) \in N^+$ which satisfies

(3.23) $$\sup_{1 \leq j \leq n_k} \Lambda(\underline{B}^*, \overline{\xi}, \pi_j) = \inf_{(\underline{B}, \overline{\xi}) \in N^+} \sup_{1 \leq j \leq n(k)} \Lambda(\underline{B}, \overline{\xi}, \pi_j)$$

To do this notice that for any $(\underline{B}, \overline{\xi}) \in N^+$ we have

(3.24) $$\Lambda(\underline{B}, \overline{\xi}, \pi_j) = \sum_{i=0}^{k} [a_{ijk} \int |B_i(\underline{x})|\, d\overline{\xi}(\underline{x})]^2 + \rho \int \sum_{i=0}^{k} B_i^2(\underline{x})\, d\overline{\xi}(\underline{x})$$

$$= \sum_{i=0}^{k} a_{ijk}^2 (\int B_i(\underline{x})\, d\overline{\xi}(\underline{x}) + \rho \int \sum_{i=0}^{k} B_i^2(\underline{x})\, d\overline{\xi}(\underline{x}).$$

Using the Cauchy-Schwartz inequality we can write

$$\Lambda(\underline{B}, \overline{\xi}, \pi_j) = a_{ojk}^2 + \rho \int B_o^2(\underline{x})\, d\overline{\xi}(\underline{x})$$

$$+ \sum_{i=1}^{k} a_{ijk}^2 \left(\int B_i(\underline{x})\, d\overline{\xi}(\underline{x})\right)^2 + \rho \int \sum_{i=1}^{k} B_i^2(\underline{x})\, d\overline{\xi}(\underline{x})$$

$$\geq a_{ojk}^2 + \rho + \sum_{i=1}^{k} a_{ijk}^2 \left(\int B_i(\underline{x})\, d\overline{\xi}(\underline{x})\right)^2$$

$$+ \rho \int \sum_{i=1}^{k} B_i^2(\underline{x})\, d\overline{\xi}(\underline{x})$$

where we have used the fact that $\int B_o(\underline{x})\, d\overline{\xi}(\underline{x}) = 1$ by (3.21). Notice that we have equality in this last equation if and only if $B_o(\underline{x}) = 1$ on supp $\overline{\xi}$.

This observation and Lemma (3.6) give us that $\Lambda(\underline{B}, \overline{\xi}, \pi_j)$ is minimized over $N^+$ by $\underline{B}^* = (B_o^*, B_1^*, \ldots, B_k^*)'$ and $\overline{\xi}^*$ where

(3.25)
$$\overline{\xi}^*(1,\ldots, 1)' = 1$$

$$B_o^*(\underline{x}) = 1$$

$$B_i^*(\underline{x}) = \theta_i\, x_i / \int x_i^2\, d\overline{\xi}^*(\underline{x}) = \theta_i\, x_i, \quad i = 1, \ldots, k.$$

This holds for each j and hence is the solution to (3.23). Furthermore if we evaluate $\Lambda(\underline{B}^*, \overline{\xi}^*, \pi_j)$ using (3.24) for $j = 1, \ldots, n(k)$ we get

(3.26)
$$\Lambda(B^*, \overline{\xi}^*, \pi_j) = \sum_{i=0}^{k} \theta_i^2\, a_{ijk}^2 + \rho \sum_{i=0}^{k} \theta_i^2$$

Recall that $\theta_o = 1$, $0 \leq \theta_i \leq 1$ for $i = 1, \ldots, k$. By Lemma 3.4 we have that $\sum_{i=0}^{k} a_{ijk}^2 \leq c^2$ for all j and in addition that $a_{01k} = c$, $a_{11k} = \cdots = a_{k1k} = 0$. Thus we see that $\sum_{i=0}^{k} \theta_i^2\, a_{ijk}^2 \leq c^2$ for all j and $\sum_{i=0}^{k} \theta_i^2\, a_{i1k}^2 = c^2$.

Hence

(3.27) $$\sup_{1\le j\le n(k)} \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_j) = \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1) = c^2 + \rho \sum_{i=0}^{k} \theta_i^2$$

Next notice that by Lemma 3.4 (iv) and the definition of P (see Lemma 3.5) $\pi_1 \varepsilon P$. We therefore have

(3.28) $$\begin{aligned}\Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1) &= \sup_{1\le j\le n(k)} \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_j) \\ &= \inf_{(\underline{B},\overline{\xi})\varepsilon N^+} \sup_{1\le j\le n(k)} \Lambda(\underline{B}, \overline{\xi}, \pi_j) \\ &\le \inf_{(\underline{B},\overline{\xi})\varepsilon N^+} \sup_{\pi\varepsilon P}{}_* \Lambda(\underline{B}, \overline{\xi}, \pi) \\ &\le \sup_{\pi\varepsilon P}{}_* \Lambda(\underline{B}^*, \overline{\xi}^*, \pi) \\ &= \sup_{1\le j\le n(k)} \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_j) \\ &= \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1)\end{aligned}$$

where the next to last equality comes from the fact that $\overline{\xi}^*$ puts all its mass on $(1, \ldots, 1)'$ and only one of the $H_j$ in any partition $\pi = \{H_1, \ldots, H_{n(k)}\}$ in $P^*$ can contain this point. Therefore all the inequalities in (3.28) are equalities and we get

$$\begin{aligned}\Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1) &= \inf_{(\underline{B},\overline{\xi})\varepsilon N^+} \Lambda(\underline{B}, \overline{\xi}, \pi_1) \\ &\le \inf_{(\underline{B},\overline{\xi})\varepsilon N^+} \sup_{\pi\varepsilon P} \Lambda(\underline{B}, \overline{\xi}, \pi) \\ &\le \inf_{(\underline{B},\overline{\xi})\varepsilon N^+} \sup_{\pi\varepsilon P^*}{}_* \Lambda(\underline{B}, \overline{\xi}, \pi) \\ &= \Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1)\end{aligned}$$

so

$$\Lambda(\underline{B}^*, \overline{\xi}^*, \pi_1) = \inf_{(\underline{B},\overline{\xi})\in N^+} \sup_{\pi\in P} \Lambda(\underline{B}, \overline{\xi}, \pi) \tag{3.29}$$

Utilizing (3.29), Lemma 3.5, Lemma 3.7, and the discussion around (3.13) we see that $\underline{B}^*$ and the probability measure $\xi^*$ on $I^k$ which puts equal mass $1/2^k$ on all $2^k$ points $\underline{x} = (x_1, \ldots, x_k)'$ with $|x_1| = \ldots = |x_k| = 1$ solves the original minimax problem stated before Lemma 3.1.

Summarizing we have the following theorem.

THEOREM 3.1. Let $B_o^*(\underline{x}) = 1$, $B_i^*(\underline{x}) = \theta_i x_i$, for $i = 1, \ldots, k$, and let $\xi^*$ be the probability measure on $I^k$ putting mass $1/2^k$ on each of the $2^k$ points $\underline{x} = (x_1, \ldots, x_k)'$ such that $|x_1| = \ldots = |x_k| = 1$. Suppose M is the set of all ordered pairs $(\underline{B}, \xi)$ where $\underline{B} = (B_o, B_1, \ldots, B_k)'$ and the $B_i$ are real valued functions on $I^k$, $\xi \in D$ is a probability measure on $I^k$, and the $B_i$ and $\xi$ satisfy (3.6). Then

$$\sup_{g\in G} L(\underline{B}^*, \xi^*, g) = \inf_{(\underline{B},\xi)\in M} \sup_{g\in G} L(\underline{B}, \xi, g) = c^2 + \rho \sum_{i=0}^{k} \theta_i^2$$

Q.E.D.

Recalling the discussion around equations (3.2) through (3.6) we see that when $\theta_o = 1$ and $0 < \theta_i \leq 1$ for $i = 1, \ldots, k$ the estimators

$$\hat{\beta}_i^* = \int Y(\underline{x})\, b_i^*(\underline{x})\, d\xi^*(\underline{x}), \quad i = 0, 1, \ldots, k,$$

where $b_i^* = B_i^*/\theta_i$, minimize

$$\sup_{g\in G} \sum_{i=0}^{k} \theta_i^2\, E(\hat{\beta}_i - \beta_i)^2$$

subject to (3.6). Notice these $\hat{\beta}_i^*$ are just the usual best linear unbiased estimates with respect to $\xi^*$ and $\xi^*$ is an optimal design, under a broad class of criteria, for k-variate linear regression on $I^k$.

Now we restrict to the case $\theta_o = 1 \geq \theta_1 = \dots = \theta_k = \theta > 0$. The minimax mean square error $\sup_{g \varepsilon G} L(\underline{B}^*, \xi^*, g)$ can be written as

$$\sup_{g\varepsilon G} L(\underline{B}^*, \xi^*, g) = \sup_{g\varepsilon G(c)} \left[\left(\int g(\underline{x})\, d\xi^*(\underline{x})\right)^2 + \theta^2 \sum_{i=1}^{k} \left(\int x_i\, g(\underline{x})\, d\xi^*(\underline{x})\right)^2\right] + \rho \int \left(1 + \theta^2 \sum_{i=1}^{k} x_i^2\right) d\xi^*(\underline{x})$$

Let $\underline{x}_j = (x_{1j}, \dots, x_{kj})' \varepsilon I^k$ for $j = 1, \dots, m$ be any set of m distinct points in $I^k$. Let $\xi_o$ be the probability measure putting mass 1/m on each of these m points. Define

$$X(\xi_o) = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \underline{x}_1 & \cdot & \cdot & \cdot & \underline{x}_m \end{bmatrix} \tag{3.30}$$

$$g(\xi_o) = (g(\underline{x}_1), \dots, g(\underline{x}_m))'.$$

Notice $X(\xi_o)$ is a $(k+1) \times m$ matrix and $g(\xi_o)$ is a $m \times 1$ column vector. If we use $J_m$ to denote the $m \times m$ matrix all of whose entries are 1, we get

$$\sup_{g\varepsilon G} L(\underline{B}^*, \xi_o, g) = \sup_{g\varepsilon G(c)} \left[\left(\int g(\underline{x})\, d\xi_o(\underline{x})\right)^2 + \theta^2 \sum_{i=1}^{k} \left(\int x_i\, g(\underline{x})\, d\xi_o(\underline{x})\right)^2\right] + \rho \int \left(1 + \theta^2 \sum_{i=1}^{k} x_i^2\right) d\xi_o(\underline{x})$$

$$= \sup_{g\varepsilon G(c)} \left[(1 - \theta^2)\left(\int g(\underline{x})\, d\xi_o(\underline{x})\right)^2 + \theta^2\left(\int g(\underline{x})\, d\xi_o(\underline{x})\right)^2 + \theta^2 \sum_{i=1}^{k} \left(\int x_i\, g(\underline{x})\, d\xi_o(\underline{x})\right)^2\right] + \rho(1 - \theta^2) + \theta^2 \rho \int \left(1 + \sum_{i=1}^{k} x_i^2\right) d\xi_o(\underline{x})$$

$$= \sup_{g \varepsilon G(c)} [(1 - \theta^2) g'(\xi_o) J_m g(\xi_o)$$

$$+ \theta^2 g'(\xi_o) X'(\xi_o) X(\xi_o) g(\xi_o)]/m^2$$

$$+ (1 - \theta^2)\rho + (\theta^2\rho \text{ tr } X(\xi_o) X'(\xi_o))/m$$

$$\geq \sup_{g \varepsilon G(c)} L(\underline{B}^*, \xi^*, g)$$

$$= \sup_{g \varepsilon G} L(\underline{B}^*, \xi^*, g)$$

$$= c^2 + \rho + k\theta^2\rho.$$

If k+1 is such that a (k+1) × (k+1) Hadamard matrix X exists (in standard form so that the first row and column are all +1) then any exact design $\psi$ on k+1 points whose support is such that $X(\psi) = X$, where $X(\psi)$ is as in (3.30), satisfies

$$\sup_{g \varepsilon G} L(\underline{B}^*, \psi, g) = \sup_{g \varepsilon G(c)} [(c - \theta^2) g'(\psi) J_{k+1} g(\psi)$$

$$+ \theta^2 g'(\psi) X'(\psi) X(\psi) g(\psi)]/(k+1)^2$$

$$+ (1 - \theta^2)\rho + (\theta^2\rho \text{ tr } X(\psi) X'(\psi))/(k+1).$$

Recalling that since $X(\psi)$ is a Hadamard matrix $X(\psi) X'(\psi) = X'(\psi) X(\psi) =$ (k+1) diag (1, ..., 1) = the diagonal matrix all of whose diagonal entries are k+1, we have that

$$\sup_{g \varepsilon G} L(\underline{B}^*, \psi, g) = \sup_{g \varepsilon G(c)} [(1 - \theta^2) g'(\psi) J_{k+1} g(\psi) + \theta^2(k + 1) g'(\psi) g(\psi)]/(k+1)^2$$

$$+ (1 - \theta^2)\rho + \theta^2\rho(k + 1)^2/(k + 1)$$

$$= [(1 - \theta^2)c^2 (k+1)^2 + \theta^2(k+1)^2c^2]/(k+1)^2$$

$$+ (1 - \theta^2)\rho + \theta^2\rho(k+1)$$

$$= c^2 + \rho + k\theta^2\rho$$

where we have used the fact that if $g \in G(c)$ then $|g(\underline{x})| = c$ for all $\underline{x} \in I^k$.

We see that $\psi$ gives the same minimax value as $\xi^*$. Since $\underline{B}^*$ and $\psi$ satisfy (3.6) we have:

THEOREM 3.2. Suppose $\theta_o = 1 \geq \theta_1 = \dots = \theta_k = \theta > 0$. Suppose k+1 is such that a $(k+1) \times (k+1)$ Hadamard matrix exists. Let $\underline{B}^* = (B_o^*, B_1^*, \dots, B_k^*)'$ where $B_o^*(\underline{x}) = 1$, $B_i^*(\underline{x}) = \theta x_i$ for $i = 1, \dots, k$. Let $\psi$ be an exact design supported on k+1 points in $I^k$ such that $X(\psi)$, as defined in (3.30), is a Hadamard matrix in standard form. Then we have

$$\sup_{g \in G} L(\underline{B}^*, \psi, g) = \inf_{(\underline{B},\xi) \in M} \sup_{g \in G} L(\underline{B}, \xi, g)$$

where M is as in Theorem 3.1.

Q.E.D.

Theorem 3.2 allows one to reduce the support of a minimax design in special cases.

REMARK. Suppose A is a Lebesgue measurable set in $I^k$, A is invariant under coordinate reflections, and $(1, \dots, 1)' \in A$. Then the above arguments work when we restrict $\underline{x} \in A$ and Theorems 3.1 and 3.2 again hold.

## 4. Discussion

In Theorems 2.1, 3.1 and 3.2 we have restricted ourselves to considering only designs $\xi$ which have finite support. Certainly such a restriction is necessary if we want our results to be of any practical value. From a theoretical standpoint, though, one might wish to know whether Theorems 2.1, 3.1 and 3.2 hold if we remove the restriction that $\xi$ have finite support. The answer in general to this is no, at least for the minimax problem we have posed.

To see this consider the model of section 3 for the k-dimensional cube. Using the same notation as in section 3, let $\lambda$ denote the probability measure on $I^k$ that is $1/2^k$ times the value of Lebesgue measure on $I^k$. Let $\underline{\overline{B}} = (\overline{B}_o, \overline{B}_1, \ldots, \overline{B}_k)'$ where the $\overline{B}_i$ are the best linear unbiased estimators with respect to $\lambda$, i.e., $\overline{B}_o(\underline{x}) = 1$, $\overline{B}_i(\underline{x}) = 3x_i$ for $i = 1, \ldots, k$. Notice that $\underline{\overline{B}}$ and $\lambda$ satisfy (3.6) but $\lambda \notin D$.

Now for any $g \in G$ we have that

$$\int g(\underline{x})\ d\underline{x} = \int x_i\ g(\underline{x})\ d\underline{x} = 0 \qquad i = 1, \ldots, k.$$

Here all integrals are over $I^k$. Since $d\lambda(\underline{x}) = d\underline{x}/2^k$, we have for any $g \in G$

$$\begin{aligned}
(4.1) \qquad L(\underline{\overline{B}}, \lambda, g) &= \sum_{i=0}^{k} \left(\int \overline{B}_i(\underline{x})\ g(\underline{x})\ d\underline{x}\right)^2/2^{2k} + \rho \sum_{i=0}^{k} \overline{B}_i^2(\underline{x})\ d\underline{x}/2^k \\
&= \left(\int g(\underline{x})\ d\underline{x}\right)^2/2^{2k} + \sum_{i=1}^{k} 9\left(\int x_i\ g(\underline{x})\ d\underline{x}\right)^2/2^{2k} \\
&\quad + \rho \int d\underline{x}/2^k + \rho \sum_{i=1}^{k} \int 9x_i^2\ d\underline{x}/2^k \\
&= 0 + \rho + \rho \sum_{i=1}^{k} 3 \\
&= \rho\ (3k + 1).
\end{aligned}$$

Hence $\sup_{g \in G} L(\underline{\overline{B}}, \lambda, g) = \rho(3k + 1)$

Now the best discrete design $\xi^*$ and estimators $\underline{B}^*$ yield, by Theorem 3.1, $\sup_{g \in G} L(\underline{B}^*, \xi^*, g) = c^2 + \rho(k + 1)$. We thus see that whether $\sup_{g \in G} L(\underline{\overline{B}}, \lambda, g)$ is larger or smaller than $\sup_{g \in G} l(\underline{B}^*, \xi^*, g)$ depends on the relative sizes of $c$ and $\rho$. Only when $c^2 < 2k\rho$ is $\sup_{g \in G} L(\underline{B}^*, \xi^*, g)$ smaller than $\sup_{g \in G} L(\underline{\overline{B}}, \lambda, g)$.

Hence the results of Theorems 3.1 and 3.2 are not in general valid if we do not restrict attention to designs having finite support.

The point in the development of section 3 where the finiteness of the support of the designs is crucially used is when Lemma 3.3 is applied above equation (3.14) to pass from considering the sup over G to considering the sup over G(c). If from the beginning we allow the contaminations g in our model (3.1) to be in G(c) rather than G the development of section 3 appears to go through with all but minor changes and Theorems 3.1 and 3.2 would hold for any design whether of finite support or not. The reason for restricting the contaminations g to be in G rather than G(c) is so that the parameters being estimated, the $\beta_i$, are uniquely determined. When the contamination g is allowed to be in G(c), it is not clear what is being estimated in (3.1) since the $\beta_i$ are no longer unique. For example let $g_1(\underline{x}) = c$, $g_2(\underline{x}) = -c$. Both $g_1$ and $g_2$ are in G(c) and

(4.2)
$$\beta_o + \beta' \underline{x} + g_1(\underline{x}) = (\beta_o + 2c) + \beta' \underline{x} + g_2(\underline{x})$$

for all $\underline{x} \in I^k$ and $\beta_o$ is not uniquely determined in the model.

Some restrictions must therefore be placed on the allowable contaminations. G has been chosen as the set of allowable contaminations since it seems consistent with the idea of least squares and because it allows one to prove a rather general result. Other sets of allowable contaminations are possible. For example, see Marcus and Sacks (1976). However, once some restriction is made so as to insure the uniqueness of the $\beta_i$ in (3.1) one runs into the problem of possibly finding designs which are absolutely continuous with respect to Lebesgue measure and which are better than the best discrete designs. To see why this is so notice that the "worst" contamination (the one which maximizes the inf of $L(\underline{B}, \xi, g)$) is that g which equals c (or -c) everywhere on the support of the design $\xi$. If $\xi$ has finite support then such a g can be found in G. If $\xi$ is absolutely continuous with respect to Lebesgue measure on $I^k$ then such a g would have to be identically equal to c (or -c) on $I^k$. Any set of contaminations containing this g as well as the contamination g = 0 would give rise to the sort of problem indicated in (4.2). Since it is desirable that the set of allowable contaminations contain g = 0 it must exclude g = c (or g = -c) if uniqueness of the $\beta_i$ is to be insured. Then the minimax problem for absolutely continuous designs becomes different than for discrete designs since the worst contamination for the absolutely continuous case, namely g = c, is not allowable.

Two further questions worth discussing are whether the techniques used here are applicable to regions invariant under groups other than the group of coordinate reflections or the group of coordinate permutations, or whether the technique can be applied to regions other than cubes or simplexes which are invariant under coordinate reflections and permutations.

The answer to the first questions seems to be no. This is due to the structure of our squared error (3.7). This error function transforms nicely under coordinate reflections, as in section 3, and coordinate permutations, as in section 2. It does not appear to behave so nicely under other typical group actions such as rotations. Perhaps the use of a different error function would allow the techniques used here to be extended.

The answer to the second question is not clear. For the k-dimensional cube we were able, by invariance, to reduce to a problem where the best design had only one point in its support, namely the point $(1, \ldots, 1)'$, and the best estimators were easy to find. This was the case because the problem was reduced to considering the region $H^k$ and the point $(1, \ldots, 1)'$ has all its coordinates simultaneously as large as possible. Hence the design putting all its mass on $(1, \ldots, 1)'$ has all its second moments as large as possible. Again, for the simplex it was easy to see that the design putting all its mass on the k+1 corners of the simplex had all its second moments simultaneously as large as possible for any design invariant under coordinate permutations. Also the best estimators were easy to determine on the support of such a design. In both the cases for the cube and the simplex an obvious solution presented itself after applying invariance under the group operations. Unfortunately, things don't seem to proceed so smoothly for other regions such as the k-dimensional ball. Solutions may exist for other regions but they appear difficult to obtain.

5. Acknowledgements. The author wishes to express his sincere thanks to Professor Jack Kiefer for his guidance, support, encouragement, and many helpful discussions.

# REFERENCES

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

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

[3] Kiefer, J. (1973). Optimal designs for fitting biased multiresponse surfaces. *Multivariate Analysis III*, Academic Press, New York, 245-268.

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