GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION
SPONSORED PROJECT INITIATION

Date: July 13, 1976

Project Title: "Optimum Designs When the Basic Observations are Sample Paths of a Stochastic Process"

Project No: G-37-606

Project Director: Dr. Marcus C. Spruill

Sponsor: National Science Foundation

Agreement Period: From 7/1/76 Until 12/31/77*

(*12 month budget period plus 6 months for submission of required reports,etc.)

Type Agreement: Grant No. MCS76-11040

Amount:

$5,700	NSF	
1,762	GIT	(G-37-316)
$7,462	TOTAL	

Reports Required:

Annual Letter Technical: Final Report

Sponsor Contact Person (s):

Technical Matters

James L.Bostick
Grants Officer
National Science Foundation
Washington, D. C. 20550

Contractual Matters
(thru OCA)

John R. Pasta
Division Director
Mathematical and Computer Sciences
National Science Foundation
Washington, D.C. 20550

Defense Priority Rating: none

Assigned to: Mathematics (School/Laboratory)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director–EES
Accounting Office
Procurement Office
Security Coordinator (OCA)
Reports Coordinator (OCA)
Library, Technical Reports Section
Office of Computing Services
Director, Physical Plant
EES Information Office
Project File (OCA)
Project Code (GTRI)
Other ____________

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION

Date: 12/17/80

Project Title: Optimum Designs When The Basic Observations Are Sample Paths Of A Stochastic Process

Project No: G-37-606

Project Director: Dr. Marcus C. Spruill

Sponsor: National Science Foundation

Effective Termination Date: 12/31/79

Clearance of Accounting Charges: 12/31/79

Grant/Contract Closeout Actions Remaining:

- _ Final Invoice and Closing Documents
- X Final Fiscal Report via FCTR
- X Final Report of Inventions
- _ Govt. Property Inventory & Related Certificate
- _ Classified Material Certificate
- _ Other ____________

Assigned to: Mathematics (School/~~Laboratory~~)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director—EES
Accounting Office
Procurement Office
Security Coordinator (OCA)
Reports Coordinator (OCA)
Library, Technical Reports Section
EES Information Office
Project File (OCA)
Project Code (GTRI)
Other C. E. Smith

E-37-606

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA GEORGIA 30332

MATHEMATICS

June 28, 1977

Central Processing Section
National Science Foundation
Washington, D.C. 20550

Attention: Dr. Julius R. Blum, Program Director
Statistics Program
Mathematical Sciences Section
Division of Mathematical and Computer Science

Subject: Annual Technical Letter

Dear Dr. Blum:

Some results of research conducted under NSF Grant No. MCS 76 - 11040, "Optimum designs when the basic observations are sample paths of a stochastic process" are summarized below. The summary consists essentially of the abstracts of two papers which have been submitted for publication. Preprints with more detailed results are available. If you desire them at this time, please let me know.

Let the process $\{Y(x,t) : t\varepsilon T\}$ be observable for each x in some compact set X. Assume that $Y(x,t) = \theta_0 f_0(x)(t)+\cdots+\theta_k f_k(x)(t) + N(t)$ where f_j are continuous functions from X into the reproducing kernel Hilbert space H of the mean zero random process N. The optimum designs for estimating $c'\theta$ are characterized by an Elfving's theorem with $\mathcal{R} = \overline{co}\{(\phi,\underset{\sim}{f}(x))_H : ||\phi||_H \leq 1, x\varepsilon X\}$ where $(\cdot,\cdot)_H$ is the inner product on H. It is shown that if X is convex and f_j are linear design points may be chosen from the extreme points of X. In some problems each linear functional $c'\theta$ can be optimally estimated by a design on one point $x(c)$. These problems are completely characterized. An example is worked and some partial results on minimax designs are obtained.

Elfving's Theorem is not very useful in three or more dimensions. One of the iterative procedures referred to below can be used to find an approximately optimal design. It has some desirable characteristics not possessed by existing iterative techniques.

Let K be a convex set in the Hilbert space H and the ray $\{\alpha c : \alpha \varepsilon \mathcal{R}\}$ "puncture" K at β_*. Each algorithm results in a non-increasing sequence $\{\beta_j\}$ which converges to β_*. The points $\beta_j c$ lie in successive supporting hyperplanes to K. The normal to the the $n^{\underline{th}}$ hyperplane is obtained by a minimization over a set no larger than the unit n-cube. It is assumed that the subset K which maximizes (ϕ, x) for x in K is relatively easily found.

Sincerely,

Carl Spruill

CS:mc

G-37-606

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332

June 19, 1978

MATHEMATICS

Central Processing Section
National Science Foundation
Washington, D.C. 20550

Attention: Dr. Julius R. Blum, Program Director
Statistics Program
Mathematical Sciences Section
Division of Mathematical and Computer Science

Subject: Annual Technical Letter

Dear Dr. Blum:

To date the results of research conducted under NSF Grant No. MCS76-11040A01 fall into two general categories.

The first is related to H(K)-valued estimators of the mean function $\{\sum_{j=0}^{k} \theta_j f_j(x,t) : t \in T\}$, where $\underset{\sim}{\theta}$ is unknown. We were able to find the best linear unbiased estimator $\hat{m}$ of $\underset{\sim}{\theta}' \underset{\sim}{f}(x)$, where x is fixed, and to prove an exact analogue of the Kiefer-Wolfowitz theorem on D-optimum designs.

The second category of results is those related to the estimation of linear functionals of the unknown parameter θ for the model

$$Y(x,t) = m_x(\theta,t) + \varepsilon(t) \qquad t \in T,$$

where for each $x \in X$ the maps m_x are linear from the space Θ to H(K). The space Θ is linear but otherwise arbitrary.

Preprints containing the detailed results in each of these categories are available. Please let me know if you would like copies at this time.

Sincerely,

Carl Spruill

CS:jj

G-37-606

FINAL TECHNICAL REPORT

OPTIMAL EXPERIMENTAL DESIGNS FOR SECOND ORDER PROCESSES

By

Carl Spruill

Prepared for

NATIONAL SCIENCE FOUNDATION
GRANT NO. MCS76–11049–A01

GEORGIA INSTITUTE OF TECHNOLOGY

SCHOOL OF MATHEMATICS
ATLANTA, GEORGIA 30332

1980

OPTIMAL EXPERIMENTAL DESIGNS
FOR SECOND ORDER PROCESSES

by

Carl Spruill
Georgia Institute of Technology

Final technical report of research conducted under National Science Foundation Grant No. MCS76-11049-A01.

Contents

1. Background.

Recorded below are the results of our investigations into some regression design problems. We assume a model $y = m(\theta,x) + \varepsilon$ where y is a second order process, ε is a zero mean second order process with known covariance kernel, and the mean function m is linear in the unknown parameter. The parameter x indexes a set of possible experiments. In the case that y, m, and ε are scalars and m is of the form $\sum_{i=0}^{k} \theta_i f_i(x)$, where $\{f_i\}_{i=0}^{k}$ are known functions, regression design problems have been thouroughly investigated. See, for example, Elfving (1952), (1954), (1959); Chernoff (1972), Fedorov (1972), Karlin and Studden (1966), Kiefer (1960), (1961), (1974), Kiefer and Wolfowitz (1960), and Studden (1971). The problems we consider are either direct analogues of those in the scalar case or naturally motivated by the richer structure of the observations and parameter space.

The statistical basis for our analysis is provided by the elegant work of E. Parzen (1959) whose methods utilized and further developed the theory of reproducing kernel Hilbert spaces as found in Aronszajn (1950). Using this theory to handle unbiased estimation problems convenient expressions for the variance, in terms of the known quantities and the design measure, are obtained. Modest functional analytic techniques are then employed to characterize the optimal designs.

The paper is divided into four chapters. The second and third chapters contain general theorems on various aspects of the design problems associated respectively with finite dimensional and infinite dimensional parameter spaces. Examples are provided. The paper culminates in the contents of the fourth chapter. Here the general theory is applied successfully to random differential equations.

One much studied class of problems which we do not explicitly treat arises in estimation problems involving stochastic processes by asking for the optimal sampling "times" Sacks and Ylvisaker (1966), (1968), (1969); and Wahba (1971), (1974). Another area of investigation which is not treated is that of Hilbert space methods as expounded by Pazman (1978a). R. K. Mehra (1974) has investigated some design problems which closely resemble some of ours. Criteria are phrased in terms of the information matrix. He is able to treat non-linear problems in this way. We treat only problems linear in the unknown parameter.

Finally, I would like to point out that certain portions of this paper represent the fruits of collaboration with Professor W. J. Studden [Spruill and Studden (1978), (1979)] and thank especially Professors M. J. Christensen, S. G. Demko, and Heinz Engl for their helpful conversations.

2. Finite dimensional parameter space.

2.1. Introduction. Let X be a set of functions on [0,1]. Suppose that for each $x \in X$ an experimenter can observe the stochastic process $\{Y(x,t): t \in [0,1]\}$, where

$$Y(x,t) = \theta_0 x(t) + \theta_1 x^{(1)}(t) + N(t)$$

and N(t) is a zero mean noise process with covariance min(s,t). The constants θ_0 and θ_1 are unknown. If the experimenter wishes to obtain an estimate of, say, θ_1 based on observing N outcomes $\{Y(x_i,t): t \in [0,1]\}$, $i = 1,\cdots,N$, which $x_i(t)$ should he use? If it is desired to estimate the value of the mean function at some particular time t_0 in (0,1], what design minimizes the maximum over X of the variances of the estimators of this value?

Let X be the closed convex hull of the set of functions on [0,1] such that $x(0) = x^{(1)}(0) = 0$ and the second derivative is

$$x^{(2)}(t) = \begin{cases} \varepsilon & 0 \le t \le \alpha \\ -\varepsilon & < t \le 1 \end{cases}$$

for some $\alpha \in [0,1]$ and $\varepsilon = \pm 1$. Then the answer to the first question is to take all observations at

$$x(s) = \begin{cases} s^2/2 & 0 \le s \le 1/2 \\ 1/4-(s-1)^2/2 & 1/2 < s \le 1 \end{cases} .$$

The answer to the second, for $t_0 = 1/4$, is to take all observations at

$$x(s) = \begin{cases} s^2/2 & 0 \le s \le 1/4 \\ 1/16-(1/2-s)^2/2 & 1/4 < s \le 1 \end{cases} .$$

The answers to these and other questions of interest are obtained by utilizing the characterizations of the optimal designs developed below for the more general model

$$Y(x,t) = \sum_{j=0}^{k} \theta_j f_j(x)(t) + N(t),\ t \in T ,$$

where N has zero mean and known covariance kernel $K(s,t)$. It seems intuitively reasonable that a mean of the given form should serve as an adequate approximation to the true mean in many experimental situations. If T is finite this is just the usual linear model mean. In either case it can be taken as the Taylor approximation to a differentiable mean (see Dieudonné (1960, 8.14.3).

A characterization of the solution to the first problem, estimation of a linear form in the unknown θ's, is provided by an Elfving-type theorem. The second, minimax design, problem is not solved in general. The solution for the example shows that, as one might expect, the optimal designs in some minimax problems do not necessarily coincide with the D-optimal designs.

Although the estimators are based on the values of the process Y on the set T which may be an interval or larger set, the information on the optimal design is valuable for at least two reasons. If the experimenter is only able to sample values of Y at some finite set $T_m \subset T$ and this set "fills" T sufficiently (see Sacks and Ylvisaker (1966)), the experiment based on our optimal design will be nearly optimal. Our optimal design should be easier to find since it does not depend upon how T_m "sits" in T. The second reason is that in certain cases the optimal estimators may turn out to depend on the values of the process at only a finite set of points, an integral over the interval, or some other quantity which may be realized as the output of an analogue device. The optimal estimator of θ_1 for the example above turns out to be

$$\hat{\theta}_1 = \frac{1}{N} \sum_{i=1}^{N} [2\, Y(x_i, 1/2) - Y(x_i, 1)] .$$

2.2. <u>Preliminaries</u>. Let $\underset{\sim}{f} = (f_0, f_1, \ \ , f_k)$ be a vector of mappings from a set X onto a subset of functions on the set T. That is, for each $x \in X$, $f_j(x)(\cdot)$ is a real valued function on the set T with value $f_j(x)(t)$ at $t \in T$. The points $x \in X$ are possible levels of feasible experiments. For each level some experiment can be performed whose outcome is a stochastic process $\{Y(x,t)\colon t \in T\}$. It is assumed that the process has mean function

$$\sum_{j=0}^{k} \theta_j f_j(x)(t) \quad , \quad t \in T ,$$

and known proper covariance kernel

$$K(s,t) = \operatorname{cov}[Y(x,s),Y(x,t)], \ x \in X, \ s,t \in T .$$

The constants $\theta_0,\cdots,\theta_k$ are unknowns and the function $c'\theta$, where $c(\neq 0)$ is a fixed known vector, is to be estimated on the basis of N uncorrelated observations $\{Y(x_i,t): t \in T, \ i = 1,\cdots,N\}$.

An experimental design specifies a probability measure ξ concentrating mass $p_1,\cdots,p_r$ at the points $x_1,\cdots,x_r$, where $p_i N = n_i$, $i = 1,\cdots,r$ are integers. The associated experiment involves taking n_i observations of the stochastic process $\{Y(x_i,t): t \in T\}$.

The problem confronting the experimenter is to choose the design which minimizes the variance of the minimum variance linear unbiased estimator of $c'\theta$.

Let K be as above and H(K) be the associated reproducing kernel Hilbert space of functions on T with inner product $(\cdot,\cdot)_K$ (see Parzen (1959)). The assumptions are as follows.

(A1) The functions $\{f_j\}_{j=0}^{k}$, $f_j: X \to H(K)$ are continuous on the compact set X with the norm topology of H(K).

(A2) The set X can be given a topology so that one point sets are measurable. Arbitrary Borel probability measures will be admitted as possible designs. We denote the class of designs by Ξ.

(A3) There is a measure $\xi\epsilon\Xi$ such that

$$\int \|\sum_{j=0}^{k} a_j f_j(x)\|_K^2 d\xi(x) = 0.$$

if and only if $a_0 = a_1 = \cdots = a_k = 0$.

Define for each measure $\xi\epsilon\Xi$ the matrix $M(\xi)$ whose $ij^{\underline{th}}$ entry is

$$[M(\xi)]_{ij} = \int (f_i(x), f_j(x))_K d\xi(x) \quad . \tag{2}$$

That $M(\xi)$ is well defined follows from (A1).

Consider the discrete design ξ which places masses $p_i = \frac{n_i}{N}$ at x_i, $i = 1,\cdots,r$, where $x_i \epsilon X$ and $\{n_i\}_{i=1}^r$ are integers with $\sum_{i=1}^{r} p_i = 1$. The experiment consists of taking N uncorrelated observations $\{Y_j(x_i,t): t\epsilon T\}$, $j = 1,\cdots,n_i$, $i = 1,\cdots,r$. As we vary t, $Y(x,t)$ denotes a sample path observed at level x. Set $\Gamma = \{(x_1,1),(x_1,2),\cdots,(x_1,n_1),\cdots,(x_r,1),\cdots,(x_r,n_r)\} \times T$ and define the process $\{Z(\gamma): \gamma\epsilon\Gamma\}$ by $Z(\gamma) = Y_j(x_i,t)$ if $\gamma = (x_i,j,t)$.

The covariance kernel of Z is given by

$$B(\gamma_1,\gamma_2) = \begin{cases} K(t_1,t_2) & \text{if} \quad (x_{i_1},j_1) = (x_{i_2},j_2) \\ \\ 0 & \text{O.W.} \end{cases}$$

where $\gamma_a = ((x_a,j_a),t_a)$. Denote by $\langle Z,g\rangle_B$ the random variable in $L_2[Z(\gamma): \gamma\epsilon\Gamma]$ which is the image of g in the reproducing kernel

Hilbert space $H(B)$ associated with B. Note that $H(B)$ consists of functions defined on Γ. The class of linear estimators of $c'\theta$ is $\{\langle Z,g\rangle_B : g \in H(B)\}$. The function $c'\theta$ is said to be estimable with respect to this design if there is a $g \in H(B)$ such that $E_\theta \langle Z,g\rangle_B \equiv c'\theta$. Denote the inner product on $H(B)$ by $(\cdot,\cdot)_B$. Parzen (1959) proved the following.

Theorem 10A. Let $\{Z(\gamma): \gamma \in \Gamma\}$ have known proper covariance kernel B and unknown mean value function $m(\theta) \in H(B)$. Given that $c'\theta$ is estimable there is a unique linear estimator $\langle Z,g_0\rangle_B$ which is the uniformly minimum variance linear unbiased estimator of $c'\theta$ with variance $||g_0||_B^2$. Furthermore $\langle Z,g\rangle_B$ is umvlue of $c'\theta$ if and only if g is the unique function in the closure of $\{m(\theta): \theta \in \mathbb{R}^{k+1}\}$ satisfying $c'\theta \equiv (m(\theta),g)_B$, $\theta \in \mathbb{R}^{k+1}$.

The proof of the next result is obtained by a routine application of the properties of reproducing kernel Hilbert spaces as found in Parzen (1959) Section 5.

Lemma 2.2.1. The element g is in $H(B)$ if and only if $g((x_i,j),\cdot) \in H(K)$ for each (x_i,j), $j = 1,\cdots,n_i$, $i = 1,\cdots,r$. Furthermore, if g and h are in $H(B)$

$$(h,g)_B = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (h(x_i,j),g(x_i,j))_K . \tag{2.2}$$

Still retaining ξ as defined above we have the following.

Lemma 2.2.2. (i) $c'\theta$ is estimable if and only if c is in the range of $M(\xi)$

(ii) The variance of the umvlue of $c'\theta$ is

$$N^{-1}c'M^{+}(\xi)c \tag{2.3}$$

where M^{+} is the Moore-Penrose g-inverse of M ($= M^{-1}$ if M is non-singular).

Proof: Since $E_{\theta}\langle Z,g\rangle_B \equiv (m(\theta),g)_B \equiv (m(\theta),Pg)_B$ where Pg is the projection of g onto the (closed) subspace $\{m(\theta)\colon \theta\in\mathbb{R}^{k+1}\}$, and since $Pg(\gamma) = \sum_{j=0}^{k}\alpha_j f_j(\gamma)$ for some $\alpha\in\mathbb{R}^{k+1}$, one has $E_{\theta}\langle Z,g\rangle_B \equiv \theta'NM(\xi)\alpha$ by Lemma 2.1. This last expression is identically $c'\theta$ if and only if c is in the range of $M(\xi)$. Now (ii) follows from Parzen's theorem 10A above and Lemma 2.2.1 which show that the variance of the umvlue is

$$||Pg||_B^2 = N\alpha'M(\xi)\alpha \ , \tag{2.4}$$

where $NM(\xi)\alpha = c$. Utilizing the properties of $M^{+}(\xi)$ it can be seen that the rhs of (2.4) is just $N^{-1}c'M^{+}(\xi)c$.

For arbitrary c in $\mathbb{R}^{k+1} - \{0\}$ and ξ in Ξ define

$$d(c,\xi) = c'M^{+}(\xi)c \tag{2.5}$$

if $c\in\mathcal{R}[M(\xi)]$, the range of $M(\xi)$, and $d(c,\xi) = +\infty$ otherwise.

If $c'\theta$ is estimable with respect to the particular ξ which has been the object of our attention so far, the variance of the umvlue of $c'\theta$ is $N^{-1}d(c,\xi)$.

Following Karlin and Studden (1966) we make the following definitions for arbitrary ξ in Ξ.

<u>Definition</u>. The linear form $c'\theta$ is estimable with respect to ξ if c is in $\mathcal{R}[M(\xi)]$.

<u>Definition</u>. A design ξ_0 in Ξ is said to be optimal with respect to the estimation of $c'\theta$ if $d(c,\xi_0) = \min_{\Xi} d(c,\xi)$.

For any symmetric non-negative definite matrix M and any vectors a and b

$$|a'Mb| \leq \sqrt{a'Ma}\,\sqrt{b'Mb}\ ,$$

with equality if and only if $Ma = kMb$ for some constant k.

The alternative expression below for $d(c,\xi)$ was proved by Karlin and Studden (1966).

<u>Lemma 2.2.3</u>. For any $c \in R^k - \{0\}$ and ξ in Ξ such that $M(\xi) \neq 0$

$$d(c,\xi) = \sup_{d \in U} \frac{(c'd)^2}{d'M(\xi)d}\ , \tag{2.6}$$

where $U = \{d: d'Md > 0\}$.

<u>Proof</u>: Write M for $M(\xi)$. If $c \in R(M)$ then since MM^+ is the projection onto $R(M)$ (see Nashed (1971)) one has $c = MM^+c$. Thus

$$(c'd)^2 = (c'M^+Md)^2 \le c'M^+cd'MM^+Md$$
$$= c'M^+cd'Md$$

for any d. If $d \in U$ then

$$\frac{(c'd)^2}{d'Md} \le c'M^+c \ . \tag{2.7}$$

Setting $d = M^+c$, which is in U, equality is achieved in (2.7).

If $c \notin R(M)$ then $c = c_R + c_{R^\perp}$ is the direct sum decomposition of c and $||c_{R^\perp}||^2 > 0$. Taking $d_\varepsilon = c_{R^\perp} + \varepsilon a$, where $a \in R$, $a \neq 0$, is such that $(a, c_R) \ge 0$,

$$\frac{(c', d_\varepsilon)^2}{d_\varepsilon' M d_\varepsilon} = \frac{||c_{R^\perp}||^2 + \varepsilon(c_R, a)}{\varepsilon^2 a'Ma} \ .$$

Therefore $\sup\limits_{d \in U} \dfrac{(c'd)^2}{d'Md} = +\infty$.

For each x in X and λ in $\mathbb{R}^{k+1}$ write

$$L(x,\lambda) = \sum_{j=0}^{k} \lambda_j f_j(x) \ .$$

If g and h are in H(K) denote their inner product by $(g,h)_K$ or (g,h) when no confusion is possible. When $g_0, \cdots, g_k$ are in H(K) let $(h,\underset{\sim}{g})$ be the k+1 vector whose $j^{\underline{th}}$ component is (h,g_j). Let $\mathcal{R} = \overline{co}\{(\theta,\underset{\sim}{f}(x)): x\in X,\ ||\phi|| = 1\}$, where $\overline{co}$ (A) denotes the closed convex hull of A. For each $\lambda\in \mathbb{R}^{k+1} - \{0\}$ let

$$A(\lambda) = \{x\in X: ||L(x,\lambda)|| = \max_{y\in X} ||L(y,\lambda)||\}$$

and for any subsets $G(\lambda)$ of X

$$H_G(\lambda) = \left\{\frac{(L(x,\lambda),\underset{\sim}{f}(x))}{||L(x,\lambda)||} : x\in G(\lambda),\ ||L(x,\lambda)|| > 0\right\}.$$

Define the subset $\mathcal{R}_0$ of R^{k+1} by $\mathcal{R}_0 = \bigcup_{\lambda\in S^{k+1}} H_A(\lambda)$, where S^{k+1} is the unit sphere in $\mathbb{R}^{k+1}$. It can be shown that $\mathcal{R} = \overline{co}(\mathcal{R}_0)$. This relationship is useful in the theoretical work below. It is also potentially useful for a graphical solution since $\mathcal{R}_0 \subseteq \{(\phi,\underset{\sim}{f}(x)): x\in X,\ ||\phi|| = 1\}$.

<u>Lemma 2.2.4</u>. Under the assumptions A1 - A3 $\mathcal{R}_0$ is compact.

<u>Proof</u>: Let $\{r_j\}_{j=1}^{\infty} \subset \mathcal{R}_0$, $r_j = \left(\frac{L(x_j,\lambda_j)}{||L(x_j,\lambda_j)||},\ \underset{\sim}{f}(x_j)\right)$. Since S^{k+1} is compact there is a λ_0 and a subsequence $\lambda_{j'} \to \lambda_0$. Since $\underset{\sim}{f}$ is continuous $\underset{\sim}{f}(X)$ is a compact subset of the metric space $(H(K))^{k+1}$. Therefore, there is a further subsequence $\underset{\sim}{f}(x_{j''})$

converging to $\underset{\sim}{f}(x_0)$ componentwise for some x_0 in X. By the continuity of the functions involved it follows that

$$r_{j''} \to \left(\frac{L(x_0,\lambda_0)}{||L(x_0,\lambda_0)||}, f(x_0)\right) \text{ in } R_0 \text{ since also } ||L(x_0,\lambda_0)|| =$$

$\max_X ||L(x,\lambda_0)||$.

2.3. Characterization of optimal designs. Fix $c \in \mathbb{R}^{k+1} - \{0\}$ and let $v_0 = \inf_{\Xi} d(c,\xi)$.

Lemma 2.3.1. (a) $\beta c \in R$ implies $v_0 \leq \frac{1}{\beta^2}$.

(b) $\beta c \in \partial R$ implies $v_0 \geq \frac{1}{\beta^2}$.

Proof: (a) If $\beta = 0$ the result is immediate. Otherwise, since $\beta c \in R$, one has by Caratheodory's theorem that

$$\beta c = \sum_{j=0}^{k+1} \alpha_j(\phi_{0j}, \underset{\sim}{f}(x_{0j})) , \qquad (2.8)$$

where $\alpha_j > 0$, $\sum\alpha_j = 1$, and $(\phi_{0j}, \underset{\sim}{f}(x_{0j})) \in R_0$ for $j = 0,\cdots,k+1$. One has for any z in $\mathbb{R}^{k+1}$ that

$$(c'z)^2 = \frac{(\beta c'z)^2}{\beta^2} = \frac{[\sum\alpha_j(\phi_{0j}, L(x_{0j},z))]^2}{\beta^2}$$

$$\leq \frac{\sum\alpha_j ||L(x_{0j},z)||^2}{\beta^2} = \frac{\sum\alpha_j z'M(x_{0j})z}{\beta^2} = \frac{z'M(\xi)z}{\beta^2}$$

where ξ is the measure which places masses α_j at x_{0j} and $M(x_{0j})$ is the matrix for the measure placing all mass at x_{0j}. This shows $v_0 \le \frac{1}{\beta^2}$ by Lemma 2.2.3

(b) Since $\beta c \in \partial R$ there is a point $\lambda \in \mathbb{R}^{k+1}$ such that

$$\lambda'\beta c \ge \lambda' r \tag{2.9}$$

for all $r \in R$. This implies $\lambda'\beta c \ge |\lambda' r|$ for all $r \in R$ since R is symmetric about 0. Consider $r_0 = \left(\frac{L(x_0,\lambda)}{||L(x_0,\lambda)||}, \underset{\sim}{f}(x_0) \right)$ where

$$||L(x_0,\lambda)|| = \max_X ||L(x,\lambda)|| . \tag{2.10}$$

From (2.8) one has

$$\begin{aligned} \lambda'\beta c &= \sum \alpha_j (\phi_{0j}, L(x_{0j},\lambda)) \\ &= \frac{\sum \alpha_j a'_j M(x_{0j}) \lambda}{\sqrt{a'_j M(x_{0j}) a_j}} \end{aligned} \tag{2.11}$$

$$\le \sum \alpha_j \sqrt{\lambda' M(x_{0j}) \lambda} = \sum \alpha_j ||L(x_{0j}, \lambda)|| \le \max_X ||L(x,\lambda)|| .$$

From (2.9), (2.10), and (2.11) one has

$$\max ||L(x,\lambda)|| = \lambda' r_0 \le \lambda'\beta c \le \max ||L(x,\lambda)||$$

with $\lambda'\beta c < \max ||L(x,\lambda)||$ unless

$$||L(x_{0j},\lambda)|| = \max||L(x,\lambda)||$$

for $j = 0,\cdots,k+1$. Consequently

$$(\lambda'\beta c)^2 = \max||L(x,\lambda)||^2 = ||L(x_{0j},\lambda)||^2 = \lambda' M(\xi)\lambda$$

where ξ places masses α_j at x_{0j}, $j = 0,1,\cdots,k+1$. Since $v_0 = \inf_{\xi} d(c,\xi)$ there is a sequence ξ_n such that $d(c,\xi_n) \to v_0$. By A3 there is at least one design for which c is estimable. We may assume that $c \in R[M(\xi_n)]$ for all n. It follows that λ has a non-zero component in $N^{\perp}[M(\xi_n)]$ for every n, for otherwise one would have $\lambda \in N[M(\xi_n)] = R^{\perp}[M(\xi_n)]$ so that $(\lambda' c)^2 = 0$. This latter equality contradicts (A3) in view of $(\beta\lambda' c)^2 = \sup_x ||L(x,\lambda)||^2$. Therefore

$$\frac{1}{\beta^2} = \frac{(c'\lambda)^2}{\lambda' M(\xi)\lambda} \leq \frac{(c'\lambda)^2}{\lambda' M(\xi_n)\lambda} \leq d(c,\xi_n)$$

for all n and the lemma is proved.

<u>Theorem 2.3.1</u>. Under the assumptions A1 - A3 there is an optimal design ξ_0 for estimating $c'\theta$. Furthermore ξ_0 is optimal if and only if there is a function ϕ: $X \to H(K)$, $||\phi(x)|| \equiv 1$, such that

$$\int (\phi(x),\ \underset{\sim}{f}(x))\, d\xi_0(x) \qquad \text{is}$$

(i) proportional to c and

(ii) a boundary point of R.

Proof: Since R is compact there is a β, $0 \le \beta < \infty$, such that $\beta c \in \partial R$. From Lemma 2.3.1 $\beta^2 = \frac{1}{v_0}$ and since v_0 is finite by A3 β must be strictly positive. If ξ_0 denotes the measure which places masses α_j at x_{0j} as in (2.8) it follows that for any z in $\mathbb{R}^{k+1}$

$$(c'z)^2 = \frac{(\beta c'z)^2}{\beta^2} = v_0(\beta c'z)^2 \le v_0 \sum \alpha_j (\phi_{0j}, L(x_{0j}, z))^2 \qquad (2.12)$$
$$\le v_0 z' M(\xi_0) z \ .$$

Because v_0 is not zero $N^{\perp}[M(\xi_0)]$ is non-void. For $z'M(\xi_0)z > 0$ (2.12) yields

$$\frac{(c'z)^2}{z'M(\xi_0)z} \le v_0 \ .$$

Thus $d(c, \xi_0) = v_0$ and it has been demonstrated that ξ_0 is optimal.

Now suppose ξ_0 is optimal. Then $c'M^+(\xi_0)c = v_0$ and $c \in R[M(\xi_0)]$. Set $\lambda_0 = M^+(\xi_0)c$. Since $c \in R[M(\xi_0)]$ and MM^+ is the projection onto $R[M]$ one has $M(\xi_0)\lambda_0 = M(\xi_0)M^+(\xi_0)c = c$. Thus

$$\text{(a)} \quad M(\xi_0)\lambda_0 = c \qquad \text{and} \qquad (2.13)$$
$$\text{(b)} \quad \lambda'_0 M(\xi_0)\lambda_0 = v_0 \ .$$

Let $\{z : \lambda'(z - v_0^{-1/2}c) = 0\}$ be a supporting hyperplane to R at $v_0^{-1/2}c$. Since $v_0^{-1/2}c$ is a boundary point of R

$$v_0^{-1/2}c = \sum_{j=1}^{k+1} \gamma_j(\phi_{1j}, \underset{\sim}{f}(x_{1j})), \quad (\phi_{1j}, \underset{\sim}{f}(x_{1j})) \in R_0 \ . \qquad (2.14)$$

From the proof of Lemma 2.3.1 part (b) one finds

$$\sqrt{\lambda' M(\xi_1)\lambda} = \max||L(x,\lambda)|| = \lambda' v_0^{-1/2}c \ ,$$

where ξ_1 places masses $\gamma_j > 0$ at x_{1j} , $j = 1,\cdots,k+1$. Furthermore, on the support of ξ_1 , $S(\xi_1)$,

$$\lambda'(\phi_{1j}, \underset{\sim}{f}(x_{1j})) = \max||L(x,\lambda)|| \ . \qquad (2.15)$$

Since $M(\xi_0)\lambda_0 = c$ one has from (2.14) and (2.15) that $\lambda' M(\xi_0)\lambda_0 v_0^{-1/2} = \max||L(x,\lambda)||$. Thus

$$\begin{aligned} \max||L(x,\lambda)||^2 &= v_0^{-1}(\lambda' M(\xi_0)\lambda_0)^2 \\ &\le v_0^{-1}\lambda' M(\xi_0)\lambda\lambda'_0 M(\xi_0)\lambda_0 \ . \end{aligned}$$

Using (2.13b) this shows that $\max||L(x,\lambda)||^2 \le \lambda' M(\xi_0)\lambda$ with strict inequality unless $M(\xi_0)\lambda_0 = kM(\xi_0)\lambda$. Since the opposite inequality always holds $||L(x,\lambda)|| = \max||L(x,\lambda)||$ a.e. ξ_0 . Now set

$$\phi(x) = \begin{cases} \dfrac{L(x,\lambda)}{||L(x,\lambda)||} & x \in S(\xi_0) \cap D \\ \phi_0 & \text{o.w.} \end{cases}$$

where $D = \{x: ||L(x,\lambda)|| = \max||L(x,\lambda)||\}$ and $||\phi_0|| = 1$. According to (2.13a)

$$\int(\phi(x),f(x))d\xi_0(x) = \frac{M(\xi_0)\lambda}{\max||L(x,\lambda)||} = \frac{M(\xi_0)\lambda_0}{k\max||L(x,\lambda)||}$$
$$= \frac{c}{k\max||L(x,\lambda)||}\ .$$

It follows from

$$v_0 = \lambda'_0 M(\xi_0)\lambda_0 = k\ \lambda'_0 M(\xi_0)\lambda = k\ \sqrt{\lambda'_0 M(\xi_0)\lambda_0}\ \sqrt{\lambda' M(\xi_0)\lambda}$$

that $v_0^{1/2} = k\max||L(x,\lambda)||$. This shows the necessity of the conditions (i) and (ii).

Next suppose (i) and (ii) hold. Then for any z in $\mathbb{R}^{k+1}$

$$\begin{aligned} v_0^{-1}(c'z)^2 &= (\int(\phi(x),L(x,z))d\xi_0(x))^2 \\ &\le \int(\phi(x),L(x,z))^2 d\xi_0(x) \\ &\le \int||L(x,z)||^2 d\xi_0(x) = z'M(\xi_0)z \end{aligned}$$

showing ξ_0 to be optimal.

In certain cases it is possible to have an optimal design concentrating all mass on a single point of X. In these cases the optimal approximate theory design coincides with the optimal exact theory design. The points c at which this phenomenon occurs is exactly the set $\{\alpha\mathcal{R}_0: \alpha > 0\}$. If $\mathcal{R}_0$ has no "holes" this set is $\mathbb{R}^{k+1} - \{0\}$. A useful result in this connection is given by the following.

Theorem 2.3.2. There is an optimal one point design for each $c \in \mathbb{R}^{k+1} - \{0\}$ if and only if $H_A(\lambda)$ is convex for each $\lambda \in S^{k+1}$.

Proof: Assume that there is a one point optimal design for each c but that $H(\lambda_0)$ is not convex for some $\lambda_0 \in S^{k+1}$. Let $z \in \overline{co}H(\lambda_0) - H(\lambda_0)$. Then since $z \in \partial R$,

$$z = \left(\frac{L(x,\lambda)}{||L(x,\lambda)||} \;,\; \underset{\sim}{f}(x) \right) \tag{2.16}$$

for some x and λ. However, since $z \in \overline{co}H(\lambda_0)$

$$z = \sum \alpha_j \left(\frac{L(x_j,\lambda_0)}{||L(x_j,\lambda_0)||} , \underset{\sim}{f}(x_j) \right)$$

where $\lambda_0' z = \max||L(x,\lambda_0)||$. By (2.16)

$$\lambda_0' z = \frac{\lambda_0' M(x)\lambda}{\sqrt{\lambda' M(x)\lambda}} \le \sqrt{\lambda_0' M(x)\lambda_0} = ||L(x,\lambda_0)|| \tag{2.17}$$

Thus $x \in A(\lambda_0)$. This implies equality in (2.17) which shows

$$M(x)\lambda = k\, M(x)\lambda_0 \; . \tag{2.18}$$

From (2.18) it follows that $||L(x,\lambda)|| = k\; ||L(x,\lambda_0)||$ and

$$z = \frac{M(x)\lambda}{||L(x,\lambda)||} = \frac{k\ M(x)\lambda_0}{k||L(x,\lambda_0)||} = \left(\frac{L(x,\lambda_0)}{||L(x,\lambda_0)||}\ ,\ \underset{\sim}{f}(x)\right) \in H(\lambda_0)\ .$$

This contradiction establishes the necessity of the convexity of $H(\lambda)$.

Let $H(\lambda)$ be convex for each λ let c be given. If $\{z\colon \lambda'(z-v^{-1/2}c) = 0\}$ is a supporting hyperplane to $\mathcal{R}$ at $v^{-1/2}c$

$$v^{-1/2}c = \sum_{j=1}^{k+1} \alpha_j \left(\frac{L(x_j,\lambda)}{||L(x_j,\lambda)||}\ ,\ \underset{\sim}{f}(x_j)\right)\ ,$$

where $||L(x_j,\lambda)|| = \max||L(x,\lambda)||$ (see the proof of Lemma 2.3.1). Thus $v^{-1/2}c \in \overline{\text{co}}H(\lambda) = H(\lambda)$, or equivalently, there is an $x \in A(\lambda)$ such that

$$v^{-1/2}c = \left(\frac{L(x,\lambda)}{||L(x,\lambda)||}\ ,\ \underset{\sim}{f}(x)\right)\ .$$

2.4. <u>Linear</u> f_j. In the linear case certain simplifications are possible if (A4) holds.

(A4) The set X is a compact convex subset in a locally convex topological vector space.

To motivate these results consider the example alluded to in the introduction.

Example 2.4.1. Let

$$Y(x,t) = \theta_0 x(t) + \theta_1 x^{(1)}(t) + N(t),\ t \epsilon [0,1]$$

be observable, where N(t) is a zero mean noise process with K(s,t) = min (s,t) and x is an element of the closed convex hull of the perfect splines on [0,1] of degree 2 with one knot satisfying

$$x(0) = x^{(1)}(0) = 0\ . \tag{2.19}$$

That is, X is the closed convex hull of the set of functions on [0,1] satisfying (19) and

$$x^{(2)}(t) = \begin{cases} \varepsilon & 0 \le t \le \alpha \\ -\varepsilon & \alpha < t \le 1 \end{cases} \tag{2.20}$$

for some $\alpha \epsilon [0,1]$, $\varepsilon = \pm 1$. The problem is to find the optimal design for estimating $c'\theta$.

To verify assumption (A1) begin by noting that the reproducing kernel Hilbert space H(K) consists of all functions of the form

$$F(t) = \int_0^t f(s)ds$$

for f in $L_2[0,1]$ with inner product

$$(F,G)_K = \int_0^1 f(s)g(s)ds\ ,$$

(see Kuelbs (1970).

Let B be the set of all functions h on [0,1] such that

$$\int_0^1 [|h(t)|^2 + |h^{(1)}(t)|^2 + |h^{(2)}(t)|^2]dt < \infty . \qquad (2.21)$$

Define $||h||^2$ by (2.21). The operators f_j defined on B by $f_j(x) = x^{(j)}$ are continuous as functions from B to H(K). The set of functions satisfying (2.19) and (2.20) is compact as a subset of B so by Mazur's theorem (Dunford and Schwarz (1958)) X is also compact.

Assumption (A2) is immediate. Computation of $\det[M(x_\alpha)]$ shows any design concentrating mass at any x_α satisfies (A3) since $M(x_\alpha)$ are all non-singular.

We defer verification of (A4) and the solution of the problem until presentation of the results of this section. Since the set X is so large it would conceivably be a difficult or impossible task to plot R_0 directly. This is true even if one observes from the form of R_0 that it suffices to consider only those points x in the boundary of X.

It is the purpose of this section to prove the more useful result, at least for the example, that only points x in E(X), the extreme points of X, need be considered. If

$$R_1 = \bigcup_{\lambda \in S^{k+1}} \bigcup_{x \in B(\lambda)} \left\{ \left(\frac{L(x,\lambda)}{||L(x,\lambda)||} , \underset{\sim}{f}(x) \right) \right\} ,$$

where $B(\lambda) = A(\lambda) \cap E(X)$ we shall show that $R = \overline{co}(R_1)$.

One can prove the following.

<u>Theorem 2.4.1</u>. If ψ is a continuous convex functional on K, a compact convex subset of a locally convex topological vector space, then ψ achieves its maximum at an extreme point of K.

If X satisfies (A4) then by Theorem 2.4.1 the sets $B(\lambda)$ are never empty for $\lambda \in S^{k+1}$. This follows from the convexity of the functions $||L(\cdot,\lambda)||$ on X.

<u>Theorem 2.4.2</u>. Under the assumptions A1 - A4 $R = \overline{co}(R_1)$.

<u>Proof</u>: If suffices to prove $R_0 \subset \overline{co}(R_1)$. Suppose R_0 is not contained in $\overline{co}(R_1)$. Then there is a point

$$r_0 = \left(\frac{L(x_0,\lambda_0)}{||L(x_0,\lambda_0)||} , \underset{\sim}{f}(x_0) \right)$$

in R_0, a number a, and a vector γ such that for all $r \in R_1$

$$\gamma' r < a < \gamma' r_0 . \tag{2.22}$$

This follows from the compactness of $\overline{co}(R_1)$. Let

$$r = \left(\frac{L(x,\gamma)}{||L(x,\gamma)||} , \underset{\sim}{f}(x) \right)$$

be in R_1. Then from (2.22)

$$||L(x,\gamma)|| < a < \frac{\gamma' M(x_0)\lambda_0}{\sqrt{\lambda_0' M(x_0)\ {}_0}}$$

$$\leq ||L(x_0,\gamma)|| .$$

This impossible since $||L(x,\gamma)|| = \max_X ||L(x,\gamma)||$.

Corollary 2.4.1. Under A1 - A4 for each $c \in \mathbb{R}^{k+1} - \{0\}$ there is an optimal design on no more than k+1 points of E(X).

Corollary 2.4.2. If A1 - A4 hold, then there is an optimal one point design for each $c \in \mathbb{R}^{k+1} - \{0\}$ if and only if $H_B(\lambda)$ is convex for each $\lambda \in S^{k+1}$.

Example 2.4.1. (Solution). Assumption (A4) is satisfied since every Hilbert space is locally convex and B is a Hilbert space. By Lemma V.8.5 of Dunford and Schwarz (1958) every extreme point of X must satisfy (2.19) and (2.20). It is therefore possible to obtain R_1 by maximizing $\lambda' M(x_\alpha)\lambda$ over $\alpha \in [0,1]$ for each $\lambda \in S^{k+1}$ and plotting the points $\dfrac{M(x_\alpha)\lambda}{\sqrt{\lambda' M(x_\alpha)\lambda}}$. However, for this particular example it is easier to solve the problem by noting that $H_B(\lambda)$ are are always convex. If

$$\lambda = \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \text{ and } \lambda_0 \neq 0,\ \lambda' M(x_\alpha)\lambda$$

is maximized by x_α or $-x_\alpha$ for a unique $\alpha \in (0,1]$. Hence $H_B(\lambda)$ is a one-point set for these λ. If $\lambda = \begin{bmatrix} 0 \\ \lambda_1 \end{bmatrix}$ then $\lambda' M(x_\alpha)\lambda$ is constant for $\alpha \in [0,1]$. In this case

$$H_B(\lambda) = \left\{ \frac{\lambda_1}{|\lambda_1|} \begin{bmatrix} 1/2(1-2\alpha)^2 \\ 1 \end{bmatrix} : \alpha \in [0,1] \right\}$$

is also convex. Therefore for each $c \in \mathbb{R}^{k+1} - \{0\}$ the design which concentrates all mass at x_α minimizing $c' M^{-1}(x_\alpha) c$ over $\alpha \in (0,1]$ is optimal for the estimation of $c'\theta$. Using the methods of Parzen it can be shown that if the optimal design concentrates all mass at α_0 in $[0,1]$ then

$$\hat{c'\theta} = u_0 \sum_i \{ Y_i(1)[2\alpha_0 - 1] - \int_0^{\alpha_0} Y_i(s)ds + \int_{\alpha_0}^{1} Y_i(s)ds \}$$

$$+ u_1 \sum_i \{ 2Y_i(\alpha_0) - Y_i(1) \}$$

where u is any solution to

$$N\, M(x_{\alpha_0}) u = c .$$

<u>Remark</u>: The linearity of the f_j combined with the convexity of X does not always entail one point optimal designs for each $c \in \mathbb{R}^{k+1} - \{0\}$. For example, if in the example above one takes X

to be the convex hull of the two functions which satisfy (2.19) and (2.20) for $\alpha = \frac{1}{2}$ and $\alpha = 1$, then the design for estimating θ_0 concentrates mass on both of these.

However there is no simple relationship between the size of X and the existence of one point optimal designs for all of S^{k+1}. Taking X to be the convex hull of the two functions satisfying (2.19) and (2.20) for $\alpha = 1/4$ and $\alpha = 3/4$ one finds that the optimal design for each c in S^{k+1} places all mass at the function corresponding to 3/4.

The solution to the problem posed in the next example may be obtained without all the machinery developed above. It provides, however, a convenient solution.

The problem has certain features in common with some problems discussed in Joiner and Campbell (1976) and Kiefer (1961). It is the problem of optimal estimation of $\theta_1 - \theta_2$ (see below) in the presence of a time trend whose form is known up to a multiplicative constant.

It differs from those discussed by Joiner and Campbell (1976) and from most of those discussed by Keifer (1961) in the criterion of optimality. It should also be pointed out that our model above does not cover the William's models discussed in Kiefer since in that problem the covariance depends on the unknown parameter.

Example 2.4.2. Let

$$Y(x,t) = \theta_0+\theta_1 x_1(t)+\cdots+\theta_k x_k(t)+\theta_{k+1}\phi(t)+\varepsilon(t) ,$$

where $x_j(t)$ $j = 1,\cdots,k$ is zero or one for each $t \in T = \{1,\cdots,M\}$, $E[\varepsilon(t)] \equiv 0$, and $E[\varepsilon(s)\varepsilon(t)] = \delta_{st}$. The function $\phi(t)$ is assumed to be known.

One may visualize $\{Y(\underset{\sim}{x},t)\}_{t=1}^{M}$ as one day's observations taken under conditions which change with time as reflected by $\phi(t)$. The function $\underset{\sim}{x}$ specifies the level of each factor to be run at each time $t \in T$. The next day's observations with factor levels $\underset{\sim}{x}^1$, are $\{Y(\underset{\sim}{x}^1,t)_{t=1}^{M}\}$. An optimal experimental design specifies the types of daily experiments $\underset{\sim}{x}$ which should be run and the proportion of days on which they should be run. We shall assume zero cost for all factor-level changes.

Let $\phi(t_1) \leq \phi(t_2) \leq \cdots \leq \phi(t_M)$ and assume, for simplicity's sake, that M is even. Let $L = \{t_1,t_2,\cdots,t_{M/2}\}$ and $U = \{t_{M/2+1},\cdots,t_M\}$. Since $2\sum_{t\in U} \phi(t) \geq \sum_{t=1}^{M} \phi(t)$ and the opposite inequality holds when U is replaced by L, there is an $\alpha_0 \in [0,1]$ such that

$$\alpha_0 \left[\sum_{t\in U} \phi(t) - \frac{\sum\phi(t)}{2} \right] + (1-\alpha_0) \left[\sum_{t\in L} \phi(t) - \frac{\sum\phi(t)}{2} \right] = 0 .$$

Let $\underset{\sim}{x}_i(t) = (x_{i1}(t),\cdots,x_{ik}(t))$, $i = 1,2$, be such that

$$x_{11}(t) = \begin{cases} 1 & \text{for } t\in U \\ 0 & \text{for } t\in L \end{cases}$$

and

$x_{12}(t) = 1 - x_{11}(t)$ for $t \in \{1,\cdots,M\}$. Set $x_{ij}(t) \equiv 0$ for $i = 1,2$, $j \geq 3$, $x_{21}(t) = 1 - x_{11}(t)$, and $x_{22}(t) = 1 - x_{21}(t)$. It will be shown that the design which places mass α_0 at $\underset{\sim}{x}_1$ and $1 - \alpha_0$ at $\underset{\sim}{x}_2$ is optimal for estimating $\theta_1 - \theta_2$ and $v_0 = \frac{4}{M}$.

It is not hard to verify that, if F is the space of all functions g from $\{1,\cdots,M\} = T$, to $\mathbb{R}^{k+2}$ with $||g|| = \sum_{j=0}^{k+1} ||g_j||$, then $X = \{1,x_1,\cdots,x_k,\phi)\}$ is a compact convex subset thereof. If $f_j(g) = g_j$ then all assumptions are satisfied since $H(K)$ consists of all real-valued functions h on T with inner product $(h_1,h_2) = \sum_{t=1}^{M} h_1(t)h_2(t)$. Therefore, see Studden and Tsay (1976), if one can find $\alpha_1,\cdots,\alpha_r$, $\underset{\sim}{x}_1,\cdots,\underset{\sim}{x}_r$ and $\lambda_0 \in \mathbb{R}^{k+2}$ such that

(i) $$\max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda}_0)|| = \min_{\lambda'c=1} \max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda})||,$$

where $c' = (0,1,-1,0,\cdots,0)$, and

(ii) $$\sum_{j=1}^{r} \alpha_j(\underset{\sim}{f}(\underset{\sim}{x}_j),L(\underset{\sim}{x}_j,\underset{\sim}{\lambda}_0)) = \max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda}_0)||^2 c,$$

then $\alpha_1,\cdots,\alpha_r,\underset{\sim}{x}_1,\cdots,\underset{\sim}{x}_r$ is an optimal design and

$$v_0 = \frac{1}{\max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda}_0)||^2}.$$

Consider solving (i). One has $||L(x,\underset{\sim}{\lambda})||^2 = \sum_{t=1}^{M} (\lambda_0+\lambda_1 x_1(t)+ \cdots+\lambda_k x_k(t)+\lambda_{k+1}\phi(t))^2$.

Set $h_\lambda(\underset{\sim}{\varepsilon}) = \lambda_0+\lambda_1\varepsilon_1+\cdots+\lambda_k\varepsilon_k$, where $\varepsilon_j = 0$ or 1, $j = 1,\cdots,k$, $b_\lambda = \max_{\underset{\sim}{\varepsilon}} h_\lambda(\underset{\sim}{\varepsilon})$ and $a_\lambda = \min_{\underset{\sim}{\varepsilon}} h_\lambda(\underset{\sim}{\varepsilon})$. Then since

$$(b_\lambda+\lambda_{k+1}\phi(t))^2 > (a_\lambda+\lambda_{k+1}\phi(t))^2$$

if and only if $\lambda_{k+1}\phi(t) > -\dfrac{(a_\lambda+b_\lambda)}{2}$, and one has

$$\max_X ||L(x,\lambda)||^2 = \sum_{t\epsilon W_\lambda} (b_\lambda+\lambda_{k+1}\phi(t))^2 + \sum_{t\epsilon V_\lambda} (a_\lambda+\lambda_{k+1}\phi(t))^2 ,$$

where $W_\lambda = \left\{t\colon t\epsilon T,\ \lambda_{k+1}\phi(t) > -\dfrac{(a_\lambda+b_\lambda)}{2}\right\}$ and $V_\lambda = T - U_\lambda$. Now write

$$b_\lambda + \lambda_{k+1}\phi(t) = \lambda_{k+1}\phi(t) + \frac{a_\lambda + b_\lambda}{2} + \frac{b_\lambda - a_\lambda}{2}$$

and $$a_\lambda + \lambda_{k+1}\phi(t) = \lambda_{k+1}\varepsilon(t) + \frac{a_\lambda + b_\lambda}{2} + \frac{a_\lambda - b_\lambda}{2}$$

to see that on W_λ, $(b_\lambda+\lambda_{k+1}\phi(t))^2 \geq \left(\dfrac{b_\lambda-a_\lambda}{2}\right)^2$ and on V_λ, $(a_\lambda+\lambda_{k+1}\phi(t))^2 \geq \left(\dfrac{b_\lambda-a_\lambda}{2}\right)^2$. Thus $\max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda})||^2 \geq M\dfrac{(b_\lambda-a_\lambda)^2}{4}$.

Now among $\underset{\sim}{\lambda}'c = 1$, that is, $\underset{\sim}{\lambda}$ for which $\lambda_1 - \lambda_2 = 1$, $h_\lambda(\underset{\sim}{\varepsilon}) = \lambda_0+(1+\lambda_2)\varepsilon_1+\lambda_2\varepsilon_2+\cdots+\lambda_k\varepsilon_k$ so $b_\lambda - a_\lambda = (1+\lambda_2)_+ + (\lambda_2)_+ - ((1+\lambda_2)_- + (\lambda_2)_-) \geq 1$. Therefore, $\max_X ||L(\underset{\sim}{x},\underset{\sim}{\lambda})||^2 \geq \frac{M}{4}$. Setting

$\underset{\sim}{\lambda}_0' = (-\frac{1}{2},1,0,\cdots,0)$, one finds $\max_{X} ||L(x,\underset{\sim}{\lambda}_0)||^2 = \frac{M}{4}$ is achieved whenever any $\underset{\sim}{x}_1$ is used.

Using the particular $\underset{\sim}{x}_1, \underset{\sim}{x}_2$ and α_0 given above one finds

$$\alpha_0 \left(\begin{bmatrix} 1 \\ x_{11}(t) \\ x_{12}(t) \\ \vdots \\ x_{1k}(t) \\ \phi(t) \end{bmatrix}, \left(-\frac{1}{2} + x_{11}(t)\right) \right) + (1-\alpha_0) \left(\begin{bmatrix} 1 \\ x_{21}(t) \\ x_{22}(t) \\ \\ x_{2k}(t) \\ \phi(t) \end{bmatrix}, -\frac{1}{2} + x_{21}(t) \right)$$

$$= \alpha_0 \begin{bmatrix} -\frac{M}{2} + \frac{M}{2} \\ -\frac{M}{4} + \frac{M}{2} \\ -M/4 \\ 0 \\ \vdots \\ 0 \\ -\frac{\sum\phi(t)}{2} + \sum_{U}\phi(t) \end{bmatrix} + (1-\alpha_0) \begin{bmatrix} -\frac{M}{2} + \frac{M}{2} \\ -\frac{M}{4} + \frac{M}{2} \\ -M/4 \\ 0 \\ \vdots \\ 0 \\ -\frac{\sum\phi(t)}{2} + \sum_{L}\phi(t) \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ M/4 \\ -M/4 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \frac{M}{4}\underset{\sim}{c} .$$

2.5. Algorithms. In general it may be difficult to obtain an optimal design analytically. Algorithms for finding optimal designs of various descriptions are given in Atwood, C. L. (1976), Fedorov (1972), Gribik and Kortanek (1977), Pazman (1978b), Studden and Tsay (1976).

We describe below a somewhat different approach which exploits the geometry of the set $\mathcal{R}$ and the characterization provided by Elfving's theorem. The algorithm given is exact (i.e.; not yet developed to be useful when numerial errors are present) and is given in terms of the more general goal of obtaining the puncture point of a convex set K. At each stage a new normal for a supporting hyperplane to K is generated. The successive hyperplanes have the property that the point at which they intersect the ray puncturing K converges to the point at which the ray punctures K. The general idea is that if one knows the normal of the supporting hyperplane to K and the point of puncture, the optimal design can be found.

Let K be a closed bounded subset of the finite dimensional Hilbert space H. Let $c \neq 0$ be a fixed point in H and define

$$\beta_* = \inf_{(\phi,c)>0} \sup_{x \in K} \frac{(\phi,x)}{(\phi,c)} ,$$

where $(\cdot,\cdot)$ is the inner product on H. If the ray $\{\beta c: \beta \in R\}$ intersects the closed convex hull of K, then

$$\beta_* = \max\{\beta: \beta c \in \overline{co}(K)\} .$$

Otherwise, $\beta_* = -\infty$.

Some iterative techniques for finding β_* are developed below. They are based on the assumption that for each ϕ in H the subset of points in K which maximize $(\phi,\cdot)$ is relatively easily found.

It can be assumed, without loss of generality, that $K = \overline{co}(K)$ since β_* is not affected. This assumption is made throughout. Other specific assumptions are stated as needed.

For each ϕ in H let $\beta(\phi)$ be that value of β for which βc lies on the supporting hyperplane to K with normal ϕ. If $(\phi,c) > 0$

$$\beta(\phi) = \max_{x \in K} \frac{(\phi,x)}{(\phi,c)} .$$

Otherwise let $\beta(\phi) = -\infty$. Let $M(\phi) = \{x \in K: (\phi,x) = \max_{y \in K} (\phi,y)\}$.

When a closed subspace is denoted by a script capital the orthogonal projection onto that subspace is denoted by the corresponding Roman capital. Set

$$A_k = \{\alpha: \sum_{j=0}^{k} \alpha_j = 1, \alpha_j \geq 0\} \subset R^{k+1} .$$

For any set $B \subset H$ denote the extreme points of B by E(B). Denote by I(c) the ray $\{\alpha c: \alpha \in R\}$ and by int(A) the interior of any subset A.

Procedure I

For this procedure the following assumption is made.

Assumption: The set K is a closed, bounded, convex subset of the real finite dimensional Hilbert space H and $\text{int}(D) \cap I(c) \neq \phi$.

Let $\phi_0 = c$. If $x_0 \epsilon M(\phi_0)$ then clearly
$\beta_1 = \min_{\underset{\sim}{\alpha} \epsilon A_1} \beta(\alpha_0\phi_0 + \alpha_1(\beta_0 c - x_0)) \leq \beta_0$ and $(\phi_0, \beta_0 c - x_0) = 0$. If

$P_0 = \mathrm{span}\{\phi_0, \beta_0 c - x_0\}$ then $P_0\beta_1 c$ is a boundary point of $P_0 K$. Therefore, there is an $x_1 \epsilon K$ (if dim H > 2) such that $P_0\beta_1 c = P_0 x_1$.
Since

$$(\beta_1 c - x_1, \beta_0 c - x_0) = (\beta_1 c - x_1, P_0(\beta_0 c - x_0))$$

$$= (P_0(\beta_1 c - x_1), \beta_0 c - x_0) = 0$$

and similarly, $(\phi_0, \beta_1 c - x_1) = 0$, the set $\{\phi_0, \beta_0 c - x_0, \beta_1 c - x_1\}$ is an orthogonal set. If dim H > 3 then since

$\beta_2 = \min_{\underset{\sim}{\alpha} \epsilon A_3} \beta(\alpha_0\phi_0 + \sum_{i=1}^{2} \alpha_i(\beta_{i-1} c - x_{i-1})) \leq \beta_1$, x_2 can be found in

K such that $\{\phi_0, \beta_0 c - x_0, \cdots, \beta_2 c - x_2\}$ is an orthogonal set. Continuing in this way a sequence $\{\beta_j\}_{j=0}^{\infty}$ is generated which satisfies
$\beta_0 \geq \beta_1 \geq \cdots \geq \beta_*$.

<u>Theorem 2.5.1</u>. If dim H = n, then $\beta_{n-1} = \beta_*$.

<u>Proof</u>: Since $\beta_* > -\infty$ there is a $\psi \epsilon H$ such that $\beta(\psi) = \beta_*$.

If dim H = n then

$$\psi = a_0\phi_0 + \sum_{j=1}^{n-1} a_j(\beta_{j-1} c - x_{j-1}) .$$

Since for any positive scalar t one has $\beta(t\psi) = \beta(\psi)$ the result will follow if it can be shown that all a_j's are non-negative. Clearly $a_0 > 0$ since $(\psi,c) = a_0||c||^2$. For $1 \le k \le n - 1$, $(\beta_{k-1}c-x_{k-1},\psi) = a_k||\beta_{k-1}c-x_{k-1}||^2$. If $\beta_{k-1}c-x_{k-1} = 0$ then $\beta_{k-1} = \beta_*$. Otherwise, one has $\beta_{k-1}(c,\psi) < (\psi,x_{k-1})$. But this implies $\beta_{k-1} < \beta_*$ which is impossible.

Procedure II

Throughout this section the following assumption is made. Assumption: There is a compact set K_0 such that $K = \overline{co}(K_0)$ and $int(K) \cap I(c) \neq \phi$.

The motivation for the following modification of Procedure I is that all computations can be performed in K_0 rather than K.

Suppose $\phi_0,\cdots,\phi_k,\beta_0,\cdots,\beta_k,x_0,\cdots,x_k$

satisfy (i) $$\beta_j = \min_{\underset{\sim}{\alpha}\in A_j} \beta(\alpha_0\phi_0+\cdots+\alpha_j(\beta_{j-1}c-x_{j-1})) = \beta(\alpha_0^{(j)}\phi_0+\cdots+\alpha_j^{(j)}(\beta_{j-1}c-x_{j-1}))$$

(ii) $\phi_0 = c,\ \phi_j = \alpha_0^{(j)}\phi_0+\cdots+\alpha_j^{(j)}(\beta_{j-1}c-x_{j-1})$ for $j = 1,2,\cdots,k$

and (iii) $x_j \in M(\phi_j)$ for $j = 0,\cdots,k$.

Define $\phi_{k+1} = \alpha_0^{(k+1)}\phi_0+\cdots+\alpha_{k+1}^{(k+1)}(\beta_k c-x_k)$ and $\beta_{k+1} = \beta(\phi_{k+1})$.

Lemma 2.5.1. If (i) - (iii) hold and $\{\phi_0, \beta_0 c - x_0, \cdots, \beta_k c - x_k\}$ is a linearly independent set with $\beta_{k+1} > \beta_*$, then

(a) dim H > k+2 and

(b) there is an $x \in M(\phi_{k+1}) \cap K_0$

such that $\{\phi_0, \cdots, \beta_{k+1} c - x_{k+1}\}$ is a linearly independent set.

Proof: The first conclusion follows from the fact that if dim H = k+2 then the given set spans H and the coefficients for the normal to achieve β_* would all be non-negative.

To obtain the second, note that $M(\phi_{k+1})$ is a compact convex supporting set for K. It follows that $E[M(\phi_{k+1})] \subseteq E(K)$. So if

$$\beta_{k+1} c - M(\phi_{k+1}) \cap E(K) \subseteq P \ , \tag{2.23}$$

where $P = \text{span}\{\phi_0, \beta_0 c - x_0, \cdots, \beta_k c - x_k\}$, then

$$\beta_{k+1} c \ - \ E[M(\phi_{k+1})] \subseteq P \ . \tag{2.24}$$

Since $P(\beta_{k+1} c)$ is a boundary point of PK one has $z \in M(\phi_{k+1})$ such that $P(\beta_{k+1} c - z) = 0$. This, with (2.24) shows $\beta_{k+1} c - z = 0$ contradicting $\beta_{k+1} > \beta_*$. The second claim now follows from $E(K) \subset K_0$ and the fact that (2.23) cannot hold. □

Let $\phi_0 = c$ and $\beta_0 = \max\limits_{x \in K_0} \dfrac{(\phi_0, x)}{(\phi_0, c)} = \dfrac{(\phi_0, x_0)}{(\phi_0, c)}$. If dim H > 2 then by Lemma 2.5.1 an $x_1 \in K_0 \cap M(\phi_1)$ can be found for which

$\{\phi_0, \beta_0 c - x_0, \beta_1 c - x_1\}$ is a linearly independent set. One can continue this process until a spanning set $\{\phi_0, \beta_0 c - x_0, \cdots, \beta_{n-2} c - x_{n-2}\}$ has been found for H, with dim H = n.

Theorem 2.5.2. The sequence $\{\beta_j\}_{j=0}^{n-1}$ generated above satisfies $\beta_j \geq \beta_{j+1} \geq \beta_*$ for all $j = 0, \cdots, n-2$ and $\beta_{n-1} = \beta_*$.

Procedure III

Throughout this section the following assumption is made.
Assumption: The set K is a compact, strictly convex set and $\mathrm{int}(K) \cap I(c) \neq \phi$.

A strictly convex set C has the property that if $c_1, c_2 \in C$ then $\alpha c_1 + (1-\alpha) c_2$ is in the interior of C for all $\alpha \in (0,1)$.

The motivation for this modification is that if K is strictly convex then all minimizations can be carried out over A_1. That is, at any stage the next normal can be found by a one-dimensional minimization on the interval [0,1]. The next lemma is more general than needed. The set K need only be convex and H need not be finite dimensional.

Lemma 2.5.2. If $\phi \in H$ is such that $(\phi, c) > 0$, $\beta = \beta(\phi) > \beta_*$, $x_0 \in M(\phi)$, and

$$||\beta c - x_0|| = \min_{y \in M(\phi)} ||\beta c - y||$$

then there is an $\varepsilon > 0$ such that

$$\beta(b\phi+(1-b)(\beta c-x_0)) < \beta \text{ for } b\in(1-\varepsilon,1) . \tag{2.25}$$

Proof: Let ϕ_1 be such that $\beta(\phi_1) = \beta_* < \beta$. Let $R = \text{span}\{\beta c-x_0,\phi\}$. The inequality (2.25) is equivalent to $R(\beta c)$ being in the complement of the set RK. The supporting hyperplane to K with normal ϕ contains the point βc. Under the action of R this hyperplane becomes a supporting line to RK. Therefore, to show that $R(\beta c)$ lies in the complement of RK it suffices to prove that it is not a boundary point of RK.

Assume that $R(\beta c)$ is a boundary point of RK. Then there is an $x\in K$ such that $R(\beta c-x) = 0$. Since $(\beta c-x,\phi) = (\beta c-x,R\phi) = (R(\beta c-x),\phi) = 0$, x is in $M(\phi)$. Introduce the set $T = \text{span}\{\beta c-x_0,\phi,\phi_1\}$ and let $\{\psi,\phi,\zeta\}$ be an orthonormal basis for T. Since $(\beta c-x_0,\phi) = 0$, one has

$$\beta c-x_0 = b_1\psi + b_3\zeta ,$$

and since $(\beta c-x,\phi) = 0$, one has

$$\beta c-x = a_1\psi + a_3\zeta + T^{\perp}(\beta c-x) .$$

Also $(\beta c-x,\beta c-x_0) = (\beta c-x,R(\beta c-x_0)) = 0$ shows that

$$a_1b_1 + a_3b_3 = 0 .$$

Consider the points $x(\alpha)$, $\alpha\in[0,1]$ defined by $x(\alpha) = \alpha x_0 + (1-\alpha)x$. Note that $x(\alpha)\in M(\phi)$ for all α. Set

$$v(\alpha) = \beta c - x(\alpha) = \alpha(\beta c - x_0) + (1-\alpha)(\beta c - x) .$$

Differentiating the square of the norm of $v(\alpha)$ with respect to α and taking the limit as α tends to one yields

$$\begin{aligned}\lim_{\alpha\to 1}(||v(\alpha)||^2)' &= 2[b_1(b_1-a_1) + b_3(b_3-a_3)] \\ &= 2||\beta c - x_0||^2 .\end{aligned}$$

Since $v(1) = \beta c - x_0$, a contradiction to the minimality of $||\beta c - x_0||$ has been obtained. □

Let $\phi_0 = c$ and $\beta_0 = \max_{x\in K} \frac{(\phi,x)}{(\phi,c)} = \frac{(\phi,x_0)}{(\phi,c)}$. By Lemma 2.5.2, since $M(\phi_0) = \{x_0\}$, $\beta_1 = \min_{\alpha\in[0,1]} \beta(\alpha\phi_0 + (1-\alpha)(\beta_0 c - x_0)) < \beta_0$ unless $\beta_0 = \beta_*$. Let α_0 be chosen to minimize $\beta(\alpha\phi_0+(1-\alpha)(\beta_0 c-x_0))$. Set $\phi_1 = \alpha_0\phi_0 + (1-\alpha_0)(\beta_0 c - x_0)$. If $\beta_1 > \beta_*$ then

$$\beta_2 = \min_{\alpha\in[0,1]} \beta(\alpha\phi_1+(1-\alpha)(\beta_1 c-x_1)) < \beta_1$$

where $\{x_1\} = M(\phi_1)$. Proceeding in this way a sequence $\{\beta_j\}_{j=0}^{\infty}$ is generated which has the property that either $\beta_N=\beta_{N+1}=\cdots=\beta_*$ for some $N < \infty$ or $\beta_0>\beta_1>\beta_2>\cdots>\beta_*$.

<u>Theorem 2.5.3</u>. The sequence $\{\beta_j\}_{j=0}^{\infty}$ generated above converges to β_*.

<u>Proof</u>: The only non-trival case is when $\beta_0 > \beta_1 > \cdots > \beta_*$. Let $\beta = \lim_{n\to\infty} \beta_n$. It suffices to prove $\beta \le \beta_*$. Assume that $\beta > \beta_*$. By the compactness of the unit shell in H and K there is a subsequence $\{\phi_{n'}\}$ of $\{\phi_n\}$ which converges to ϕ and a subsequence $\{x_{n'}\}$ of $\{x_n\}$ which converges to $x \in K$. It follows that $\{x\} = M(\phi)$. By Lemma 2.5.1 there is an $a \in [0,1]$ such that $\beta(a\phi + (1-a)(\beta c - x)) < \beta$. But this contradicts $\beta > \beta_*$, for

$$\beta(a\phi_{n'} + (1-a)(\beta_{n'}c - x_{n'})) \ge \beta_n \quad \text{for all } n'$$

implies $\beta(a\phi + (1-a)(\beta c - x)) \ge \beta$.

2.6. <u>Minimax designs</u>. Let C be a subset of $\mathbb{R}^{k+1}$. If the design ξ^* satisfies

$$\sup_{c \in C} d(c, \xi^*) = \inf_{\Xi} \sup_{C} d(c, \xi)$$

it is said to be a minimax design with respect to C.

It is easily seen that if for some $c_0 \in C$, ξ_0 is optimal for estimating $c_0'\theta$ and

$$\sup_{C} d(c, \xi_0) = d(c_0, \xi_0) \tag{2.26}$$

then ξ_0 is minimax. This follows from

$$\sup_C d(c,\xi_0) = d(c_0,\xi_0) \le d(c_0,\xi) \le \sup_C d(c,\xi)$$

for any ξ.

Example 5.1. The minimax design for estimating the value of the mean function of Example 4.1 at $t = 1/4$ is the minimax design with respect to the set

$$C = \left\{ \begin{pmatrix} x(1/4) \\ x^{(1)}(1/4) \end{pmatrix} : x \in X \right\}.$$

Define the mapping $c\colon X \to R^2$ by $c(X) = \begin{pmatrix} x(1/4) \\ x^{(1)}(1/4) \end{pmatrix}$. Then $c(X) = C$, and for any design ξ_α which concentrates all mass on x_α, $\sqrt{d(c(\cdot),\xi_\alpha)}$ is continuous and convex on X. Thus Theorem 4.1 shows

$$\max_C d(c,\xi_\alpha) = \max_X d(c(x),\xi_\alpha) = \max_{E(X)} d(c(x),\xi_\alpha) .$$

A plot of the set $c[E(X)]$ suggests that the design optimal for estimating $1/32\ \theta_0 + 1/4\ \theta_1$ is minimax. This design, ξ^*, concentrates all mass at x_α for $\alpha = \frac{1}{4}$. Computation shows that for $c_0 = \begin{pmatrix} 1/32 \\ 1/4 \end{pmatrix}$,

$$\frac{1}{16} = d(c_0,\xi^*) = \max_{E(X)} d(c(x),\xi^*) \qquad \text{so that}$$

ξ^* is minimax.

Example 5.2. For a fixed input $\{x_0(t): t\in T\}$ one may wish to estimate quantities depending on the unknown mean $L(x_0,\theta)$, such as the value of the mean at certain times $t\in T$, or possibly its derivative there. If it is desired to hold the maximum variance of these estimators to a minimum we have a minimax design problem. The particular design problem is dictated by the collection of estimators. A collection of interest is the set of estimators of the quantities $(z,L(x_0,\theta))$ where z is in the unit ball of $H(K)$. That is, the estimators of certain continuous linear functionals of the mean. Since $L(x_0,\theta) \in H(K)$ and the evaluation functional is continuous in $H(K)$ this set always includes the estimators of the values of the mean at any time $t\in T$. Depending upon K, it may also include derivatives or integrals of the mean evaluated at points in T. With regard to the setup as described in Example 4.1, we find the minimax design with respect to the set $c(\Omega) = C$ where $\Omega = \{z\in H(K): ||z|| \le 1\}$ and $c(z) = (z,\underset{\sim}{f}(x_{1/3}))$. That is, we find the minimax design for the collection of estimators of the quantities $\{(z,L(x_{1/3},\theta)): ||z|| \le 1\}$, where $L(x_{1/3},\theta) = \theta_0 x_{1/3}(t) + \theta_1 x_{1/3}^{(1)}(t)$. The convex set $c(\Omega)$ may be plotted by plotting its boundary

$$\left\{ \frac{M(x_{1/3})\lambda}{\sqrt{\lambda' M(x_{1/3})\lambda}} : \lambda \epsilon S^{k+1} \right\}.$$

The plot suggests that a design optimal for estimating $3/54\ \theta_0 + \theta_1$ is minimax. Such a design ξ^* is the one which places all mass at $x_{2/3}$ and satisfies $d(\binom{3/54}{1}, \xi^*) = 1$. It can be seen that ξ^* is minimax by (2.23) since

$$\max_{\lambda \epsilon S^{k+1}} \frac{\lambda' M(x_{1/3}) M^{-1}(x_{2/3}) M(x_{1/3})\lambda}{\lambda' M(x_{1/3})\lambda} = 1 \ .$$

The preceding examples demonstrate that there are minimax designs which are distinct from designs which maximize $|M(\xi)|$, the determinant of $M(\xi)$. Furthermore $\min_{\Xi} \max_{C} d(c,\xi)$ is not, for these examples equal to k+1. Thus a useful theorem developed by Kiefer and Wolfowitz (1960) for the case when T is a one-point set does not apply to the minimax problems described above. There is however a natural design criterion which yields a direct analogue of the Kiefer-Wolfovitz theorem. A general description of this criterion is as follows. Suppose for each design $\xi \epsilon \Xi$ and x in X we can find a minimum "variance" linear unbiased estimator $\hat{m}_x$ of the entire mean function $\{\sum_{i=0}^{k} \theta_j f_j(x,t) : t \epsilon T\}$. That is, our estimator $\hat{m}_x$ is H(K) valued, its expected value is $m_x = \sum_{j=0}^{k} \theta_j f_j(x)$, and its "variance" is $E||\hat{m}_x - m_x||_K^2$. Defining the (new) function $d: X \times \Xi$ by

$$N^{-1}d(x,\xi) = E||\hat{m}_x - m_x||^2$$

one has the equivalence of the following:

i) ξ^* maximizes $|M(\xi)|$

ii) ξ^* minimizes $\sup_X d(x,\xi)$

iii) $\sup_X d(x,\xi^*) = k+1.$

We state these results more precisely below. First we investigate H-valued linear random variables defined on the observable process.

Let V be a Hilbert space with inner product $(\cdot,\cdot)_V$. Let Δ be an arbitrary set and consider a zero-mean stochastic process $\{Q(\delta): \delta\in\Delta\}$ defined on a probability space (Ω,A,P) with known covariance kernel C. Denote by H(C) the reproducing kernel Hilbert space of functions on Δ induced by C and by $(\cdot,\cdot)_C$ the inner product on this space. Let $L_2(V)$ be the space of all V-valued random variables Z on (Ω,A) for which

$$||Z||^2_{L_2(V)} = \int [||Z||^2_V]dP < \infty .$$

The space $L_2(V)$ is a Hilbert space with inner product $(Z_1,Z_2)_{L_2(V)} = \int (Z_1,Z_2)_V dP$. Let $V(C,\Delta) = \{\sum_{j=1}^{n} v_j Q(\delta_j): n < \infty,\ v_j\in V,\ \delta_j\in\Delta\}$ and $V_2(C,\Delta)$ be the closure of $V(C,\Delta)$ in $L_2(V)$.

Lemma 2.6.1. Let $R \in V_2(C,\Delta)$ satisfy

$$||R - \sum_{j=1}^{n} v_{nj} Q(\delta_{nj})||_{L_2(V)} \to 0 \text{ as } n \to \infty .$$

Then if $h \in H(C)$

$$\lim_{n\to\infty} \left(\sum_{j=1}^{n} v_{nj} (h+Q)(\delta_{nj}) \right)$$

exists in $L_2(V)$.

Proof: It suffices to prove that

$$S_n = \sum_{j=1}^{n} v_{nj} h(\delta_{nj})$$

is a Cauchy sequence in $L_2(V)$. Let m and n be arbitrary. By redefining coefficients properly one may write

$$S_n - S_m = \sum_{j=1}^{n} (v^*_{Nj} - w^*_{Nj}) h(\delta^*_{Nj}) \;\; ,$$

$N \le n+m$. For any $Z \in L_2(V)$, $||Z||_{L_2(V)} = 1$,

$$\begin{aligned} |(Z,S_n-S_m)_{L_2(V)}|^2 &= |\int (Z(\omega),S_m-S_n)_V dP(\omega)|^2 \\ &= |\sum_{j=1}^{N} [(\mu_Z,v^*_{Nj})_V - (\mu_Z,w^*_{Nj})_V] h(\delta^*_{Nj})|^2 \end{aligned} \tag{2.27}$$

where $\mu_z = \int ZdP$ (Bochner-integral). By the reproducing property of C the expression in (2.27) may be written

$$|(\sum_{j=1}^{N}[(\mu_z,v_{Nj}^*)_v-(\mu_z,w_{Nj}^*)_v]C(\cdot,\delta_{Nj}^*),h(\cdot))_c|^2 . \quad (2.28)$$

Using the Schwarz inequality for the inner product $(\cdot,\cdot)_c$ it can be seen that (2.28) is

$$\leq ||h||_c^2 \sum_{i=1}^{N}\sum_{j=1}^{N}(\mu_z,v_{Ni}^*-w_{Ni}^*)_v(\mu_z,v_{Nj}^*-w_{Nj}^*)_v C(\delta_{Ni}^*,\delta_{Nj}^*) . \quad (2.29)$$

It can be checked that the righthand sum in (2.29) is just

$$E|(\mu_z,\sum_{j=1}^{n}v_{nj}Q(\delta_{nj}) - \sum_{i=1}^{m}v_{mi}Q(\delta_{mi}))_v|^2 \quad (2.30)$$

so that using the Schwarz inequality on (2.30) for the inner product $(\cdot,\cdot)_v$ one obtains

$$|(Z,S_n-S_m)_{L_2(V)}|^2 \leq ||h||_c^2||\mu_z||_v^2||\sum_{j=1}^{n}v_{nj}Q(\delta_{nj})-\sum_{i=1}^{m}v_{mk}Q(\delta_{mi})||_{L_2(V)}^2 .$$

Since $||\mu_z||_v^2 \leq 1$ for all Z in the unit ball of $L_2(V)$ one has

$$||S_n-S_m||_{L_2(V)}^2 \leq ||h||_c^2||\sum_1^n v_{nj}Q(\delta_{nj}) - \sum_1^m v_{mi}Q(\delta_{mi})||_{L_2(V)}^2 . \quad (2.31)$$

It now follows from the assumptions that S_n is Cauchy in $L_2(V)$. □

IF $S(\Delta,V) = \bigcup_{n=1}^{\infty}\bigcup_{j=1}^{n}\{(v_{nj},\delta_{nj}):v_{nj}\in V,\delta_{nj}\in\Delta\}$ and $L_2(\Delta,C,V)$ is that subset of $S(\Delta,V)$ for which $\sum_{j=1}^{n} v_{nj}Q(\delta_{nj})$ converges in $L_2(V)$ then for each $L \in \mathcal{L}_2(\Delta,C,V)$ define $L(Q) = \lim \sum_{j=1}^{n} v_{nj}Q(\delta_{nj})$. Fix $x \in X$ and let $m(\underset{\sim}{\theta}) = \underset{\sim}{\theta}'f(\underset{\sim}{x}) \in H(K)$. Set $Z = X + \underset{\sim}{\theta}'\underset{\sim}{f}$ when $\underset{\sim}{\theta} \in R^{k+1}$, holds, where Z is as above with covariance B.

<u>Definition</u>. The set of linear estimators of the mean $m(\underset{\sim}{\theta}) \in H(K)$ is

$$\mathcal{L}_2(\Gamma,B,H(K)) \ .$$

The lemma above shows that if $L \in \mathcal{L}_2(\Gamma,B,H(K))$ then for any $\underset{\sim}{\theta} \in R^{k+1}$, $L(Z+\underset{\sim}{\theta}'\underset{\sim}{f})$ is in $L_2[H(K)]$.

<u>Definition</u>. $L_0 \in \mathcal{L}_2(\Gamma,B,H(K))$ is unbiased for $m(\underset{\sim}{\theta})$ if for all $\theta \in R^{k+1}$

$$||E(L_0(X+\underset{\sim}{\theta}'f)] - m(\underset{\sim}{\theta})||_K^2 \equiv 0 \ .$$

<u>Definition</u>. The estimator $L_0 \in \mathcal{L}_2(\Gamma,B,H(K))$ is said to be best linear unbiased for $m(\underset{\sim}{\theta})$ if

(i) L_0 is unbiased

and (ii) for any other linear unbiased estimator $L \in \mathcal{L}_2(\Gamma,B,H(K))$ and all $\underset{\sim}{\theta} \in R^{k+1}$

$$E||L_0(X+\underset{\sim}{\theta}'\underset{\sim}{f})-m(\theta)||_K^2 \le E||L(X+\underset{\sim}{\theta}'\underset{\sim}{f})-m(\underset{\sim}{\theta})||_K^2 \ .$$

Keeping x fixed let $M = \{\underset{\sim}{\theta}'\underset{\sim}{f}(x) \in H(K): \underset{\sim}{\theta} \in R^{k+1}\}$ and P_M be the orthogonal projection in H(K) onto M.

Lemma 2.6.2. If $L \in L_2(\Gamma, B, H(K))$ is unbiased for $m(\underset{\sim}{\theta})$ then so is $P_M L$ and for all $\underset{\sim}{\theta} \in R^{k+1}$

$$E||P_M L(X+\theta'f)-m(\theta)||_K^2 \le E||L(X+\theta'f)-m(\theta)||_K^2$$

Proof: Clearly $P_M L(X+\underset{\sim}{\theta}'\underset{\sim}{f}) \in L_2(\Gamma, B, H(K))$ and $P_M L(X+\underset{\sim}{\theta}'\underset{\sim}{f})$ is unbiased. Fix $\underset{\sim}{\theta}$ and set $X+\underset{\sim}{\theta}'\underset{\sim}{f} = U$. One has

$$\begin{aligned} E||LU-m||_K^2 &= E||LU-P_M LU + P_M LU - m||_K^2 \\ &= E||(I-P_M)LU||_K^2 + E((I-P_M)LU, P_M(LU-m))_K \\ &\qquad + E||P_M LU-m||_K^2 \\ &\ge E||P_M LU-m||_K^2 \, . \end{aligned}$$

Therefore, if there is a best linear unbiased estimator, it must have values in M with probability one.

Lemma 2.6.3. Let $L \in L_2(\Gamma, B, H(K))$ satisfy $L(X(\omega)+\underset{\sim}{\theta}'\underset{\sim}{f}) \in M$ for $\omega \in \Omega_0$, $P(\Omega_0) = 1$. Then there is $L^* \in L_2(\Gamma, B, R^{k+1})$ such that

$$||L(X+\underset{\sim}{\theta}'\underset{\sim}{f})-[L^*(X+\underset{\sim}{\theta}'\underset{\sim}{f})]'\underset{\sim}{f}(x)||_K = 0$$

with probability one.

Proof: Fix $\underset{\sim}{\theta}$ and set $U(\omega) = L(X(\omega)+\underset{\sim}{\theta}'\underset{\sim}{f})$.

Define $\underset{\sim}{V}(\omega) = \begin{bmatrix} (f_0, U(\omega))_K \\ \vdots \\ (f_k, U(\omega))_K \end{bmatrix}$ and

$\underset{\sim}{\tilde{\theta}}(\omega) = \underset{\sim}{M}^+ \underset{\sim}{V}(\omega)$ where $\underset{\sim}{M}_{ij} = (f_i(x), f_j(x))_K$.

Then for each $\omega \in \Omega_0$

$$||\underset{\sim}{\tilde{\theta}}'(\omega)\underset{\sim}{f}(x) - \underset{\sim}{U}(\omega)||_K = \sup_{||z||_K=1} |(z, \underset{\sim}{\tilde{\theta}}'(\omega)\underset{\sim}{f}(x) - U(\omega))_K| \tag{2.32}$$

$$= \sup_{\underset{\sim}{\alpha} \in A} |(\sum_{j=0}^{k} \alpha_j f_j(x), \underset{\sim}{f}'\underset{\sim}{\tilde{\theta}}(\omega) - U(\omega))_K| , \tag{2.33}$$

where $A = \{\underset{\sim}{\alpha} \in R^{k+1}: \underset{\sim}{\alpha}'\underset{\sim}{M}\underset{\sim}{\alpha} \le 1\}$. Thus (2.33) is

$$\sup_{\underset{\sim}{\alpha} \in A} |\underset{\sim}{\alpha}'\underset{\sim}{M}\underset{\sim}{\theta}(\omega) - \underset{\sim}{\alpha}'\underset{\sim}{V}(\omega)|$$

$$= \sup_{\underset{\sim}{\alpha} \in A} |\underset{\sim}{\alpha}'[\underset{\sim}{M}\underset{\sim}{M}^+\underset{\sim}{V}(\omega) - \underset{\sim}{V}(\omega)]| . \tag{2.34}$$

For $\omega \in \Omega_0$, $U(\omega) = \underset{\sim}{f}'\underset{\sim}{\theta}$ for some $\underset{\sim}{\theta} \in \mathbb{R}^{k+1}$ so that $\underset{\sim}{V}(\omega) = \underset{\sim}{M}\underset{\sim}{\theta}$. Thus for $\omega \in \Omega_0$, $\underset{\sim}{V}(\omega)$ is in the range of $\underset{\sim}{M}$. Since $\underset{\sim}{M}\underset{\sim}{M}^+$ is the projection onto the range of $\underset{\sim}{M}$ the expression in (2.34), and hence in (2.32), is zero for $\omega \in \Omega_0$. It follows that $||\underset{\sim}{\tilde{\theta}}'\underset{\sim}{f} - U||_K = 0$ with probability one. To finish the proof take $L^* \in L_2(\Gamma, B, R^{k+1})$ to correspond to the sequences

$$\bigcup_{n=1}^{\infty}\bigcup_{j-1}^{n}\{(\underset{\sim}{M}^{+}\begin{bmatrix}(f_0,v_{nj})_K\\ \vdots \\ (f_k,v_{nj})_K\end{bmatrix},\gamma_{nj})\} \text{ where } L\in\mathcal{L}_2(\Gamma,B,H(K))$$

corresponds to $\bigcup_{n=1}^{\infty}\bigcup_{j=1}^{n}\{(v_{nj},\gamma_{nj})\}$.

In our notation the random vector

$$N^{-1}(<z,f_0>_B,\cdots,(<z,f_k>_B)\underset{\sim}{M}^{+}(\xi)\underset{\sim}{f}(x)$$

is written

$$N^{-1}\underset{\sim}{M}^{+}(\xi)\begin{bmatrix}L_0(X+\theta'f)\\ \vdots \\ L_k(X+\theta'f)\end{bmatrix}, \tag{2.35}$$

where $L_j\in\mathcal{L}_2(\Gamma,B,R)$, $j = 0,\cdots,k$. The operator L_j corresponds to $\bigcup_{n=1}^{\infty}\bigcup_{i=1}^{n}\{(a_{ni}^{(j)},\gamma_{ni}^{(j)})\}$ where $||f_j(\cdot)-\sum_{i=1}^{n}a_{ni}^{(j)}B(\cdot,\gamma_{ni}^{(j)})||_B \to 0$ as $n\to\infty$.

For any matrix $\underset{\sim}{A}$ let $\mathcal{R}(\underset{\sim}{A})$ denote the range of $\underset{\sim}{A}$.

<u>Lemma 2.6.4</u>. If $L\in\mathcal{L}_2(\Gamma,B,H(K))$ is unbiased for $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$ then

$$\{(y,\underset{\sim}{f}(x))_K: y\in H(K)\} \subseteq \mathcal{R}[\underset{\sim}{M}(\xi)] .$$

Proof: One may check that for any $y \in H(K)$, $(y,L)_K \in L_2(\Gamma,B,R)$ and is unbiased for $(y,\underset{\sim}{f}(x))'_K\underset{\sim}{\theta}$. Also, $(y,L)_K$ is the image of some g in H(B) (see Lemma 2.2.2). Hence the image of Pg, where P is the orthogonal projection onto the subspace $\{\underset{\sim}{\theta}'\underset{\sim}{f}: \underset{\sim}{\theta} \in \mathbb{R}^{k+1}\}$, is also unbiased and has expected value

$$\underset{\sim}{\theta}'\underset{\sim}{M}(\xi)N\underset{\sim}{\alpha} \tag{2.36}$$

for some $\underset{\sim}{\alpha} \in R^{k+1}$. Since for all $\underset{\sim}{\theta}$ (2.36) equals $(y,\underset{\sim}{f}(x))'\underset{\sim}{\theta}$ the assertion has been proven.

Lemma 2.6.5. If $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$ is linearly estimable then

$$\underset{\sim}{f}'(x)N^{-1}\underset{\sim}{M}^{+}(\xi)\begin{bmatrix} L_0 \\ \vdots \\ L_k \end{bmatrix} \tag{2.37}$$

is in $L_2(\Gamma,B,H(K))$ and is unbiased for $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$. The estimator

$$N^{-1}\underset{\sim}{M}^{+}(\xi)\begin{bmatrix} L_0 \\ \vdots \\ L_k \end{bmatrix} = L_0^*$$

is in $L_2(\Gamma,B,R^{k+1})$ and for any other unbiased estimator $L \in L_2(\Gamma,B,H(K))$ which has values in M almost surely the covariance of the associated $L^* \in L_2(\Gamma,B,R^{k+1})$ satisfies

$$\underset{\sim}{c}'\underset{\sim}{cov}(L_0^*)\underset{\sim}{c} \le \underset{\sim}{c}'\underset{\sim}{cov}(L^*)\underset{\sim}{c}$$

for all $\underset{\sim}{c} \in \{(y,\underset{\sim}{f}(x))_K : y \in H(K)\}$.

Proof: The reader may check that (2.37) is in $L_2(\Gamma,B,H(K))$. The unbiasedness follows from Lemma 2.6.4 above and Lemma 2.2.2

Let L_j correspond to $\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{n} \{(a_{ni}^{(j)},\gamma_{ni}^{(j)})\}$, $j = 0,1,\cdots,k$. By defining the components in an appropriate way it can be checked that L_0^* corresponds to

$$\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{M(n)} \{(b_{Mi}^*,\gamma_{Mi}^*)\} ,$$

where $b_{Mi}^* \in R^{k+1}$ for each i, $\gamma_{Mi}^* \in \Gamma$, and $M = M(n) \le (k+1)n$. It may be shown that L_0^* is in $L_2(\Gamma,B,R^{k+1})$ using this correspondence.

If $L \in L_2(\Gamma,B,H(K))$ is unbiased for $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$ then $(y,L)_K$ is in $L_2(\Gamma,B,R)$ and is unbiased for $(y,\underset{\sim}{f}(x))'\underset{\sim}{\theta}$. Since $(y,\underset{\sim}{f}(x))_K' L^* = (y,L)_K$ a.s. if L has values in M with probability one, the variance of $(y,L)_K$ is

$$N^{-1}\underset{\sim}{c}'\underset{\sim}{cov}(L^*)\underset{\sim}{c} \quad \text{where} \quad \underset{\sim}{c} = (y,\underset{\sim}{f}(x))_K .$$

The variance of $\underset{\sim}{c}'L_0^*$ is, from Lemma 2.2.2, $N^{-1}\underset{\sim}{c}'\underset{\sim}{cov}(L_0^*)\underset{\sim}{c}$. The conclusion now follows from the fact that $\underset{\sim}{c}'L_0^*$ is the umvlue of $\underset{\sim}{c}'\underset{\sim}{\theta}$. $\square$

The main result of this section can now be proved under the additional assumption below.

(A4) There is a sequence of finite subsets $T_n \subseteq T_{n+1} \subseteq T$, $n = 1,2,\cdots$ such that for all f and g in H(K)

$$(f,g)_{T_n} = \underset{\sim}{f}'_{T_n} \underset{\sim}{K}^+(T_n) \underset{\sim}{g}_{T_n} , \qquad (2.38)$$

$$\lim_{n\to\infty} (f,g)_{T_n} = (f,g)_K \text{ , and}$$

$$||f||^2_{T_n} = (f,f)_{T_n} \text{ , is non-decreasing in n.}$$

The notation $\underset{\sim}{f}'_{T_n}$ means the n-vector $(f(t_{n1})),\cdots,f(t_{nN(n)}))$ where $T_n = \{t_{n1},\cdots,t_{nN(n)}\}$ and $\underset{\sim}{K}(T_n)$ has $i,j^{\underline{th}}$ entry $K(t_{ni},t_{nj})$. Parzen (1959) in his Theorem 6E shows that, for example, if T is a separable metric space and K is weakly continuous on T then (A4) is satisfied for any non-decreasing sequence T_n whose union $\bigcup_n T_n$ is dense in T.

<u>Theorem 2.6.1</u>. Under the assumptions (A1) - (A4), and if there is an unbiased estimator in $L_2(\Gamma,B,H(K))$ of $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$, then the estimator

$$\underset{\sim}{f}'(x)L_0^* = (N^{-1}\underset{\sim}{M}^+(\xi)\begin{bmatrix} L_0 \\ \vdots \\ L_k \end{bmatrix})'\underset{\sim}{f}(x) \text{ ,}$$

where L_0^* is described in (2.35), is the uniformly best unbiased estimator in $L_2(\Gamma,B,H(K))$.

Proof: By Lemma 2.6.5 the only part to be proved is that $\underset{\sim}{f}'(x)L_0^*$ is best. Let L in $L_2(\Gamma,B,H(K))$ be an arbitrary unbiased estimator for $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$. We may as well assume that $L = P_M L$ (see Lemma 2.6.2). Let L^* be the associated estimator in $L_2(\Gamma,B,R^{k+1})$ given by Lemma 2.6.3. If T_n is a finite subset as in (A4) then

$$E||f'(x)L_0^*(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||_{T_n}^2$$

$$= tr\{\underset{\sim}{K}^+(T_n)\underset{\sim}{F}'_{T_n}(x)\ \underset{\sim}{cov}(L_0^*)\underset{\sim}{F}_{T_n}(x)\}$$

where the matrix $\underset{\sim}{F}_{T_n}(x)$ has $ij^{\underline{th}}$ entry

$$f_i(x,t_{nj}),\ i = 0,\cdots,k,\ j = 1,\cdots,N(n)\ .$$

For L^* one has

$$E||f'(x)L^*(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||_{T_n}^2$$

$$= tr\{\underset{\sim}{K}^+(T_n)\underset{\sim}{F}_{T_n}(x)\ \underset{\sim}{cov}(L^*)\underset{\sim}{F}_{T_n}(x)\}\ .$$

Now for any vector $\underset{\sim}{a}\in R^{N(n)}$

$$\underset{\sim}{F}_{T_n}(x)\underset{\sim}{a} = \sum \begin{bmatrix} f_0(x,t_{nj})a_j \\ f_k(x,t_{nj})a_j \end{bmatrix}$$

$$= (\sum a_j K(\cdot,t_{nj}), \underset{\sim}{f}(x))_K$$

so that, since $\sum a_j K(\cdot,t_{nj}) \in H(K)$, one has by Lemma 2.6.5 that

$$\underset{\sim}{F}'_{T_n}(x)[\underset{\sim}{cov}(L^*)-\underset{\sim}{cov}(L_0^*)]\underset{\sim}{F}_{T_n}(x)$$

is non-negative definite. Therefore, for any finite subset T_n

$$E||\underset{\sim}{f}'(x)L_0^*(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||^2_{T_n}$$

$$\leq E||\underset{\sim}{f}'(x)L^*(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||^2_{T_n} .$$

The conclusion of the theorem now follows from (A4) and the monotone convergence theorem. □

An especially simple treatment of the above results is available if the estimators are H(K) valued and H(K) is separable. In this case the space of linear H(K)-valued random variables defined on Z consists of elements of the form $\sum_{j\geq 1} <Z,g_j>_B \phi_j$, where $\{\phi_j\}$ is a complete orthonormal system for H(K), g_j are each in H(B), and $\sum_{j\geq 1} E[<Z,g_j>_B^2] < \infty$.

In this context an unbiased estimator U of $\theta'\underset{\sim}{f}(x)$ is one which satisfies the following equivalent conditions.

$$||E_\theta[U]-\theta'\underset{\sim}{f}(x)||_K = 0$$

$$E_\theta[U(t)] = \sum_{j=0}^{k} \theta_j f_j(x,t) \quad \text{for } t\epsilon T .$$

A "best" linear unbiased estimator U of $\theta'\underset{\sim}{f}(x)$ is one which is unbiased and for any other unbiased estimator V one has

$$E_\theta||V-\theta'\underset{\sim}{f}(x)||_K^2 \geq E_\theta||U-\theta'\underset{\sim}{f}(x)||_K^2 .$$

Theorem 2.6.1'. If an unbiased estimator of $\theta'\underset{\sim}{f}(x)$ exists then the estimator (2.35) is the best unbiased estimator.

Proof: Using extensions of Parzen (1959) or Lemma 2.2.2 it follows that the estimator in (2.35) is unbiased and for any finite set $T_n = \{t_1,\cdots,t_n\} \subset T$ $(\sum a_j K(\cdot,t_j),\hat{m})$ is the umvlue of $(\sum a_j K(\cdot,t_j),\theta'\underset{\sim}{f}(x))$. Let $U = \hat{m} - \theta'\underset{\sim}{f}(x)$ and V be the corresponding quantity for any unbiased estimator. For $h\epsilon H(K)$ of the form $h = \sum a_j K(\cdot,t_j)$

$$E(h,U)^2 = a' \operatorname{cov}(U_{T_n})a .$$

Therefore $E(h,U)^2 \leq E(h,V)^2$ holds for a dense subset of $h\epsilon H(K)$. From this it can readily be deduced that $E||U||_K^2 \leq E||V||_K^2$. □

We assume below that (A1) - (A4) hold. Let $L_0(x)\epsilon L_2(\Gamma,B,H(K))$ be best unbiased for $\underset{\sim}{f}'(x)\underset{\sim}{\theta}$. From above $L_0(x) = \underset{\sim}{f}'(x)L_0^*$ where $L_0^*\epsilon L_2(\Gamma,B,R^{k+1})$. Let $\underset{\sim}{M}(x) = \underset{\sim}{M}(\xi_x)$ where ξ_x places all mass at x.

Lemma 2.6.5. If $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$ is linearly estimable then $L_0(x)$ is best unbiased for $\underset{\sim}{\theta}'\underset{\sim}{f}(x)$ and

$$E||L_0(x)(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||_K^2 = N^{-1}\mathrm{tr}\{M^+(\xi)M(x)\}$$

Proof: Let $T_n \subset T$ be a finite subset as in (A4). Then

$$E||L_0(x)(X+\underset{\sim}{\theta}'\underset{\sim}{f})-\underset{\sim}{f}'(x)\underset{\sim}{\theta}||_{T_n}^2 = \mathrm{tr}\{\underset{\sim}{K}^+(T_n)\underset{\sim}{F}'_{T_n}(x)\underset{\sim}{\mathrm{cov}}(L_0^*)\underset{\sim}{F}_{T_n}(x)\} \quad (2.39)$$

Since $\underset{\sim}{\mathrm{cov}}(L_0^*) = N^{-1}\underset{\sim}{M}^+(\xi)$, the expression in (2.39) is just

$$\begin{aligned} & N^{-1}\mathrm{tr}\{\underset{\sim}{K}^+(T_n)\underset{\sim}{F}'_{T_n}(x)\underset{\sim}{M}^+(\xi)\underset{\sim}{F}_{T_n}(x)\} \\ &= N^{-1}\mathrm{tr}\{\underset{\sim}{M}^+(\xi)\underset{\sim}{F}_{T_n}(x)\underset{\sim}{K}^+(T_n)\underset{\sim}{F}'_{T_n}(x)\} \\ &= N^{-1}\mathrm{tr}\{\underset{\sim}{M}^+(\xi)\underset{\sim}{M}_{T_n}(x)\} \end{aligned}$$

where the $i,j^{\underline{th}}$ entry of $M_{T_n}(x)$ may be written, using (2.38), as $(f_i(x),f_j(x))_{T_n}$. The conclusion of the lemma now follows using (A4) and taking the limit as $n \to \infty$. $\square$

In all the above ξ has been a discrete design. The following definitions are made for arbitrary designs $\xi \in \Xi$.

<u>Definition</u>: The mean $\underset{\sim}{f}'(x)\underset{\sim}{\theta}$ is said to be estimable with respect to the design ξ if $\{(y,\underset{\sim}{f}(x))_K : y \in H(K)\} \subseteq \mathcal{R}[\underset{\sim}{M}(\xi)]$.

<u>Definition</u>: Let $d(x,\xi) = \mathrm{tr}\{\underset{\sim}{M}^+(\xi)\underset{\sim}{M}(x)\}$ if $\underset{\sim}{f}'(x)\underset{\sim}{\theta}$ is estimable with respect to ξ and $+\infty$ otherwise.

<u>Definition</u>: The design ξ^* is said to be a minimax design if

$$\inf_{\xi \in \Xi} \sup_{x \in X} d(x,\xi) = \sup_{x \in X} d(x,\xi^*) \ .$$

<u>Theorem 2.6.2</u>. Under (A1) - (A4) the coditions

(i) ξ^* maximizes $|\underset{\sim}{M}(\xi)|$

(ii) ξ^* minimizes $\sup_{x} d(x,\xi)$

(iii) $\sup_{x} d(x,\xi^*) = k+1$

are equivalent. The set Γ of all ξ^* satisfying these conditions is convex and closed and $\underset{\sim}{M}(\xi^*)$ is the same for all $\xi^* \in \Gamma$.

<u>Proof</u>: If there is an $x \in X$ such that $\underset{\sim}{f}'(x)\underset{\sim}{\theta}$ is not estimable $d(x,\xi) = +\infty$. By (A3) there is a design ξ_0 for which $|\underset{\sim}{M}(\xi_0)| \neq 0$. Thus $\mathcal{R}[\underset{\sim}{M}(\xi_0)] = R^{k+1}$ and $\underset{\sim}{f}'(x)\underset{\sim}{\theta}$ is estimable for all $x \in X$. Therefore attention may be restricted to those $\xi \in \Xi$ for which $|\underset{\sim}{M}(\xi)| > 0$. The remainder of the proof is exactly, except that the matrixes $\underset{\sim}{M}(\xi)$ differ, as it appears in Karlin and Studden (1966). □

For finding the D-optimal design the iterative process given below can be shown to converge in exactly the same manner as in Feforov (1972), Theorem 5.2.2.

1. Let ξ_0 be such that $|\underset{\sim}{M}(\xi_0)| > 0$.

2. Find x_0 to maximize

   $\mathrm{tr}\{\underset{\sim}{M}^{-1}(\xi_0)\underset{\sim}{M}(x)\}$.

3. The design $\xi_1 = (1-\alpha_0)\xi_0 + \alpha_0\xi(x_0)$ is constructed.

4. The matrix $\underset{\sim}{M}(\xi_1)$ is found and its inverse is computed.

Operations 2-4 are repeated with ξ_1, then with ξ_2, and so on as long as one of the inequalities

$$\max_x\{\mathrm{tr}\, M^{-1}(\xi_s)M(x)\} - (k+1) \le \delta_1$$

$$\frac{|M(\xi_{s+1})| - |M(\xi_s))|}{|M(\xi_{s+1})|} \le \delta_2 ,$$

where δ_1 and δ_2 are small positive preassigned numbers, is violated. The quantities α_s are conveniently chosen to satisfy $\alpha_s \ge 0$, $\lim_{s\to\infty} \alpha_s = 0$, and $\sum\alpha_s = +\infty$.

3. General linear parameter space.

3.1. Introduction. The action of an unknown variable force $\{\theta(t): t \in [0,1]\}$ on a particle of known mass $x \in [a,b]$ which is initially at rest may be observed repeatedly. The results are marred by an observational error of zero mean and known covariance. In an experiment with N uncorrelated observations $\{Y(x_i,t): t \in [0,1]\}_{i=1}^{N}$ of the position function of the particle over time, what is the "best" selection of masses $\{x_1,\cdots,x_N\}$ for estimating $\int_0^1 \theta(s)ds$ (see Example 2.1)?

The problem described above is a particular case of a type of design problem whose solution is characterized below. In the case of the general problem, for each x in a set X of possible levels of feasible experiments an experiment can be performed whose outcome is a stochastic process $\{Y(x,t): t \in T\}$. It is assumed that the process has a mean function $m(x,\theta,t)$ of known form, linear in the unknown parameter θ. The parameter θ is an element of a linear, but otherwise arbitrary, space Θ. For each x in X and θ in Θ the function $m(x,\theta)$ on T is a member of the reproducing kernel Hilbert space $H(K)$ generated by the known covariance kernel $K(s,t) = \mathrm{cov}[Y(x,s),Y(x,t)]$, $x \in X, s,t \in T$. The value $\tau(\theta)$, where τ is a linear (not necessarily continuous) functional on Θ, is to be estimated on the basis of N uncorrelated observations $\{Y(x_i,t): t\epsilon T,\ i = 1,\cdots,N\}$. The problem to be solved is to find the experimental design which minimizes the variance of the best linear unbiased estimator of $\tau(\theta)$.

This model generalizes that of chapters 1 and 2 by allowing a more general mean and parameter space. For the most part, the development below parallels that above. There is at least one major difference however. Under assumptions similar to those in the finite-dimensional case, the variance of the blue may fail to be minimized by any probability measure on the factor space which concentrates on a finite number of points. A design is called optimal below only if it minimizes the variance of the blue and is supported on a finite number of points.

3.2. Preliminary results. The class of linear estimators of $\tau(\theta)$ is $\{\langle Z,g\rangle_B: g\epsilon H(B)\}$. The functional $\tau(\theta)$ is said to be (linearly) estimable with respect to this design if there is a $g\epsilon H(B)$ such that $E_\theta\langle Z,g\rangle_B \equiv \tau(\theta)$. Denote by $(\cdot,\cdot)_B$ the inner product on $H(B)$ and for ξ as given above fixed, by m the map from Θ to $H(B)$ defined by $m(\theta)(\gamma) = m(x_i,\theta,t)$ if $\gamma = ((x_i,j),t)$. Parzen (1959) proved the following.

Theorem (10A). Given that $\tau(\theta)$ is estimable, there is a unique linear estimator $\langle Z,g_0\rangle_B$ which is the uniformly minimum variance linear unbiased estimator (umvlue) of $\tau(\theta)$ with variance $||g_0||_B^2$.

Furthermore $\langle Z,g\rangle_B$ is the umvlue of $\tau(\theta)$ if and only if g is the unique function in the closure of $R(m) = \{m(\theta): \theta\epsilon\Theta\}$ satisfying $\tau(\theta) \equiv (m(\theta),g)_B$, $\theta\epsilon\Theta$.

Let m' be the map from H(B) to the space of linear functionals Θ' on Θ defined by

$$m'(v)(\theta) = (v,m(\theta))_B .$$

The map m' is the restriction of the usual transpose to H(K) (see Taylor (1958)). Setting

$$M = N^{-1}m'm \tag{3.1}$$

it will be shown below that an expression for the variance of the umvlue of $\tau(\theta)$ may be obtained which is analogous to that obtained for $\Theta = R^{k+1}$ above. If $\Theta = R^{k+1}$ the variance of the umvlue of $\tau(\theta)$ is $N^{-1}\tau'M^{+}\tau$ where M^{+} is the Moore-Penrose generalized inverse of M.

The notion of a generalized inverse extends to mappings between arbitrary linear spaces (see Nashed and Votruba (1976)). In particular, we use the notion of an algebraic generalized inverse (A.G.I.). The reader is referred to Proposition 1.16 and the material preceding Proposition 1.17 of Nashed and Votruba (1976) which shows that given the linear operator $L: \Theta \rightarrow \Theta'$ there are (algebraic) projectors P and Q, P defined on Θ and Q on Θ', such that

$$\Theta = N(L) \dot{+} M ,$$

$$\Theta' = R(L) \dot{+} S ,$$

$$R(P) = M(L), \ R(I-P) = M , \tag{3.2}$$

$$R(Q) = R(L) \ , \text{ and } R(I-Q) = S .$$

The symbol $\dot{+}$ means algebraic direct sum so that P and Q are not necessarily continuous projectors. Indeed, we should emphasize that no topological assumptions have been made, or will be made in this section. For a mapping A, $R(A)$ denotes the range of A and $N(A)$ denotes the null space of A. Nashed and Votruba demonstrate the existence of a unique (for P and Q fixed) linear operator $L^{\#}: \Theta' \to \Theta$ called the A.G.I. of L which satisfies

$$LL^{\#}L = L$$
$$L^{\#}LL^{\#} = L^{\#}$$
$$L^{\#}L = I - P$$
$$LL^{\#} = Q .$$

We note that if $\theta' = L\theta$ then $\theta - L^{\#}\theta' = (I-L^{\#}L)\theta$ or

$$\theta = L^{\#}\theta' + (I-L^{\#}L)\theta . \tag{3.3}$$

If the form $\tau(\theta)$, $\tau\epsilon\Theta'$, is estimable then there is a umvlue which corresponds to, say $g_0 \in \overline{R(m)}$ (the closure of $R(m)$). Let $\theta_n \epsilon \Theta$ be such that $\lim_{n\to\infty} ||m(\theta_n)-g_0||_B = 0$ and define $\theta_n' = L\theta_n$, where $L = NM$.

Lemma 3.2.1. (i) The form $\tau(\theta)$ is estimable if and only if $\tau\epsilon R(m')$.

(ii) If $\tau(\theta)$ is estimable then $||g_0||_B^2 = N^{-1} \lim_{n\to\infty} \theta_n'(M^{\#}\theta_n')$.

Proof: (i) $\tau(\theta) \equiv (m(\theta),g_0)_B$ if and only if $\tau(\theta) \equiv m'g_0(\theta)$.

(ii)
$$
\begin{aligned}
||g_0||_B^2 &= \lim_{n\to\infty} ||m(\theta_n)||_B^2 \\
&= \lim_{n\to\infty} m'm(\theta_n)(\theta_n) \\
&= \lim_{n\to\infty} L(\theta_n)(\theta_n) \\
&= \lim_{n\to\infty} \theta_n'(L^{\#}\theta_n' + (I-L^{\#}L)\theta_n) \\
&= \lim_{n\to\infty} \theta_n'(L^{\#}\theta_n') + \theta_n'(I-L^{\#}L)\theta_n .
\end{aligned}
$$

The result will follow as soon as it is shown that $\theta_n'(I-L^{\#}L)\theta_n = 0$. To see this, suppose $u\epsilon N(L)$. Then for all $y\epsilon\Theta, 0 = Lu(y) = N(m(u),m(y))_B$. In particular, $m(u) = 0$ so $N(L)\subset N(m)$. Since $(I-L^{\#}L)\theta_n \in N(L)$ and

$$\theta_n'(I-L^{\#}L)\theta_n = N(m(\theta_n), m(I-L^{\#}L)\theta_n)_B$$

the lemma has been proved.

Corollary. If $\tau(\theta)$ is estimable and $R(m)$ is closed in H(B) the variance of the umvlue of $\tau(\theta)$ is given by

$$V = N^{-1}\tau(M^{\#}\tau) .$$

Denote by Ξ the set of all probability measures on X with finite support. Let H_ξ be the Hilbert space of real valued functions f on $S(\xi)xT$, where $S(\xi)$ is the support of ξ, such that $f(x_i,\cdot) \in H(K)$ for $x_i \in S(\xi)$. The inner product of f and g is

$$(f,g)_\xi = \sum_{x \in S(\xi)} \xi(x)(f(x),g(x))_K .$$

Denote by m_ξ the map from Θ to H_ξ defined by $m_\xi(\theta)(x_i,t) = m(x_i,\theta,t)$ for $x_i \in S(\xi)$, $t \in T$. Let

$$M(\xi) = m_\xi' m_\xi . \tag{3.4}$$

If ξ has all its mass at a point $x \in X$ write m_x rather than m_ξ.

When ξ has rational probabilities at its support points (3.1) and (3.4) coincide. We are now in a position to make the following definitions for $\xi \in \Xi$.

Definition. The linear form $\tau(\theta)$ is estimable with respect to the design $\xi \in \Xi$ if $\tau \in R(m_\xi')$.

Let $M_\xi = \{m\ (\theta): \theta \in \Theta\}$. Then $M_\xi \subset H_\xi$ and if $\tau(\theta)$ is estimable there is a unique $g \in \bar{M}_\xi$ (the closure of M_ξ) such that

$(m_\xi(\theta),g)_\xi \equiv \tau(\theta)$ and a sequence $\{\theta_n\} \subset \Theta$ such that $||m_\xi(\theta_n)-g||_\xi \to 0$ as $n \to \infty$. Let $\theta_n' = M(\xi)\theta_n$ and define

$$d(\tau,\theta) = \lim_{n\to\infty} \theta_n' M^{\#}(\xi)\theta_n' . \tag{3.5}$$

If τ is not estimable let $d(\tau,\xi) = +\infty$.

Definition. The design $\xi_0 \in \Xi$ is said to be optimal for estimating $\tau(\theta)$ if $d(\tau,\xi_0) = \inf_{\Xi} d(\tau,\xi)$.

Lemma 3.2.2. For $\xi \in \Xi$ the operator $M(\xi)$ is given by $M(\xi) = \int m_x' m_x d\xi(x)$.

Proof: For u and v in Θ

$$\begin{aligned}(M(\xi)u)v = (m_\xi' m_\xi u)v &= (m_\xi(u), m_\xi(v))_\xi \\ &= \int (m(x,u), m(x,v))_K d\xi(x) \\ &= \int (m_x'(m_x u))v d\xi(x) .\end{aligned}$$

Lemma 3.2.3. For any $\theta_1, \theta_2 \in \Theta$ and $\xi \in \Xi$

$$|(M(\xi)\theta_1)\theta_2| \le \sqrt{(M(\xi)\theta_1)\theta_1}\ \sqrt{(M(\xi)\theta_2)\theta_2}$$

with equality if and only if $m_\xi(\theta_1) = km_\xi(\theta_2)$ for some constant k.

Proof: Set $||m_\xi(\theta_1)-sm_\xi(\theta_2)||_\xi^2 = f(s)$. Then the real valued function f defined on $(-\infty,+\infty)$ has a minimum at

$s_0 = \dfrac{(M(\xi)\theta_2)\theta_1}{(M(\xi)\theta_2)\theta_2}$ if $m_\xi(\theta_2) \neq 0$. Using the fact that $f(s) \geq 0$,

with equality if and only if $m_\xi(\theta_1) = km_\xi(\theta_2)$, the result is proved as in the usual finite dimensional case.

In addition to the algebraic generalized inverse of a linear operator T it is useful to have another notion of the transpose. It differs somewhat from both the usual transpose (see Taylor (1958), Chapter 1) and the transpose defined above. In particular, we suppose that $T: \Theta \to \Theta'$, where as above, Θ is a linear space and Θ' is the space of all linear functionals on Θ.

Definition. The t-transpose of T, written T^t, is the linear operator mapping Θ into Θ' defined by $Tx(y) = T^t y(x)$, for all x,y in Θ. The following lemma is true for T^t but is not in general true for the usual transpose. Some additional notation is required. For an arbitrary subset $A \subset \Theta$ let $A^\perp = \{\theta' \subset \Theta': \theta'(\theta) = 0 \text{ for all } \theta \in A\}$. For an arbitrary subset $B \subset \Theta'$ let $B^\perp = \{\theta \in \Theta: \theta'(\theta) = 0 \text{ for all } \theta' \in B\}$.

Lemma 3.2.4. (i) $\mathcal{R}^\perp(T) = \mathcal{N}(T^t)$

(ii) $\mathcal{R}^\perp(T^t) = \mathcal{N}(T)$

(iii) $\mathcal{R}(T^t) = \mathcal{N}^\perp(T)$

(iv) $\mathcal{R}(T) = \mathcal{N}^\perp(T^t)$.

Proof: Parts (i) and (ii) are proved in a similar manner, so we only prove (i). Let $x \in R^{\perp}(T)$. Then $(Ty)x = 0$ for all $y \in \Theta$. Thus $T^t x(y) = 0$ for all $y \in \Theta$. Therefore $T^t x = 0$. The reverse inclusion follows from the same argument.

If W is any subspace contained in Θ then $W = W^{\perp\perp}$ (see Taylor (1958), Thm. 1.9-A). Parts (iii) and (iv) follow immediately.

Lemma 3.2.5. For any $\theta' \in \Theta'$, $\theta' \neq 0$, and $\xi \in \Xi$

$$d(\theta', \xi) = \sup_{\theta \in U} \frac{(\theta'(\theta))^2}{(M(\xi)\theta)\theta}, \tag{3.6}$$

where $U = \{\theta\colon (M(\xi)\theta)\theta \neq 0\}$.

Proof: As in (3.2), write $\Theta = N(M) \dot{+} M$. First, suppose $\theta' \in R(M)$. Then $\theta' = MM^{\#}\theta'$ since $MM^{\#}$ is the projection onto $R(M)$. Thus

$$\begin{aligned}
(\theta'(\theta))^2 &= [(MM^{\#}\theta')\theta]^2 \\
&= (M(M^{\#}\theta')\theta)^2 \\
&\leq [M(M^{\#}\theta')M^{\#}\theta'][(M\theta)\theta] \\
&= [\theta'(M^{\#}\theta')][(M\theta)\theta],
\end{aligned}$$

for any θ. If $\theta \in U$ then $(M\theta)\theta \neq 0$ so

$$\frac{(\theta'(\theta))^2}{(M\theta)\theta} \leq (M^{\#}\theta')\theta'. \tag{3.7}$$

Setting $\theta = M^{\#}\theta'$, which is in U, equality is achieved in (3.7), proving the result for $\theta' \in R(M)$.

If $\theta' \in \mathcal{R}(m') - \mathcal{R}(M)$ then $\theta'(\theta)$ is estimable with umvlue corresponding to some $g_0 \in \bar{M}_\xi \subset H(\xi)$. Let $\{\theta_n\} \subset \Theta$ be such that $||m(\theta_n)-g_0||_\xi \to 0$ and set $\theta_n' = M\theta_n$. The linear form $\theta_n'(\theta)$ is estimable and its umvlue corresponds to $m(\theta_n)$ since

$$\theta_n'(\theta) = (m(\theta_n), m(\theta))_\xi$$

and $m(\theta_n) \in M_\xi$. Furthermore $\theta_n' \in \mathcal{R}(M)$, so by the above argument

$$d(\theta_n', \xi) = \sup_{\theta \in U} \frac{(\theta_n'(\theta))^2}{(M\theta)\theta} \text{ and}$$

$$d(\theta_n', \xi) = ||m(\theta_n)||^2 \to ||g_0||^2 = d(\theta', \xi) . \tag{3.8}$$

Notice that for fixed θ

$$\begin{aligned}
\left|\frac{[\theta_n'(\theta)]^2}{(M\theta)\theta} - \frac{[\theta'(\theta)]^2}{(M\theta)\theta}\right| &= \left|\frac{(m(\theta_n),m(\theta))^2 - (g_0,m(\theta))^2}{||m(\theta)||_\xi^2}\right| \\
&= \left|\frac{(m(\theta_n) - g_0, m(\theta))_\xi (m(\theta_n) + g_0, m(\theta))_\xi}{||m(\theta)||_\xi^2}\right| \\
&\le ||m(\theta_n) - g_0||_\xi \; ||m(\theta_n) + g_0||\xi ,
\end{aligned} \tag{3.9}$$

so that given $\varepsilon > 0$ it follows from (3.9) that for n sufficiently large $\left|d(\theta_n', \xi) - \sup_{\theta \in U} \frac{[\theta'(\theta)]^2}{(M\theta)\theta}\right| < \varepsilon/2$ and from (3.8) that $|d(\theta_n', \xi) - d(\theta', \xi)| < \varepsilon/2$. Hence for n sufficiently large

$$|d(\theta',\xi) - \sup_{\theta\in U} \frac{[\theta'(\theta)]^2}{(M\theta)\theta}| < \varepsilon .$$

This proves the lemma for $\theta'\in R(m')$.

If $\theta'\notin R(m')$ then $\theta'_R = \theta'_{R(m')} + \theta'_T$ where $\Theta' = R(m') \dotplus T$ and $\theta'_T \neq 0$. If $\theta'_T(\theta) \equiv 0$ for $\theta\in N(M)$ then $\theta'_T\in N^{\perp}(M)$. By lemma 3.2.4 $\theta'\in R(M^t)$. Since $M\theta_1(\theta_2) = (m(\theta_1),m(\theta_2))_\xi$ and $M^t\theta_1(\theta_2) = M\theta_2(\theta_1) = (m(\theta_1),m(\theta_2))_\xi$ we observe that $M = M^t$ and hence that $R(M^t) = R(M)$. Since $R(M) \subset R(m')$, we have arrived at the contradictory conclusion that $\theta'_T\in R(m')$ unless θ_0 in $N(M)$ can be found for which $\theta'_T(\theta_0) \neq 0$. Fix $\theta_1\in M$ and let $\theta_n = \theta_0 + \frac{\theta_1}{n}$. Then $(M\theta_n)\theta_n = \frac{1}{n^2}(M\theta_1)\theta_1$ and $\theta'(\theta_n) = \theta'_{R(m')}(\theta_0) + \theta'_T(\theta_0) + \theta'\left(\frac{\theta_1}{n}\right)$. Since $\theta'_{R(m')} = m'h$ for some $h\in H(B)$ one has $\theta'_{R(m')}(\theta_0) = (h,m(\theta_0))_\xi$. The latter expression is zero by virtue of its being in $N(M)$ so

$$\frac{(\theta'(\theta_n))^2}{(M\theta_n)\theta_n} = n^2\left[\frac{(\theta'_T(\theta_0))^2}{(M\theta_1)\theta_1} + O(n^{-1})\right]$$

and the lemma has been proven. □

<u>Example 3.2.1</u>. With reference to the introduction, let the unknown variable force be $\{\theta(t): t\in[0,1]\}$. The observation is $\{Y(x,t): t\in[0,1]\}$ when mass $x\in[a,b]$ is used, where

$Y(t) = \int_0^1 (t-u)_+ \frac{\theta(u)}{x}\, du + N(t)$ and $N(t)$ is a zero mean process with covariance $K(s,t) = \int_0^1 (t-u)_+(s-u)_+du$. As usual $a_+ = \max\{0,a\}$. The quantity to be estimated is $\int_0^1 \theta(s)ds = \tau(\theta)$. The space $H(K)$ is the set of all functions f on [0,1] for which $f(0) = f'(0) = 0$ and $f''\epsilon L_2[0,1]$. The inner product is

$$(f,g)_K = \int_0^1 f''(s)g''(s)ds \ .$$

For any $\theta \neq 0$ in $\Theta = C[0,1]$ and $\xi\epsilon\Xi$

$$\begin{aligned}(M(\xi)\theta)\theta &= \sum\xi(x_j)\,||m_{x_j}(\theta)||_K^2 \\ &= \sum\frac{\xi(x_j)}{x_j^2}\int_0^1 \theta^2(s)ds > 0\end{aligned}$$

so

$$d(\tau,\xi) = \sup_{\theta\neq 0} \frac{\left(\int_0^1 \theta(s)ds\right)^2}{\sum_{j=1}^{r}\frac{\xi(x_j)}{x_j^2}\int_0^1 \theta^2(s)ds} .$$

From the Schwarz inequality one has $d(\tau,\xi) = \frac{1}{\sum\frac{\xi(x_j)}{x_j^2}}$ so the optimal design takes all observations at x = a.

Example 3.2.2. Assume the same conditions as in example 2.1 except that we wish to estimate $\theta(1/2)$. We shall show that $\theta(1/2)$ is not estimable with respect to any design $\xi\epsilon\Xi$.

First we observe that for any $\xi\epsilon\Xi$ and $\theta\epsilon C[0,1]$

$$||m_\xi(\theta)||_\xi^2 = \sum_{x\epsilon S(\xi)} \xi(x) \int_0^1 \frac{\theta^2(s)}{x^2} ds$$

$$\leq \left(\sum \frac{\xi(x)}{x^2} \; ||\theta||_\infty^2\right)$$

so that m_ξ is continuous from $C[0,1]$ to $H(K)$. Thus $m_\xi^!$ is actually the adjoint m_ξ^* mapping $H(K)$ into the topological dual of $C[0,1]$. Therefore, given $v\epsilon H(K)$ there is a function g of bounded variation on $[0,1]$ such that for all $\theta\epsilon C[0,1]$

$$(m_\xi^* v)\theta = \int_0^1 \theta(s)dg(s) \quad .$$

If $\theta(1/2)$ were estimable one would have $v\epsilon H(K)$ such that $m_\xi^* v$ corresponds to

$$g(s) = \begin{cases} 0 & s < 1/2 \\ 1 & 1/2 \leq s \end{cases} \quad . \qquad (3.10)$$

However, using the facts that

$$(m_\xi^* v)\theta = (v, m_\xi(\theta))\theta = \sum_{x\in S(\xi)} \xi(x) \int_0^1 v''(s)\, \frac{\theta(s)}{x}\, ds$$

$$= \left(\sum \frac{\xi(x)}{x}\right) \int_0^1 v''(s)\,\theta(s)\,ds \tag{3.11}$$

and $\int_0^1 (v''(s))^2 ds < \infty$ it is clear that (3.11) implies $m_\xi^* v$ can not correspond to the g in (3.10).

3.3. Characterization of optimal designs. The assumptions which at various times we shall have need of in this section are

(B1) The parameter space Θ is a linear topological space.

(B2) The mappings $\{m_x\}_{x\in X}$ from Θ to $H(K)$ are all linear and continuous and $\mathcal{R}(m_\xi)$ is closed for $\xi\in\Xi$.

Since (B2) implies m_ξ' has values in Θ^*, the topological dual of Θ, we shall write m_ξ^* rather than m_ξ'. Let F denote the set of all functions from X into the unit ball of H(K) and

$$\mathcal{R} = \{\int m_x^* \phi(x)\, d\xi(x) :\ \xi\in\Xi, \phi\in F\} .$$

(B3) For each $r_0 \in \partial\mathcal{R}$ (the boundary of $\mathcal{R}$) there is a $\theta_0\in\Theta$ such that $r_0(\theta_0) \geq r(\theta_0)$ for all $r\in\mathcal{R}$.

(B4) For each $\theta\in\Theta$, $\theta \neq 0$, $\sup_X ||m_x(\theta)|| > 0$.

The notation and assumptions may be simplified if instead the parameter space Θ is assumed to be a Hilbert space. The details appear in Spruill (1980). Fix $\theta^* \epsilon \Theta^*$ and let $v_0 = \inf_{\Xi} d(\theta^*, \xi)$.

Lemma 3.3.1. Let B1 - B4 hold. Then

(a) $\beta\theta^* \epsilon R$ implies $v_0 \leq \frac{1}{\beta^2}$, and

(b) $\beta\theta^* \epsilon \partial R$ implies $v_0 \geq \frac{1}{\beta^2}$.

Proof: (a) If $\beta = 0$ the result is trivially true. Otherwise

$$\beta\theta^* = \sum_{j=1}^{n} \alpha_j m^*_{x_j} \phi_{x_j} .$$

One has for any $\theta \epsilon \Theta$ that

$$\begin{aligned}[\theta^*(\theta)]^2 = \frac{[\beta\theta^*(\theta)]^2}{\beta^2} &= \frac{[\sum \alpha_j (\phi_{x_j}, m_{x_j}(\theta))_K]^2}{\beta^2} \\ &\leq \frac{\sum \alpha_j M_{x_j} \theta(\theta)}{\beta^2} \\ &= \frac{M(\xi)\theta(\theta)}{\beta^2} .\end{aligned}$$

Therefore, by lemma 2.7, $v_0 \leq \frac{1}{\beta^2}$.

(b) Since $\beta\theta^* \epsilon \partial R$ there is a $\theta \epsilon \Theta$ such that $\beta\theta^*(\theta) \geq |r(\theta)|$ for all $r \epsilon R$. This is a consequence of (B3). Let $x_n \epsilon X$ be such that $||m_{x_m}(\theta)|| \to \sup_X ||m_x(\theta)||$. Let $r_n = m^*_{x_n}\left(\frac{m_{x_n}(\theta)}{||m_{x_n}(\theta)||}\right)$

to obtain

$$\beta\theta^*(\theta) \geq |r_n(\theta)| = ||m_{x_n}(\theta)|| \ . \tag{3.12}$$

We conclude that $\beta\theta^*(\theta) \geq \sup_X ||m_x(\theta)||$. Also, since R is symmetric and convex, for $\varepsilon > 0$ sufficiently small $(\beta-\varepsilon)\theta^* \in R$ ($\beta > 0$ by (3.12) and (B4)). One has $(\beta-\varepsilon)\theta^* = \sum \alpha_j m^*_{x_j}(\phi_j)$ so

$$\beta\theta^*(\theta) - \varepsilon\theta^*(\theta) = \sum \alpha_j (\phi_j, m_{x_j}(\theta))$$

$$\leq \sum \alpha_j \ ||m_{x_j}(\theta)|| \leq \sup_X ||m_x(\theta)|| \ . \tag{3.13}$$

Combining (3.12) and (3.13) we have

$$\beta\theta^*(\theta) = \sup_X ||m_x(\theta)|| \ . \tag{3.14}$$

Let ξ_n be such that $d(\theta^*,\xi_n) \to v_0$. If v_0 is $+\infty$ the result is trivially true. Assume, therefore that $\theta^* \in R(m^*_{\xi_n})$ for each n.

Then θ must be such that $(M(\xi_n)\theta)\theta > 0$ for all n. To see this, suppose that it is not for some n. Then $\theta \in N[M(\xi_n)] = R^{\perp}[M^t(\xi_n)]$, and since $M(\xi_n) = M^t(\xi_n)$ (see lemma (3.2.5) one has $\theta \in R^{\perp}[M(\xi_n)]$.

From the argument in lemma (3.2.5) there is a sequence $\theta_k^* \epsilon R[M(\xi_n)]$ such that $\lim_{k\to\infty} |d(\theta_k^*,\xi_n)-d(\theta^*,\xi_n)| = 0$. But $\theta_k^*(\theta) \equiv 0$ by definition of $R^{\perp}[M(\xi_n)]$ so that $\theta^*(\theta) = 0$. This, with (3.14) contradicts (B4). Since $(M(\xi_n)\theta)\theta > 0$ for all n one has

$$\frac{1}{\beta^2} = \frac{[\theta^*(\theta)]^2}{\sup_X ||m_x(\theta)||_K^2} \le \frac{[\theta^*(\theta)]^2}{(M(\xi_n)\theta)\theta} \le d(\theta^*,\xi_n). \quad \Box$$

<u>Theorem 3.3.1</u>. Let $\theta^*(\theta)$ be estimable with respect to some design and B1 - B4 hold. The design ξ_0 is optimal for the estimation of $\theta^*(\theta)$ if and only if there is a function ϕ: $X \to H(K)$ such that $||\phi(x)|| \equiv 1$ and

$$\int m_x^* \phi(x) d\xi_0(x)$$

is (i) proportional to θ^* and (ii) in $R \cap \partial R$.

<u>Proof</u>: Suppose ξ_0 is optimal. Then setting $\lambda_0 = M^{\#}(\xi_0)\theta^*$ one has $M_0\lambda_0 = \theta^*$ and $M_0\lambda_0(\lambda_0) = v_0$. Let $\{z: \lambda(z-v_0^{-1/2}\theta^*) = 0\}$ be a supporting hyperplane to R at $v_0^{-1/2}\theta^*$ where $\lambda \neq 0$. From above $\theta^*(v_0^{-1/2}\lambda) = \sup_X ||m_x(\lambda)||$. Since $M_0\lambda_0 = \theta^*$ one has $v_0^{-1/2}M_0\lambda_0(\lambda) = \sup_X ||m_x(\lambda)||$. Therefore

$$\sup\|m_x(\lambda)\|^2 = v_0^{-1}[(M_0\lambda_0)\lambda]^2 \tag{3.15}$$

$$\leq v_0^{-1}[M_0(M_0^{\#}\theta^*)M_0^{\#}\theta^*][(M_0\lambda)\lambda]$$

$$= v_0^{-1}[\theta^* M_0^{\#}\theta^*][(M_0\lambda)\lambda] = (M_0\lambda)\lambda$$

with strict inequality unless $m_{\xi_0}(\lambda_0) = km\xi_0(\lambda)$ for some scalar k. Since one always has $(M_0\lambda)\lambda \leq \sup_X\|m_x(\lambda)\|^2$ equality holds in (3.15) and we have $\|m_x(\lambda)\| = \sup_X\|m_x(\lambda)\|$ on the support $S(\xi_0)$ of ξ_0. Set $\phi(x) = \dfrac{m_x(\lambda)}{\|m_x(\lambda)\|}$ for $x\epsilon S(\xi_0)$. Then

$$\int m_x^*\phi(x)d\xi_0(x) = \frac{M_0\lambda}{\sup\|m_x(\lambda)\|} = \frac{M_0\lambda_0}{k\ \sup\|m_x(\lambda)\|} .$$

From above

$$v_0 = M_0\lambda_0(\lambda_0) = k\ M_0\lambda(\lambda_0) = k\ \sqrt{M_0\lambda_0(\lambda_0)}\ \sqrt{M_0\lambda(\lambda)}$$

$$= k\ v_0^{1/2}\ \sup_X\|m_x(\lambda)\| ,$$

so that $\int m_x^*\phi(x)d\xi_0(x) = v_0^{-1/2}\theta^*$.

If (i) and (ii) hold then for any $\theta\epsilon\Theta$

$$v_0^{-1}[\theta^*(\theta)]^2 = (\int(m_x(\theta),\phi(x)d\xi_0(x))^2$$

$$\leq \int \|m_x(\theta)\|^2 d\xi_0(x) = (M_0\theta)\theta$$

showing ξ_0 to be optimal. □

Another theorem of interest whose proof we leave to the reader (see Studden and Tsay (1976)), is the following. Let $\theta_0^* \epsilon \Theta^*$ be fixed, Θ be a reflexive Banach space, and $\Delta = \{\delta \epsilon \Theta: \theta_0^*(\delta) = 1\}$.

<u>Theorem 3.3.2</u>. Under B1 - B4, if $\delta_0 \epsilon \Delta$ satisfies

(i) $\inf_{\Delta} \sup_{X} ||m_x(\delta)||_K = \sup_{X} ||m_x(\delta_0)||$,

(ii) $S(\xi_0) \subset \{x: ||m_x(\delta_0)|| = \sup_{X} ||m_x(\delta_0)||\}$,

and (iii) $\int m_x^* m_x(\delta_0) d\xi_0(x)$ is proportional to θ_0^*

then ξ_0 is optimal for estimating θ_0^* and $v_0 = (\sup_{X} ||m_x(\delta_0)||)^{-2}$. Furthermore, if there is an optimal design ξ_0 , then there is a δ_0 satisfying (i), (ii), and (iii).

Some other theorems of possible interest may be obtained from section 2 above. Their statements and proofs require certain additional assumptions and modifications along the lines of those presented above.

3.4. <u>Verifying the assumptions</u>. Verification of assumptions B1 - B4 may be a non-trivial task. We present herein some sufficient conditions which relate especially to (B3), the most difficult to verify. The assumption (B3) is trivial if $\Theta = R^n$ for some n. We introduce some terminology.

The point a is said to be an internal (not necessarily interior) point of the set $A \subset \Theta^*$ if $a \in A$ and for each $\theta^* \in \Theta^*$ there is an $\alpha_0 > 0$ such that $\alpha\theta^* + (1-\alpha)a \in A$ for $0 \leq \alpha < \alpha_0$. Since $\bar{R}$ is symmetric, $\bar{R}$ has an internal point if and only if 0 is an internal point. Furthermore, since Θ^* is a Banach space, an application of the Baire category theorem shows that $\bar{R}$ has an internal point if and only if it has an interior point.

The next result may be found in Holmes (1975).

Lemma 3.4.1. If $A \subset \Theta^*$ contains an interior point and is convex then at each point $b \in \partial A$ there is a non-zero element $\theta_b^{**} \in \Theta^{**}$ such that

$$\theta_b^{**}(b) \geq \theta_b^{**}(a) \quad \text{for all} \quad a \in A .$$

We prove the following.

Theorem 3.4.1. If Θ is a reflexive Banach space, $\sup_X ||m_x(\theta)|| \geq k||\theta||$ for some $k > 0$, and $\{m_x\}_{x \in X}$ are continuous then (B3) holds.

Proof: It suffices to prove that 0 is an internal point of $\bar{R}$. Let $\theta_0^* \in \Theta^*$ be non-zero and suppose that $\alpha\theta_0^*$ is not in $\bar{R}$. Then there is a $\lambda_\alpha \in \Theta$ such that $||\lambda_\alpha|| = 1$, and $\alpha\theta_0^*(\lambda_\alpha) \geq a(\alpha) > b(\alpha) \geq r(\lambda_\alpha)$ for all $r \in \bar{R}$ (see Dunford and Schwarz (1959), V. 2.11). Therefore,

$$\alpha||\theta_0^*|| \geq \alpha\theta_0^*(\lambda_\alpha) > ||m_x(\lambda_\alpha)|| \tag{3.16}$$

as in seen by choosing $r = m_x^*\left(\frac{m_x(\lambda_\alpha)}{||m_x(\lambda_\alpha)||}\right)$, for some x which makes $||m_x(\lambda_\alpha)|| > 0$. In view of our assumptions, (3.16) implies 0 is an internal point of $\bar{R}$.

Lemma 3.4.2. If (B2) holds and $\theta^*(\theta)$ is estimable for each $\theta^*\epsilon\Theta^*$ then 0 is an internal point of $\bar{R}$.

Proof: Let $\theta^* \neq 0$. Then, since $\theta^*(\theta)$ is estimable there is a $\xi\epsilon\Xi$ such that $\theta^* = m_\xi^*h$ for some $h\epsilon H_\xi$. Since $R^\perp(m_\xi) \subset N(m_\xi^*)$ and $R(m_\xi)$ is closed there is a $\theta_0\epsilon\Theta$ such that $\theta^* = m_\xi^*m_\xi(\theta_0)$ or

$$\theta^* = \sum_{x\epsilon S(\xi)} \xi(x)m_x^*m_x(\theta_0) \quad . \tag{3.17}$$

Since $\max_{x\epsilon S(\xi)} ||m_x(\theta_0)|| > 0$, the result follows from (3.17) upon dividing both sides by $\max_{x\epsilon S(\xi)} ||m_x(\theta_0)||$.

Lemma 3.4.3. If $\theta^*(\theta)$ is estimable for each $\theta^*\epsilon\Theta^*$ and (B2) holds then (B4) holds.

Proof: Suppose $\theta_0 \neq 0$ and $\sup_X||m_x(\theta_0)|| = 0$. Let θ_0^* be such that $||\theta_0|| = \theta_0^*(\theta_0)$. Since $\theta_0^*(\theta)$ is estimable an argument similar to that in lemma 3.4.2 shows that there is a sequence $\{\theta_n\} \subset \Theta$ such that

$$\lim_{n\to\infty} \left|\left|\theta_0^* - \sum_{x\in S(\xi)} \xi(x) m_x^* m_x(\theta_n)\right|\right| = 0 .$$

But then

$$||\theta_0|| = \theta_0^*(\theta_0) = \lim_{n\to\infty} \sum \xi(x)(m_x(\theta_n), m_x(\theta_0)) = 0 .$$

This contradiction establishes the lemma. We have proved the following.

<u>Theorem 3.4.2</u>. If Θ is a reflexive Banach space, (B2) holds and $\theta^*(\theta)$ is estimable for each $\theta^* \epsilon \Theta^*$ then (B3) and (B4) hold.

<u>Example 3.4.1</u>. Let the observed process be $Y(x,t) = \int_0^1 (t-u)_+ \theta(u) x(u) du + N(t)$, $t \epsilon [0,1]$ where $N(t)$ is as in example 3.2.1 above and the function x is of the form

$$x(u) = \begin{cases} 0 & 0 \le u < \alpha \\ 1 & \alpha \le u \le 1 \end{cases} . \qquad (3.18)$$

Let X consist of all functions of the form (3.18) for some $\alpha \epsilon [0,1]$. We suppose that $\Theta = L_2[0,1]$ and that we are to optimally estimate $\tau(\theta) = \int_0^1 (s-\varepsilon)_+ \theta(s) ds$. We shall employ theorem 3.3.2 to solve this problem.

To verify that B1 - B4 hold observe that the hypotheses of Theorem 3.4.1 hold (and therefore B4 holds also) with $k = 1$.

Let δ satisfy

$$1 = \tau(\delta) = \int_0^1 \delta(s)(s-\varepsilon)_+ ds . \qquad (3.19)$$

We seek to minimize

$$\sup_X ||m_x(\delta)||^2 = \sup_X \int_0^1 x^2(s)\delta^2(s)ds = \int_0^1 \delta^2(s)ds$$

subject to (3.19). From Schwarz's inequality

$$\int_0^1 \delta^2(s)ds \geq \frac{1}{\int_0^1 (s-\varepsilon)_+^2 ds}$$

with equality if and only if $\delta(s) = k(s-\varepsilon)_+$. Therefore, any design which concentrates all its mass at x's satisfying $x(\varepsilon) = 1$ is optimal and $v = \frac{(1-\varepsilon)^3}{3}$.

<u>Example 3.4.2</u>. Let

$$Y(x,t) = \int_0^1 \theta(s)e^{-xs}ds + W(t), \; t\in[0,1]$$

where $x\in[a,b]$ and $W(t)$ is the Wiener process. We suppose that $\Theta = L_2[0,1]$ and that we wish to estimate

$$\int_0^{t_0} \theta(s)e^{-ys}ds$$

for some $t_0 \in (0,1]$, and some $y \in R$. We note that $H(K)$ consists of all functions f on $[0,1]$ for which $f(0) = 0$ and $f' \in L_2[0,1]$, with

$$(f,g)_K = \int_0^1 f'(s)g'(s)ds \ .$$

Since L_2 is reflexive and $\int_0^1 (\chi_{[0,t_0]}(s)\ e^{-ys})^2 ds < \infty$ the function τ, $\tau(\theta) = \int_0^{t_0} \theta(s)e^{-ys}ds$, is in L_2^*. The map m_x: $L_2 \to H(K)$ is continuous since

$$||m_x(\theta)||^2 = \int_0^1 \theta^2(s)e^{-2xs}ds \leq \int_0^1 \theta^2(s)e^{-2as}ds$$
$$\leq e^{-2a}||\theta||_2^2 \ .$$

Also, the conditions of theorem 3.4.1 are satisfied since

$$\sup_{a \leq x \leq b} ||m_x(\theta)||^2 \geq e^{-2b}||\theta||_2^2 \ .$$

Therefore, we may apply theorem 3.3.2. Let

$$1 = \tau(\delta) = \int_0^1 \chi_{[0,t_0]}(s)\ e^{-ys}\delta(s)ds \ . \qquad (3.20)$$

We shall minimize

$$\sup_{a \le x \le b} \int_0^1 \delta^2(s) e^{-2xs} ds = \int_0^1 \delta^2(s) e^{-2as} ds$$

subject to (3.20). We have, by Schwarz,

$$1 = \left(\int_0^1 \delta(s) c^{-ys} \chi_{[0,t_0]}(x)\, ds \right)^2 \le \int_0^1 \delta^2(s) e^{-2as} ds \int_0^1 e^{2(a-y)s} \chi_{[0,t_0]}(x)\, ds$$

with equality if and only if

$$\delta(s) e^{-as} = k e^{as} e^{-ys} \chi_{[0,t_0]}(s) \quad .$$

Therefore $\delta_0(s) = \left(\dfrac{2(a-y)}{e^{2(a-y)t_0} - 1} \right) e^{(2a-y)s} \chi_{[0,t_0]}(x)$

and

$$\sup_{a \le X \le b} ||m_x(\delta_0)||^2 = \frac{2(a-y)}{e^{2(a-y)t_0} - 1} \quad .$$

To find m_x^* note that for $v \epsilon H(K)$ and λ in L_2

$$m_x^* v(\lambda) = \int_0^1 m_x^* v(s) \lambda(s) ds$$

and $m_x^* v(\lambda) = (v, m_x(\lambda))_K = \int_0^1 v'(s) \lambda(s) e^{-xs} ds$

so m_x^* corresponds to $v'(s)e^{-xs}$. The support of any optimal design is x = a so we must have $m_a^* m_a(\delta_0)$ proportional to τ. We see that

$$m_a^* m_a(\delta_0) = m_a(\delta_0)'(s)e^{-as}$$

$$= \delta_0(s)e^{-2as} = \frac{2(a-y)}{e^{2(a-y)t_0}-1} \chi_{[0,t_0]}(x)\ e^{-ys} .$$

Therefore, the design which places mass one at x = a is optimal and $v = \frac{e^{2(a-y)t_0}-1}{2(a-y)}$.

Example 3.4.3. Let $X = [0,2\pi]$, $\Theta = L_2[0,1]$, and

$$Y(x,t) = \int_0^1 \theta(s) \sin(2\pi s-x)ds + W(t)$$

for $t \in [0,1]$. Since

$$||m_x(\theta)||^2 = \int_0^1 \theta^2(t) \sin^2(2\pi t-x)dt \le ||\theta||_2^2$$

(B1) and (B2) are satisfied. Let $\phi(x) = \frac{1}{2\pi} \chi_{[0,2\pi]}(x)$. Then

$$\sup_{0\le x\le 2\tau} ||m_x(\theta)||^2 \ge \int_0^{2\pi} ||m_x(\theta)||^2 \phi(x)dx = \frac{||\theta||_2^2}{2}$$

so (B3) and (B4) are satisfied. We shall use Theorem 3.3.2 to solve the design problem for estimating $\int_0^1 \theta(s)ds$. If we view

X as the parameter space, [0,1] as the outcome space, $\mathcal{R}$ as the action space, and $\theta(\cdot)$ a decision rule in a statistical decision problem then $\theta_0(s) \equiv 1$ results in a constant risk of $\frac{1}{2}$. Viewing ϕ as a prior on the parameter space, it is easily seen that θ_0 is Bayes among those $\theta \in L_2$ such that $\int_0^1 \theta(s)ds = 1$. That is; it is minimax,

$$\inf_{\{\theta:\int_0^1\theta(s)ds=1\}} \sup_{0\le x\le 2\pi} ||m_x(\theta)||^2 = \sup_{0\le x\le 2\pi} ||m_x(\theta_0)||^2 = \frac{1}{2}$$

It remains to find a design for which

$$[\sum \xi(x) m_x^* m_x(\theta_0)](t) \equiv 1 . \tag{3.21}$$

Since $m_x^* m_x(\theta_0)(t) = \theta_0(t) \sin^2(2\pi t - x)$ the design which places masses 1/2 at $x = 0$ and 1/2 at $x = \pi/2$ satisfies (3.21) and is optimal for estimating $\int_0^1 \theta(s)ds$ with $v = 2$.

<u>Example 3.4.4.</u> Let $\Theta = L_2[0,1]$ and $Y(x,t) = \int_0^t \theta(s)x(s)ds + W(t)$, $t \in [0,1]$, where x is any function of the form

$$x(s) = as + b, \ |a| \le A, \quad |b| \le B, \ AB > 0 .$$

Since

$$||m_x(\theta)||^2 = \int_0^1 \theta^2(s)x^2(s)ds \le (A+B)^2||\theta||_2^2$$

and $\sup_x ||m_x(\theta)||^2 \ge B^2||\theta||_2^2$, B1 - B4 are satisfied. We are to find the optimal design for estimating $\int_0^1 \theta(s)ds$ so we seek to minimize

$$\sup_x ||m_x(\theta)||^2 = \int_0^1 \lambda^2(s)(As+B)^2 ds$$

among all $\lambda \epsilon L_2$ such that $\int_0^1 \lambda(s)ds = 1$. Using $\phi(s) = As + B$

$$1 = (\int\lambda)^2 = (\int \frac{\lambda}{\phi}\phi)^2 \le \int\lambda^2\phi^2\int\phi^{-2}$$

we see that $\lambda_0(s) = \dfrac{B(A+B)}{(As+B)^2}$ satisfies

$$\inf_{\int\lambda=1} \sup_X ||m_x(\lambda)||^2 = \sup_X ||m_x(\lambda_0)||^2 = B(A+B) .$$

To complete the solution to the problem we must find a design on the two functions $x_1(s) = -As - B$ and $x_2(s) = As + B$ such that

$$\alpha m_{x_1}^* m_{x_1}(\lambda_0) + (1-\alpha)m_{x_2}^* m_{x_2}(\lambda_0) \tag{3.22}$$

is constant. Since

$$\lambda_0(s)[\alpha x_1^2(s) + (1-\alpha)x_2^2(s)] = B(A+B)$$

and design which distributes all its mass between x_1 and x_2 is optimal with $v = [B(A+B)]^{-1}$.

Finally we provide some examples of how estimable functions may fail to have an associated optimal design.

Example 3.4.5. Let $\Theta = \ell_2$, T be a one point set, and $X \subset \ell_2$. Suppose $Y(x) = (\theta,x) + \varepsilon$ and X is the open unit sphere $\{||x|| < 1\}$. Since $m_x^*: R \to \ell_2$ satisfies $m_x^* u = ux$ one finds that $R = \{\pm x: x \in X\}$. Since $R \cap \partial R = \phi$ there can be no optimal designs even though every linear form is estimable.

The non-existence of optimal designs in the preceding example could be attributed to the fact that $||m_x(\theta)||$ failed to attain its supremum on X. Even in a finite dimensional space Θ an estimable function τ may fail to be optimally estimable although $||m_x(\theta_0)||$ attains its maximum on X, where

$\max_X ||m_x(\theta_0)|| = \inf_{\tau(\theta)=1} \sup_X ||m_x(\theta)||$. If however, Θ is finite dimensional and $\{m_x^*\phi(x): ||\phi(x)|| = 1, x\in X\}$ is closed and bounded, then there is an optimal design for each estimable function. The next example shows that the latter is not the case if Θ is not finite dimensional.

Example 3.4.6. As in example 3.4.5, let $\Theta = \ell_2$ and $m_x(\theta) = (x,\theta)$. Let $\{\Phi_0,\Phi_1,\cdots\}$ be a complete orthonormal basis for ℓ_2 and

$$X = \{\pm(\Phi_0+\Phi_j): j \geq 1\} \cup \left\{\frac{\pm\Phi_0}{2}\right\}.$$

Since $m_x^* u = ux$ one finds that $R = co(X)$. We observe that (Φ_0, θ) is estimable since $m^*_{\Phi_{0/2}}(2) = \Phi_0$. Also $\Phi_0 \epsilon \bar{R}$ since taking $\alpha_{n_j} = \frac{1}{n}$, $j = 1,\cdots,n$ one has

$$||\Phi_0 + \sum_{j=1}^{n} \alpha_{n_j}\Phi_j - \Phi_0||^2 = \frac{1}{n} \to 0 .$$

The fact that for no $\xi \epsilon \Xi$ or $|\varepsilon_j| \equiv 1$ is it true that $\Phi_0 = \sum_{j\epsilon J} \varepsilon_j \xi_j (\Phi_0+\Phi_j)$ implies at once that $\Phi_0 \epsilon \partial R$ but not in R.

Therefore, there is no optimal design for estimating (Φ_0, Θ).

4. Application to random differential equations.

4.1. Introduction. Suppose that $\{\theta(t)\colon t\in[0,1]\}$ is a variable force and that $\int_0^1 \theta(t)dt$ is to be measured but that any action of this force is contaminated by white noise. Three possible experiments to accomplish this task are:

1) Observe the motion $\{Y_1(t)\colon t\in[0,1]\}$ of a particle of mass m initially at rest at zero under the influence of $\{\theta(t) + \sigma W'(t)\colon t\in[0,1]\}$.

2) Observe the motion $\{Y_2(t)\colon t\in[0,1]\}$ of a particle of mass m initially at rest at 0 under the influence of $\{\theta(t) + \sigma W'(t)\colon t\in[0,1]\}$ and a restoring Hook's law spring with constant k.

3) Observe $\{Y_3(t)\colon t\in[0,1]\}$ subject to $\{\theta(t) + \sigma W'(t)\colon t\in[0,1]\}$ and a frictional force $-fY'(t)$.

In each of these cases it is possible to compute the variance of the uniform minimum variance unbiased estimator $\int_0^1 \theta(t)dt$ as discussed above. Which of the experiments described above yields the smallest such variance?

Before answering the question let us rephrase it in a form which will allow a more general answer. Notice that the observable processes are respectively the solutions to the following three random differential equations, $t\in[0,1]$,

$$mY_1''(t) = \theta(t) + \sigma W'(t) \ , \tag{4.1}$$

$$Y_1(0) = Y_1'(0) = 0 \ ;$$

$$mY_2''(t) + kY(t) = \theta(t) + \sigma W'(t) \ , \tag{4.2}$$

$$Y_2(0) = Y_2'(0) = 0 \ ;$$

$$mY_3''(t) + fY_3'(t) = \theta(t) + \sigma W'(t) \ , \tag{4.3}$$

$$Y_3(0) = Y_3'(0) = 0 \ .$$

More generally the output $Y(t)$ of a measuring instrument (see Doebelin (1975)) satisfies an equation of the form

$$(D^n + a_{n-1}D^{n-1} + \cdots + aI)Y(t) = \theta(t) + \sigma W'(t)$$

subject to $Y^{(0)}(0) = Y^{(1)}(0) = \cdots = Y^{(n-1)}(0) = 0$. The questions in this generality become:

a) Is there a best selection of coefficients $a_0, \cdots, a_{n-1}$ and an integer $n \geq 1$?

b) What are the values of these parameters if there is a best selection?

The answer can be given from the more general results given below in theorem 4.2.1. All such experiments yield the same variance. In fact, if $\int_0^1 u(s)\theta(s)ds$ is to be estimated, where θ and u are in $L_2[0,1]$, then the variance of the umvlue is just

$\sigma^2\int_0^1 u^2(t)dt$ regardless of $n \geq 1$ and the coefficients $a_0,\cdots,a_{n-1}$. Some implications of this result for the design of measurement systems are discussed in section 4.3 below.

Before describing the optimum design problem which motivates our current efforts we need some notation. Let $[a,b]$ be an arbitrary closed bounded subinterval of R, n an integer greater than or equal to one, and $\{c_j(t)\}_{j=0}^{n-1}$ be n-times continuously differentiable on $[a,b]$. As in Kimeldorf and Wahba (1970) $H_L[a,b]$ is the set of real-valued functions on $[a,b]$ for which $D^{n-1}f$ is absolutely continuous and Lf is in $L_2[a,b]$, where L is the differential operator defined by

$$L = D^n + c_{n-1}D^{n-1} + \cdots + c_0 I \ . \tag{4.4}$$

Let $\{z_i\}_{i=0}^{n-1} \subset H$ satisfy $z_i^{(j)}(a) = \delta_{ij}$, $j = 0,\cdots,n-1$ and $Lz = 0$. Denote the Green's function of L by $G(s,t)$. That is, $G(s,t) = 0$, $s < t$, $L_{(s)}G(s,t) = 0$, $s \geq t$, and $\left.\frac{\partial^j G(s,t)}{\partial s^j}\right/_{s=t} = \delta_{j,n-1}$, $j=0,\cdots,n-1$.

<u>Lemma 4.1.1</u>: (Kimeldorf and Wahba). If $\{Y(t): t\in[0,1]\}$ satisfies $LY(t) = F(t) + \sigma W'(t)$ subject to $Y^{(j)}(a) = Y_j$, $j = 0,\cdots,n-1$, where $Y_0,\cdots,Y_{n-1}$, W are uncorrelated and $F\in L_2[a,b]$ then

$$Y(t) = m(t) + N(t) \ , \ t\in[a,b]$$

where

$$m(t) = \sum_{j=0}^{n-1} \mu_j z_j(t) + \int_a^b G(t,s)F(s)ds ,$$

$$N(t) = \sum_{j=0}^{n-1} (Y_j - \mu_j) z_j(t) + \sigma \int_a^b G(t,s)dw(s) ,$$

and $\mu_j = E(Y_j)$. If $\sigma_j^2 = \mathrm{Var}(Y_j) > 0$ for all $j = 0,\cdots,n-1$ the reproducing kernel Hilbert space generated by $K(s,t) = \mathrm{cov}(N(s),N(t))$ consists of all functions f on [a,b] of the form

$$f(t) = \sum_{j=0}^{n-1} \alpha_j z_j(t) + \int_a^b G(t,s)u(s)ds$$

such that $u \in L_2[a,b]$ and the α's are real scalars. If

$$g(t) = \sum_{j=0}^{n-1} \beta_j z_j(t) + \int_a^b G(t,s)v(s)ds$$

the Hilbert space innner product of f and g is given by

$$(f,g)_K = \sum_{j=0}^{n-1} \frac{\alpha_j \beta_j}{\sigma_j^2} + \frac{1}{\sigma^2} \int_a^b u(s)v(s)ds .$$

Proof: As indicated, the proof follows from Kimeldorf and Wahba (1970). □

We remark here that the formulas continue to hold for $\sigma_j^2 = 0$ if $\mu_j = 0$ also. Just omit those indices from the summation.

A general description of the type of experimental situation to which our techniques apply is as follows (see section 5 for generalizations). For each x in a set X of possible experiments an experiment can be performed whose outcome is a stochastic process $\{Y(x,t): a \le t \le b\}$. It is assumed that the stochastic process $Y(x,t)$ is the solution to

$$L_x Y(x,t) = F(x,\theta,t) + \sigma_x W'(t)$$

on $[a,b]$, where $L_x \in \mathcal{L}$, the collection of all differential operators satisfying (4.4) for some $n \ge 1$, and the initial conditions are given by

$$\frac{d^j}{dt^j} Y(x,t)\Big/_{t=a} = Y_j(x) \quad , \; j = 0,\cdots,n-1 \; .$$

The Wiener process W and the random variables Y_j are mutually uncorrelated and $E[Y_j(x)] \equiv 0$ for $j = 0,\cdots,n-1$. The function $F(x,\theta,\cdot)$ in $L_2[a,b]$ is of known form and is linear in the unknown parameter θ. The parameter θ is an element of a linear, but otherwise arbitrary space Θ. The value $\tau(\theta)$, where τ is a linear functional on Θ, is to be estimated on the basis of N uncorrelated observations $\{Y(x_i,t): t \in [a,b],\ i = 1,\cdots,N\}$. The problem is to find the experimental design which minimizes the variance of the minimum variance unbiased estimator of $\tau(\theta)$.

Denoting by Ξ the set of probability measures ξ on X whose supports $S(\xi)$ consist of subsets of X with a finite number of elements we have the following theorem for the differential equation data problem.

<u>Theorem 4.1.1</u>. The deisgn $\xi_0 \in \Xi$ is optimal for estimating $\tau(\theta)$ if and only if

$$\sup_{\theta \in U(\xi_0)} \sum_{x \in S(\xi_0)} \frac{[\tau(\theta)]^2}{\frac{\xi_0(x)}{\sigma_x^2} \int_a^b F^2(x,\theta,t)\,dt} \le \sup_{\theta \in U(\xi)} \sum_{S(\xi)} \frac{[\tau(\theta)]^2}{\frac{\xi(x)}{\sigma_x^2} \int_a^b F^2(x,\theta,t)\,dt}$$

for all $\xi \in \Xi$, where

$$U(\xi) = \{\theta: \sum \frac{\xi(x)}{\sigma_x^2} \int_a^b F^2(x,\theta,t)\,dt > 0\} .$$

<u>Proof</u>: By lemma 4.1.1 above it follows that the solutions to the random differential equations $\{Y(x_i,t): t \in [a,b]\}$ may be expressed in the form

$$Y(x,t) = m(x,\theta,t) + N(t) \;, \; t \in [a,b]$$

where m and N satisfy all the assumptions of the general means regression design problem except that the covariance kernel depends upon x. Modification of the methods in section 3 to cover this case results in the following version of lemma 3.2.5. The function $d(\tau,\xi)$ satisfies

$$d(\tau,\xi) = \sup_{\theta \in U(\xi)} \frac{[\tau(\theta)]^2}{\sum \xi(x) ||m(x,\theta)||^2_{K_x}} \;, \text{ where}$$

$$U(\xi) = \{\theta: \sum \xi(x) ||m(x,\theta)||^2_{K_x} > 0\} .$$

The definition of optimal design as that design in Ξ which minimizes $d(\tau,\xi)$, and lemma 4.1.1 above which shows that

$$||m(x,\theta)||^2_{K_x} = \sum_{j=0}^{n(x)-1} \frac{\mu_j^2}{\sigma_j^2} + \frac{1}{\sigma_x^2}\int_a^b F^2(x,\theta,t)dt$$

$$= \frac{1}{\sigma_x^2}\int_a^b F^2(x,\theta,t)dt$$

demonstrates the validity of the assertion. $\square$

In summary, we have shown that the design problem with data from random differential equations may be analyzed in the framework of the general linear means regression design problem when the forcing function is linear in the unknown parameter. The function $d(\tau,\xi)$ defined by

$$d(\tau,\xi) = \sup_{\theta\epsilon U} \frac{[\tau(\theta)]^2}{\sum \frac{\xi(x)}{\sigma_x^2}\int_a^b F^2(x,\theta,t)dt} \tag{4.5}$$

where $U = \{\theta\epsilon\Theta: \sum \frac{\xi(x)}{\sigma_x^2}\int_a^b F^2(x,\theta,t)dt > 0\}$ determines the optimality of the deisgn $\xi \epsilon \Xi$. The optimal design $\xi_0 \epsilon \Xi$, if one exists, must satisfy $d(\tau,\xi_0) = \inf_{\xi\epsilon\Xi} d(\tau,\xi)$.

4.2. <u>Application to the design of measuring instruments.</u> In this section we shall identify measuring instruments with the differential operators L of section 4.1. That is, we shall identify a measuring instrument by the differential operator $L \in \mathcal{L}$ which describes its dynamic response. Not all measuring instruments are covered, but many are. See Doebelin (1975) for details. We shall identify the process of designing a measuring instrument with the choice of an order n of its differential operator and the selection of its (possibly variable, but throughout this work deterministic) coefficients. More precisely, we shall be interested in the optimal collection of instruments $\{L_1, \cdots, L_N\} \in \mathcal{L}$ to minimize the variance of the best unbiased estimator of a linear functional of the forcing function. We assume that:

a) the unknown forcing function θ is an element of $L_2[0,1]$.

b) we observe $\{Y_1, \cdots, Y_N\}$ where Y_j solves

$$L_j Y_j = \theta + \sigma W_j'$$

subject to $Y_j^{(i)}(0) = 0$, $i = 0, \cdots, n_j - 1$, where n_j is the order of L_j and $\{W_1, \cdots, W_N\}$ are uncorrelated.

c) The value $\tau(\theta) = \int_0^1 \theta(t)u(t)dt$, $u \in L_2[0,1]$, is to be estimated using the best unbiased estimator.

This corresponds to the problem described in section 4.1 above where $\mathcal{L} = X$, $F(x,\theta) = \theta$, and $\sigma_x \equiv \sigma > 0$.

Theorem 4.2.1. Every measuring instrument $L \in \mathcal{L}$ yields the same variance of the best unbiased estimator of $\tau(\theta)$. This variance is $\sigma^2 ||\tau||^2 \equiv d(\tau, \xi)$ for all designs ξ.

Proof: For any $\xi \in \Xi$ equation (5) shows

$$d(\tau,\xi) = \sup_{\theta\neq 0} \frac{[\tau(\theta)]^2}{\sum \frac{\xi(x)^2}{\sigma^2} \int_0 \theta^2(t)\,dt} = \sup_{\theta\neq 0} \left[\frac{|\tau\ \theta|}{||\theta||} \right]^2 \sigma^2$$

$$= \left[\sup_{\theta\neq 0} \frac{|\tau(\theta)|}{||\theta||} \right]^2 \sigma^2 = ||\tau||^2 \sigma^2 .$$

This shows that by our (very narrow) criterion all instruments of order at least one yield the same precision.

Even though the variances of the best unbiased estimators are the same, the form of the best estimator as a function of the observable process $\{Y_i(t): t \in [0,1],\ i = 1,\cdots,N\}$ depends upon the differential operator L_i, $i = 1,\cdots,N$. Since the form of the measuring instrument does not affect the precision we may attempt to design the instrument so that the final estimator is as simple a functional of the observable process as can be achieved, or perhaps as numerically stable as possible. We do not attempt here to investigate these questions any further than to offer theorem 4.2.2 below.

We first give some requisite lemmas.

Lemma 4.2.1. The Green's functions $G(t,s)$ of the constant coefficient differential operator

$$L = D^n + a_{n-1}D^{n-1} + \cdots + a_0 I$$

defined on $H_L[0,1]$ and the covariance kernel (see Kimeldorf and Wahba (1970)

$$K(s,t) = \sigma^2 \int_0^1 G(s,u)G(t,u)du$$

of the process $\{Y(t): t \in [0,1]\}$ satisfying $LY = W'$, $Y^{(i)}(0) = 0$, $i = 0,\cdots,n-1$, satisfy

a) $\frac{\partial^j}{\partial t^j} K(s,t) \in H(K)$ for $j = 0,1,\cdots,n-1$,

b) $\frac{\partial^j}{\partial t^j} K(s,t) = \sigma^2 \int_0^1 G(s,u) \frac{\partial^j}{\partial v^j} G(v,u)\Big/_{v=t} du$, $j=0,\cdots,n-1$,

and

c) $\frac{\partial^j}{\partial t^j} G(t,s)$ for $j = 0,1,\cdots,n-1$,

is a continuous function in the supremum norm on $[0,1]\times[0,1]$ of the coefficients $(a_0,\cdots,a_{n-1}) \in R^n$.

Proof: Part (a) follows from Wahba (1974). Part (b) is a consequence of the proof of theorem 3.4.2 of Naimark (1967). Finally, in the case of constant coefficients the form of G as a function of $a_0,\cdots,a_{m-1}$ can be used to show (c).

In the next lemma denote by L_2 the space of complex valued Lebesgue measurable functions g on $(-\infty,+\infty)$ such that $\int_{-\infty}^{+\infty} |g(s)|^2 ds < \infty$. For $z \in \mathbb{C}$ write

$$p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0 = \prod_{j=1}^{n} (z-z_j)$$

where $p(z_j) = 0$. For ε positive and real let

$$p_\varepsilon(z) = \prod_{j=1}^{n} (z-(z_j-\varepsilon)) \ . \tag{4.6}$$

Assume that $\text{Re}(z_j) \leq 0$ for $j = 1,\cdots,n$.

<u>Lemma 4.2.2</u>. Suppose that for each positive real ε there is a function $u_\varepsilon \epsilon L_2$ whose Fourier-Plancherel transform $\psi(u_\varepsilon)$ satisfies

$$\psi(u_\varepsilon)(-\lambda) = \frac{q(i\lambda)}{p_\varepsilon(i\lambda)} \ , \ \lambda \text{ real} \ ,$$

and there is a function $u \epsilon L_2$ such that $\psi(u)(-\lambda) = \frac{q(i\lambda)}{p(i\lambda)}$. Then $u_\varepsilon \to u$ in L_2 as $\varepsilon \to 0$.

<u>Proof</u>: Since ψ is an isometry of L_2 onto L_2(see Rudin (1973)) it suffices to prove that $\frac{q(i\lambda)}{p_\varepsilon(i\lambda)}$ converges to $\frac{q(i\lambda)}{p(i\lambda)}$ in L_2. Note that

$$\frac{q(i\lambda)}{p(i\lambda)} - \frac{q(i\lambda)}{p_\varepsilon(i\lambda)} = \frac{q(i\lambda)}{p(i\lambda)}\left(1 - \frac{p(i\lambda)}{p_\varepsilon(i\lambda)}\right) \ .$$

It is clear from the form of $\frac{p(i\lambda)}{p_\varepsilon(i\lambda)}$ that $\lim_{\varepsilon\to 0} \frac{p(i\lambda)}{p_\varepsilon(i\lambda)} = 1$ except at those points z_j of the form $i\lambda_j$. Also writing $z = a + ib$

$$\left|\frac{i\lambda - z}{i\lambda - z + \varepsilon}\right| \leq 1 + \frac{\varepsilon}{\sqrt{(\varepsilon - a)^2 + (\lambda - b)^2}} \leq 2$$

for all $\lambda \in (-\infty, +\infty)$. It follows that

$$\left|1 - \frac{p(i\lambda)}{p_\varepsilon(i\lambda)}\right| \leq 1 + 2^n \tag{4.7}$$

for all $\lambda \in R$. Since

$$\left|\frac{q(i\lambda)}{p(i\lambda)}\right| \left|\left(1 - \frac{p(i\lambda)}{p_\varepsilon(i\lambda)}\right)\right| \leq \left|\frac{q(i\lambda)}{p(i\lambda)}\right| (2^n + 1)$$

and $\frac{q(i\lambda)}{p(i\lambda)} \in L_2$ the dominated convergence theorem shows that

$$\frac{q(i\lambda)}{p_\varepsilon(i\lambda)} \to \frac{q(i\lambda)}{p(i\lambda)} \quad \text{in } L_2 \,.$$

The proof of the lemma is complete. $\square$

Let

$$\tau(\theta) = \int_0^1 \theta(s) u(s) ds$$

where $u \in L_2[0,1]$ and define for $\lambda \in R$

$$\phi(\lambda) = \int_0^1 e^{i\lambda s} u(s) ds \,.$$

That ϕ is well defined follows from the usual theory of Fourier transforms for L_1 functions since $u \in L_1[0,1]$.

Theorem 4.2.2. If there exists an integer $n \geq 1$, real constants $a_0, \cdots, a_{n-1}, b_{0u}, \cdots, b_{n-1u}$ and points $t_u \in [0,1]$, $u = 1, \cdots, r$, such that

$$\frac{\sum_{u=1}^{r} \sum_{v=0}^{n-1} [b_{vu}(i\lambda)^v] e^{i\lambda t_u}}{(i\lambda)^n + a_{n-1}(i\lambda)^{n-1} + \cdots + a_0} = \phi(\lambda) ,$$

where $z^n + a_{n-1}z^{n-1} + \cdots + a_0 = 0$ implies $\mathrm{Re}(z) \leq 0$, then the best unbiased estimator $\hat{\tau}(\theta)$ of $\tau(\theta)$ satisfies

$$\hat{\tau}(\theta) = \sum_{u=1}^{r} \sum_{v=0}^{n-1} b_{vu} Y^{(v)}(t_u)$$

where Y is the solution to

$$(D^n + a_{n-1}D^{n-1} + \cdots + a_0 I)Y = \theta + \sigma W'$$

on $[0,1]$ subject to $Y^{(j)}(0) = 0$, $j = 0, \cdots, n-1$.

Proof: First suppose that the zeros of

$$p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$$

all have negative real parts. That is, $p(z) = 0$ implies $\mathrm{Re}(z) < 0$.

Denote by $G(t,s)$ the Green's function of the differential operator

$$L = D^n + a_{n-1}D^{n-1} + \cdots + a_0 I \ .$$

That is, $G(t,s) = h(t-s)$ where $h(u) = 0$ for $u < 0$, $h^{(j)}(0) = \delta_{j,n-1}$, $j = 0,\cdots,n-1$ and $Lh = 0$ on $[0,1]$. Since the zeros of $p(\cdot)$ have negative real parts h is in $L_1(-\infty,+\infty) \cap L_2(-\infty,+\infty)$ and the Fourier transform ψh of h satisfies

$$H(\lambda) = \psi h(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda u} h(\ )du = [p(i\lambda)]^{-1} \ .$$

If $t_1,t_2,\cdots,t_t$ are points in $[0,1]$ and c_j are arbitrary scalars

$$\int_{-\infty}^{\infty} e^{i\lambda u} \sum_{j=1}^{r} c_j h(u-t_j)du = \left[\sum_{j=1}^{r} c_j e^{i\lambda t_j} \right] H(\lambda) \ .$$

It is also easily verified that

$$\int_{-\infty}^{\infty} e^{i\lambda u} h^{(v)}(t_j-u)du = (i\lambda)^v H(\lambda) e^{i\lambda t_j}$$

for $v = 1,2,\cdots,n-1$, where

$$\frac{d^v}{ds^v} h(s) \Big/_{t_j-u} = h^{(v)}(t_j-u) = G^{(v)}(t_j,u) = \frac{\partial^v}{\partial t^v} G(t,u) \Big/_{t=t_j}$$

One has therefore

$$\psi[\sum_{u=1}^{r} \sum_{v=0}^{n-1} b_{vu} G^{(v)}(t_u,t)](-\lambda) = \phi(\lambda) \ . \tag{4.8}$$

If $\bar{u}(t) = u(t)X_{[0,1]}(t)$ then $\bar{u}\epsilon L_1(-\infty,+\infty) \cap L_2(-\infty,+\infty)$ and one has from (4.8) that

$$\psi[\sum_{u=1}^{r}\sum_{v=0}^{n-1} b_{vu}G^{(v)}(t_u,t)](-\lambda) = \psi\bar{u}(-\lambda) .$$

Therefore, in the L_2 sense,

$$\bar{u}(t) = \sum_{u=1}^{r}\sum_{v=0}^{n-1} b_{vu}G^{(v)}(t_u,t) \tag{4.9}$$

on $(-\infty,+\infty)$. In particular this equality holds on [0,1].

Define the function g on [0,1] by

$$g(t) = \sigma^2\int_0^1 G(t,s)\sum_{u=1}^{r}\sum_{v=0}^{n-1} b_{vu}G^{(v)}(t_u,s)ds .$$

By lemma 4.2.1 (a) and (b) $g\epsilon H(K)$. Also by (4.9) we have

$$g(t) = \int_0^1 G(t,s)\sigma^2 u(s)ds .$$

Therefore $g\epsilon M \subset H(K)$ where M is the closed subspace of possible mean functions of the Y process with $Y^{(j)}(0) = 0$, $j = 0,\cdots,n-1$ defined by

$$M = \{m(\theta)\colon \theta\epsilon L_2[0,1]\} .$$

From Parzen's theory, if

$$\tau(\theta) \overset{\theta}{\equiv} (g, m(\theta))_k$$

then the linear operations which yield g in terms of the kernel K are the same ones which when applied to the process Y yield the umvlue of $\tau(\theta)$. Since

$$(g, m(\theta))_k = \frac{1}{\sigma^2} \int_0^1 \sigma^2 u(s)\theta(s)ds = \tau(\theta)$$

we conclude by lemma 4.2.1 that the umvlue (by normality the umvue) of $\tau(\theta)$ is

$$\hat{\tau}(\theta) = \sum_{u=1}^{r} \sum_{v=0}^{n-1} b_{vu} Y^{(v)}(t_u) .$$

In order to complete the proof of the theorem we must demand only $Re(z) \leq 0$ for z satisfying $p(z) = 0$.
From the proof above one may glean the fact that it suffices to prove that (in $L_2[0,1]$)

$$u(t) = \sum_{u=1}^{r} \sum_{v=0}^{n-1} b_{vu} G^{(v)}(t_u, t) \tag{4.10}$$

holds where $G(t,s) = h(t-s)$, but now h is not in $L_1(-\infty,+\infty)$. Let $\varepsilon > 0$ be arbitrary and define $G_\varepsilon(t,s)$ to be the Green's function of the differential operator $p_\varepsilon(D)$ where $p_\varepsilon(z)$ is defined in (4.6). Setting $q(i\lambda) = \sum_{u=1}^{r} \sum_{v=0}^{n-1} (i\lambda)^v e^{i\lambda t_u}$ we observe that

$\frac{q(i\lambda)}{p_\varepsilon(i\lambda)} = \frac{q(i\lambda)}{p(i\lambda)}\frac{p(i\lambda)}{p_\varepsilon(i\lambda)}$ so by (4.7) $\frac{q(i\lambda)}{p_\varepsilon(i\lambda)} \in L_2$ for all $\varepsilon > 0$. Since ψ is an isometric isomorphism of L_2 onto itself there exists $u_\varepsilon \in L_2$ such that

$$\psi(u_\varepsilon) = \frac{q(i\lambda)}{p_\varepsilon(i\lambda)} .$$

Also, since the polynomial p_ε satisfies $p_\varepsilon(z) = 0 \Rightarrow \mathrm{Re}(z) < 0$ the preceding analysis yields

$$\sum_{u=1}^{r} \sum_{v=0}^{n-1} b_{vu} G_\varepsilon^{(v)}(t_u, v) = u_\varepsilon(t)$$

in L_2 where G_ε is the Green's function of the differential operator $p_\varepsilon(D)$. By lemma 4.2.1 (c) we have $G_\varepsilon^{(v)}(t,s)$ converging to $G^{(v)}(t,s)$ on $[0,1]\times[0,1]$ where G is the Green's function of $p(D)$. The proof is completed by observing that lemma 4.2.2 shows that u_ε converges in L_2 to $\bar{u}$. We conclude that (4.10) holds and therefore the theorem.

<u>Example 4.2.1</u>. Suppose we wish to estimate $\int_0^1 \theta(t)dt$. We are able to observe $\{\theta(t) + \sigma W'(t) : t \in [0,1]\}$. Since $\tau(\theta) = \int_0^1 1\cdot\theta(t)dt$ we see that

$$\phi(\lambda) = \frac{e^{i\lambda} - 1}{i\lambda} .$$

It follows from theorem 4.2.2 that if we observe the output $\{Y(t): t \in [0,1]\}$ of the instrument whose differential operator is $L = D$ then the minimum variance unbiased estimator of $\tau(\theta)$ is $\hat{\tau}(\theta) = Y(1) - Y(0) = Y(1)$.

Example 4.2.2. Suppose the same conditions as in example 4.2.1 except we wish to estimate $\tau(\theta) = \int_0^1 \sin(wt)\theta(t)dt$. Computations show that

$$\phi(\lambda) = \frac{[-w \cos w + i\lambda \sin w]e^{i\lambda} + w}{(i\lambda)^2 + w^2}$$

Therefore if $\{Y(t): t \in [0,1]\}$ is the output of the instrument whose dynamic response is given by $L = D^2 + w^2 I$, $Y(0) = Y^{(1)}(0) = 0$, when the input is $\{\theta(t) + \sigma W'(t): t \in [0,1]\}$ then the umvlue of $\tau(\theta)$ is

$$\begin{aligned} \hat{\tau}(\theta) &= -w \cos w \, Y(1) + \sin w \, Y'(1) + wY(0) \\ &= -w \cos w \, Y(1) + \sin w \, Y'(1) \, . \end{aligned}$$

For the sake of simplicity we have assumed that the initial conditions on the solution Y to $(D^n + a_{n-1}D^{n-1} + \cdots + a_0 I)Y = F + \sigma W'$ satisfy $\mu_j = E[Y^{(j)}(0)] \equiv 0$, $\sigma_j^2 = \mathrm{Var}[Y^{(j)}(0)] \equiv 0$, $j = 0, \cdots, n-1$. The results stated above continue to hold true for arbitrary μ_j's and σ_j^2's whenever $\sigma_j^2 = 0$ implies $\mu_j = 0$ and the initial values are mutually uncorrelated and uncorrelated with W. The case of intercorrelations can also be treated. (See section 4.4 below).

4.3. Modulation of the forcing function. In the preceding section we studied the efficacious selection of the coefficients in the linear differential operator describing the dynamic response of the measuring instrument. In this section we consider the proper selection of the forcing term $F(x,\theta,t)$. Our objective is limited to the proper modification of the mean forcing term. We do not consider filtering or other modification of the input $F(x,\theta,t) + \sigma_x W'(t)$. We shall deal with this problem elsewhere.

A general description of the experimental situation and data follows.

a) The experimenter observes (or has available) the stochastic processes $\{Y(x_i,t): t\in[0,1],\ i = 1,\cdots,N\}$. The process $Y(x_i,t)$ solves

$$L[Y(x_i)] = F(x_i,\theta) + \sigma_{x_i} W_i'$$

subject to $Y^{(j)}(x_i,0) = 0$, $j = 0,\cdots,n-1$, where

$$L = D^n + a_{n-1}D^{n-1} + \cdots + a_0 I$$

is an element of $\mathcal{L}$.

b) The unknown parameter θ is a member of the real linear space Θ and for each x in the set of feasible experiments X the mappings F_x from Θ to $L_2[0,1]$ defined by $F_x(\theta)(t) = F(x,\theta,t)$ are linear and of known form.

c) The real number $\tau(\theta)$, where τ is a linear functional on Θ is to be estimated.

Before presenting our results for the general problem we offer the following two examples. They illustrate some cases in which the optimal design problem of estimating a linear functional in θ may be solved using procedures which have already been studied to some extent in the existing literature. A feature which they share is that the parameter space Θ is R^k for some k.

Example 4.3.1. Suppose there is a function η in $L_2[0,1]$, $\Theta = R^k$, and there are function f_j mapping X into R, $j = 1,\cdots,k$, such that

$$F(x,\theta,t) = [\sum_{j=1}^{k} \theta_j f_j(x)]\eta(t) \ .$$

Any linear functional τ on Θ has values given by $\tau(\theta) = c'\theta$ for some fixed vector $c \epsilon R^k$. Writing $c'\theta$ rather than $\tau(\theta)$ one has

$$d(c,\xi) = \sup_{\theta \ U} \frac{[c'\theta]^2}{\theta' M(\xi)\theta} \frac{1}{\int_0^1 \eta^2(t)dt} ,$$

where $U = \{\theta: \theta' M(\xi)\theta > 0\}$ and

$$M(\xi) = \sum \xi(x) \underset{\sim}{f}(x) \underset{\sim}{f}'(x) \ .$$

Therefore the problem has been reduced to a regression design problem for scalar observations. There is a large literature dealing with the solutions to such problems. See, for example, Fedorov (1972), Karlin and Studden (1966) or Kiefer (1974). Note that if $\hat{\theta}$ is the Gauss-Markov estimator of θ based on the solution process Y (see Wahba (1979) and section 2.6 above) then defining $\hat{F}(x,t) = F(x,\hat{\theta},t)$ one has

$$E_\theta ||\hat{F}(x,t)-F(x,\theta,t)||^2_{L_2} = E_\theta ||\hat{m}(x)-m(x,\theta)||^2_K ,$$

where $m(x,\theta)$ is the mean function of the solution process Y and $\hat{m}$ is given by

$$\hat{m}(x,t) = [\sum_{i=1}^{k} \hat{\theta}_i f_i(x)] \int_0^1 G(t,s)\eta(s)ds .$$

Therefore the minimax design problem studied in section 2.6 may be solved in this special case by using the techniques appropriate for scalar deservations. We see also that there is an additional interpretation of the D-optimal design. It is also the minimax design for the unbiased estimation of the forcing function $F(x,\theta)$ as an element of $L_2[0,1]$.

<u>Example 4.3.2</u>. As in example 4.3.1 assume $\Theta = R^k$. Suppose that for each $i = 1,\cdots,k$ g_i is a function defined on X with values in $L_2[0,1]$ and

$$F(x,\theta,t) = \sum_{i=1}^{k} \theta_i g_i(x,t) .$$

In this case the mean of the process $\{Y(x,t)\colon t\in[0,1]\}$ is given by

$$m(x,\theta,t) = \sum_{i=1}^{k} \theta_i f_i(x,t) ,$$

where

$$f_i(x,t) = \int_0^1 G(t,s)g_i(x,s)ds .$$

Design problems of this sort were treated in section 2. Wahba (1979) has further studied these problems when X is contained in H(K) and f_i are bounded commuting self-adjoint operators. Note that as in Example 4.3.1 the D-optimal design has the additional interpretation of being the minimax design for the forcing function $F(x,\theta)$ as an element of $L_2[0,1]$.

For each finite subset S of X define the linear mapping F_S from Θ into $L_2^m[0,1]$ by $[F_S(\theta)(t)]_i = F(x_i,\theta,t)$ where $S = \{x_1,\cdots,x_m\}$ and $[v]_i$ is the i^{th} component of the m-vector v. Write F_x when $S = \{x\}$.

Lemma 4.3.1. a) For every $x \in X$ for which F_x is bounded, m_x is a bounded linear transformation from Θ to $H(K_x)$ and $m_x^* = \sigma_x^{-2}F_x^* \circ L$.
b) If for every finite subset $S \subset X$ the mappings F_S have closed range, then for every $\xi \in \Xi$, $R(m_\xi)$ is closed in H_ξ.

Proof: From lemma 4.1.1 we have

$$||m_x(\theta)||^2 = \frac{1}{\sigma_x^2} \int_0^1 F(x,\theta,t)dt \le \frac{1}{\sigma_x^2} ||F_x||^2||\theta||^2 ,$$

proving the first part of (a). Let $v \in H(K_x)$. Then by definition $m_x^*(v)(\theta) = (v, m_x(\theta))_K$. Therefore

$$m_x^*(v)(\theta) = \frac{1}{\sigma_x^2} \int_0^1 L(v)(t) F_x(\theta)(t)dt = \frac{1}{\sigma_x^2} (Lv, F_x\theta)_{L_2} \tag{4.11}$$

where $L \in \mathcal{L}$ is the differential operator associated with the process. From (4.11) we have that $m_v^*(v)$ corresponds to $\dfrac{F_x^* \circ L(v)}{\sigma_x^2}$.

To prove the second part, suppose $\{\theta_n\} \subset \Theta$ and $m_\xi(\theta_n)$ converges to z in H_ξ. Then

$$||m_\xi(\theta_n) - z||_\xi^2 = \sum_S \frac{\xi(x)}{\sigma_x^2} \int_0^1 (F(x,\theta_n,t) - Lz_x(t))^2 dt ,$$

where $S = S(\xi)$ (the support of ξ), converging to zero shows that for each $x \in X$, $F(x,\theta_n)$ converges in measure (Lebesgue) to $Lz_x(t)$. One can extract a subsequence $\{\theta_n'\}$ such that

$$F_x(\theta_n') \overset{\text{a.e.}}{\rightarrow} Lz_x \tag{4.12}$$

for all x in S. Since the range of F_S is closed one must have a θ_0 such that

$$F_S(\theta_n') \rightarrow F_S(\theta_0) . \tag{4.13}$$

From (4.12) and (4.13) we have the existence of a point $\theta_0 \in \Theta$ such that

$$F(x,\theta_0,t) = Lz_x(t) \tag{4.14}$$

for all $x \in S$. From (4.14) one has

$$z_x(t) = \int_0^1 G(t,s)F(x,\theta_0,s)ds$$

or equivalently $z = m_\xi(\theta_0)$. This concludes the proof of the lemma. □

We are now prepared to specialize the theorems of section 3.3 to the case of interest in this section, the case of processes which arise as solutions to certain white noise random differential equations. We choose to present only the specialization of Theorem 3.3.2, the specialization of 3.3.1 can be obtained similarly. We make the following assumptions.

(C1) The parameter space Θ is $L_2[0,1]$.

(C2) There is a constant $k > 0$ such that

$$\sup_X \int_0^1 \frac{F^2(x,\theta,t)}{\sigma_x^2} dt \geq k||\theta||^2 \text{ for all } \theta \in L^2.$$

(C3) The mappings $F_S: \Theta \to L_2$ defined above are bounded linear transformations and all have closed range.

Let $\tau(\theta) = \int_0^1 u(s)\theta(s)ds$ be a fixed linear functional on Θ and define

$$\Delta = \{\theta: \tau(\theta) = 1\} .$$

Theorem 4.3.1. Under the assumptions (C) if $\delta_0 \in \Delta$ satisfies

$$\text{i)}\quad \inf_{\Delta} \sup_{X} \int_0^1 \frac{F^2(x,\theta,t)}{\sigma_x^2}\, dt = \sup_X \int_0^1 \frac{F^2(x,\delta_0,t)}{\sigma_x^2}\, dt\ ,$$

$$\text{ii)}\quad S(\xi_0) \subset \left\{ x: \int_0^1 \frac{F^2(x,\delta_0,t)}{\sigma_x^2}\, dt = \sup_X \int_0^1 \frac{F^2(x,\delta_0,t)}{\sigma_x^2}\, dt \right\}\ ,$$

and iii) $\int_0^1 \frac{F_x^* F_x}{\sigma_x^2} \delta_0 d\xi_0(x)$ is proportional to u then ξ_0 is optimal for the estimation of $\tau(\theta)$ with $d(\tau,\xi_0) = \left(\sup_X \int_0^1 \frac{F^2(x,\delta_0,t)}{\sigma_x^2}\, dt \right)^{-1}$.

Furthermore, if an optimal design ξ_0 exists then there exists a $\delta_0 \in \Delta$ satisfying i), ii), and iii).

Proof: The proof can be accomplished, in light of theorem 3.3.2, lemma 4.3.1 above, and our assumptions by verifying

$$m_x^* m_x = \frac{F_x^* F_x}{\sigma_x^2}\ . \tag{4.15}$$

The reason for this is that (C1) shows that (B1) is satisfied. Lemma 4.3.1 above and (C3) show that (B2) is satisfied. Theorem 3.4.1, (C2) and lemma 4.3.1 show that (B3) is satisfied. Finally, (B4) follows from (C2) above.

For the verification of (4.15) observe that $m_x^*(v) = \frac{1}{\sigma_x^2} F_x^* \circ L(v)$ from lemma 4.3.1 above. Since $Lm_x(\theta) = F_x(\theta)$ one has $m_x^* m_x(\theta) = \frac{1}{\sigma_x^2} F_x^* F_x(\theta)$ concluding the proof.

Example 4.3.3. Suppose $\sigma_x \equiv \sigma$, $F(x,\theta,t) = x(t)\theta(t)$, $\Theta = L_2[0,1]$, and $X = \{f \in L_\infty[0,1]: ||f||_\infty \le c\}$. Assume that we wish to estimate $\tau(\theta) = \int_0^1 u(s)\theta(s)ds$ where $u \in L_2[0,1]$ is some arbitrary function. We show that the design ξ is optimal if and only if it places all mass on the functions $x \in X$ satisfying $|x(t)| = c$ a.e. and $d(\tau,\xi) = \frac{\sigma^2||\tau||^2}{c^2}$. We bgin by verifying the assumptions (C). The first is immediate. For (C2) observe

$$\sup_X \int_0^1 x^2(t)\theta^2(t)dt \ge c^2\int_0^1 \theta^2(t)dt . \tag{4.16}$$

Verification that F_x is bounded results from

$$||F_x(\theta)||_2^2 = \int_0^1 x^2\theta^2 \le c^2||\theta||^2 . \tag{4.17}$$

Finally, let $S = \{x_1,\cdots,x_m\} \subset X$, $\{\theta_n\} \subset L_2[0,1]$ and $F_S(\theta_n)$ converge to z in $L_2^m[0,1]$. That is

$$\max_{1\le i\le m} \int_0^1 (x_i(t)\theta_n(t) - z_i(t))^2 dt \to 0$$

as $n \to \infty$. Extract a subsequence $\{\theta_n'\}$ for which $x_i\theta_n'$ converges a.e. to z_i for all values of $i = 1,\cdots,m$. On the set where x_i is nonzero θ_n' converges a.e. to $\frac{z_i}{x_i}$. On the set where all x_i are zero let θ_0 be arbitrary. Off of this set one obtains

$$\theta_0(t) = \frac{z_i(t)}{x_i(t)} \quad \text{a.e.}$$

if $x_i \neq 0$. Therefore $F_S(\theta_n) \to F_S(\theta_0)$ in $L_2^m[0,1]$ showing that (C3) is satisfied.

From (4.16) we observe that

$$\inf_{\Delta} \sup_{X} \int_0^1 F^2(x,\theta,t)dt \geq \inf_{\Delta} c^2 \int_0^1 \theta^2(t)dt \;. \tag{4.18}$$

By Schwarz's inequality, if $\theta \in \Delta$

$$1 = \left(\int_0^1 \theta u\right)^2 \leq \int \theta^2 \int u^2 \tag{4.19}$$

with equality if and only if $\theta = ku$. Combining (4.18) and (4.19) shows that

$$\inf_{\Delta} \sup_{X} \int_0^1 F^2(x,\theta,t)dt \geq c^2 \left(\int_0^1 u^2\right)^{-1} \;. \tag{4.20}$$

Taking

$$\delta(t) = \frac{u(t)}{||u||_2^2} \tag{4.21}$$

one has

$$\sup_{X} \int_0^1 F^2(x,\delta,t)dt \leq c^2 ||u||_2^{-2} \;. \tag{4.22}$$

Combining (4.22) and (4.20) shows that we may take $\delta_0(t)$ to be the δ of (4.21). We note that if $|x(t)| = c$ a.e. fails then

$$\int_0^1 F^2(x,\delta_0,t)dt = \int_0^1 x^2\delta_0^2 < c||u||_2^{-2} .$$

We must now find a measure ξ_0 on X which satisfies iii). In the case which we are considering, $F_x(\theta) = x\theta$, one can verify that $F_x^* v = \frac{xv}{\sigma_x^2}$ so, in particular, $F_x^* F_x \delta_0 = \frac{x^2\delta_0}{\sigma^2}$. Therefore we seek those measures $\xi \in \Xi$ which satisfy

$$\sum \frac{\xi(x)}{\sigma^2} x^2(t) \frac{u(t)}{||u||^2} \propto u(t) \text{ a.e. } .$$

Clearly any measure ξ satisfying $S(\xi) \subset \{|x(t)| = c \text{ a.e.}\}$ satisfies this requirement.

Example 4.3.4. Make the same assumptions as in example 4.3.3 except that

$$X = \{\frac{1}{\sqrt{2}}, \sin(2\pi t), \cdots, \sin(2K_1\pi t)\}$$

and $\tau(\theta) = \int_0^1 \theta(s)\sin(2K_2\pi s)ds$, where $K_2 > K_1$ are positive integers. We shall verify that the unique optimal design ξ_0 for this problem places all mass at the constant function $x(t) = 2^{-1/2} \in X$.

Again, we begin by verifying the assumptions (C) and again (C1) is immediate. Since $\frac{1}{\sqrt{2}} \in X$ $\sup_X \int_0^1 x^2\theta^2 \geq \frac{1}{2}||\theta||_2^2$ verifying (C2). Certainly $||F_x(\theta)||^2 \leq ||\theta||^2$ and the portion of the proof that F_S has closed range given in example 4.3.3 applies without alteration to example 4.3.4.

Let

$$\delta(t) = 2 \sin(2K_2\pi t) \ . \tag{4.23}$$

Since for any integer $m \neq 0$

$$\int_0^1 \sin^2(2m\pi t)dt = \frac{1}{2} \tag{4.24}$$

the function δ is in $\Delta = \{\theta : \tau(\theta) = 1\}$. Denote the elements of X by $\{x_0,\cdots,x_{K_1}\}$ where $x_0(t) = 2^{-1/2}$ and $x_j(t) = \sin(2j\pi t)$, $j = 1,\cdots,K_1$. Since for any integers n and m, $n \neq m$,

$$\int_0^1 \sin^2(2n\pi t)\sin^2(2m\pi t)dt = \frac{1}{4} \tag{4.25}$$

one has

$$\int_0^1 \delta^2(t)x_j^2(t)dt \equiv 1, \ j = 0,1,\cdots,K_1 \ .$$

Since $2^{-1/2} \in X$,

$$\inf_{\Delta} \frac{1}{2} \int_0^1 \theta^2 \le \inf_{\Delta} \sup_{X} \int_0^1 \theta^2 x^2 .$$

Using (4.19) and (4.24) shows that $1 = \inf_{\Delta} \frac{1}{2} \int_0^1 \theta^2$. Now (4.25) shows that we may take δ_0 to be the δ defined in (4.23) and yields no information about the possible supports of optimal designs. We seek a design ξ_0 which satisfies

$$\left[\sum_{j=0}^{K_1} \frac{\xi_0(x_j)}{\sigma^2} x_j^2(t) \right] 2 \sin(2K_2\pi t) \propto \sin(2K_2\pi t) \quad \text{a.e.} ,$$

and consequently must have

$$\sum_{j=0} \xi(x_j) x_j^2(t) \propto 1 \text{ a.e.}$$

The unique measure ξ_0 which does this places all mass at $x_0(t) = 2^{-1/2}$.

4.4. <u>Generalizations</u>. For the most part we have assumed above that the solution to

$$LY = F(x,\theta) + \sigma_x W'$$

was on [0,1], a fixed subinterval independent of x, and the initial conditions were $Y^{(j)}(0) = 0$, $j = 0,\cdots,n-1$. This was done for the sake of simplicity. In the general case one may allow variable intervals and initial conditions. For example, we may suppose that Y_x solves $L_x Y_x = F(x,\theta) + \sigma_x W'$ on $[0,b(x)]$ where $E[Y_x^{(j)}(0)] = \mu_j$, $j = 0,\cdots,n(x)-1$, and $\mathrm{Var}[Y_x^{(j)}(0)] = \sigma_j^2(x)$. In this case we view $\tilde{\Theta} = \Theta\times R^\infty$, $\tilde{\theta} = (\theta,\mu_0,\mu_1,\cdots)$, as the parameter space and for estimators of $\tau(\theta)$ we have

$$d(\tau,\xi) = \sup_{\tilde{\theta}\in\tilde{U}} [\tau(\theta)]^2 / \sum \xi(x) \sum_{j=0}^{n(x)-1} \frac{\mu_j^2}{\sigma_j^2(x)} + \frac{1}{\sigma_x^2} \int_0^{b(x)} F^2(x,\theta,t)\,dt \tag{4.26}$$

$$= \sup_{\theta\in U} [\tau(\theta)]^2 / \sum \frac{\xi(x)}{\sigma_x^2} \int_0^{b(x)} F^2(x,\theta,t)\,dt \tag{4.27}$$

as above. However, we may also investigate functionals of the form $\tilde{\tau}(\tilde{\theta}) = c'\mu + \tau(\theta)$ where $c\in R^k$ for some finite number k. In this case (4.26) does not simplify to (4.27).

Another model results if it is assumed that some or all of the μ_j's are under the control of the experimenter. In order to preserve the linearity assumption made throughout the paper we may subtract these terms from the observable process. For example, if $\mu_0(x)$ is assumed to be under the control of the experimenter and $n(x) \equiv n > 2$ one has

$$Y(x,t) = \mu_0(x) z_0(t) + \sum_{j=1}^{n-1} \mu_j z_j(t) + \int_0^1 G(t,s)F(x,\theta,s)\,ds + N(t).$$

Setting $\tilde{Y}(x,t) = Y(x,t) - \mu_0(x)z_0(t)$, the process $\tilde{Y}$ is observable and the mean function is a linear mapping from $\Theta \times R^{n-1}$ into $H(K_x)$ if $F(x,\cdot)$ is linear.

From the relationship between (4.26) and (4.27) it can be seen that all results given in sections 4.1 through 4.3 hold regardless of the random initial conditions, as long as the variances of the initial conditions are positive.

If the variance of an initial condition, say $Y^{(j)}(0)$ is zero and the mean is unknown, the mean of the solution process Y_x is not in the reproducing kernel Hilbert space $H(K_x)$. However, taking $Y_x^{(j)}(0)$ for one observation yields μ_j which can then be subtracted to form $\tilde{Y}_x$ satisfying the assumptions of this section.

References

Aronszajn, N., (1950). Theory of reproducing kernels. Trans. Amer. Math. Soc. 68, 337-404.

Atwood, C. L., Convergent design sequences for sufficiently regular optimality criteria, The Annals of Statistics. Vol. 4, No. 6, 1976.

Box, G. E. and Jenkins, G. M., (1970). Time Series Analysis. Holden-Day, San Francisco.

Chernoff, Herman (1972). Sequential Analysis and Optimal Design. Siam, Philadelphia.

Dieudonné, J. (1960). Foundations of Modern Analysis. Academic Press, New York.

Doebelin, Ernest O. (1975). Measurement Systems, Application and Design, McGraw-Hill, New York.

Dolph, C. L., and Woodbury, M. A. (1952). On the relationship between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction. Trans. Amer. Math. Soc. 72 519-550.

Dubova, I. S., Fedorov, V. V., and Fedorova, G. S. (1977). Choice of the optimum trajectory when the response depends upon time.

Dunford, N. and Schwartz, J. T. (1958). Linear Operators Part I. General Theory. Interscience, New York.

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Statist. 23, 255-262.

Elfving, G. (1954). Geometric allocation theory. Skand. Aktuarietidsker, 170-190.

Elfving, G. (1959). Design of linear experiments. In Probability and Statistics, The Cramér Volume, U. Grenander ed., John Wiley, New York.

Fedorov, V. V., (1959). The design of experiments in the multiresponse case. Theor. Probability Appl. 16, 323-000.

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

Gribik, P. R., and Kortanek, K. O., Equivalence theorems and cutting plane algorithms for a class of experimental design problems, Siam Journal of Applied Mathematics, Vol. 32, No. 1, 1977.

Holmes, R. B. (1975). Geometric Functional Analysis and its applications. Springer-Verlag, New York.

Joiner, B. and Campbell, Cathy. (1976). Designing experiments when run order is important. Technometrics. **18** No. 3, 249-259.

Kalaba, R. E. and Spingarn, K. (1978). On the numerical determination of optimal inputs. J. of Optimization Theory and Applications **25** 219-227.

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

Kiefer, J. (1960). Optimum experimental designs V. Proc. Fourth Berkeley Symp. Math. Statist. Prob., Univ. of Calif. Press, **1** 381-405.

Kiefer, J., (1961). Optimum designs in regression problems II. Ann. Math. Statist. **32**, 298-325.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. **2** 849-879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Can. J. Math. **12**, 363-366.

Kimeldorf, George S. and Wahba, Grace (1970). Spline functions and stochastic processes. Sankhya Ser. A Vol. 32 173-180.

Kuelbs, J. (1970). Gaussian measures on a Banach space. J. Functional Analysis **5**, 354-367.

Mehra, Raman K. (1974). Optimal input signals for parameter estimation in dynamic systems survey and new results IEEE Trans. Automat. Contr. AC-19, 753-768.

Naimark, M. A. (1967). Linear Differential Operators. Frederick Ungar Publishing Co., New York.

Nashed, M. Z. (1971). Generalized inverses, normal solvability, and iteration for singular operator equations. In Nonlinear Functional Analysis and Applications, L. B. Rall ed. Academic Press, New York.

Nashed, M. Zuhair and Votruba, G. F. (1976). A unified operator theory of generalized inverses. Generalized Inverses and Applications. Ed. M. Z. Nashed, Academic Press, New York.

Pazman, Andrej (1978a). Hilbert-space methods in experimental design. Kybernetika **14** 73-84.

Pazman, Andrej (1978b). Computation of the optimum designs under singular information matrices. Ann. Statist. **6** 465-467.

Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I. Technical Report No. 23, Stanford University.

Rudin, Walter (1973). Functional Analysis. McGraw-Hill, New York.

Sacks, Jerome and Ylvisaker, Donald. (1966). Designs for regression with correlated errors. Ann. Math. Statist. **37**, 66-89.

Sacks, Jerome and Ylvisaker, Donald (1968). Designs for regression problems with correlated errors; many parameters. Ann. Math. Statist. **39**, 49-69.

Sacks, Jerome and Ylvisaker, Donald (1969). Designs for regression problems with correlated errors, III. Ann. Math. Statist. **41**, 2057-2074.

Spruill, M. C. (1979). Some exact iterative procedures for finding the puncture point of a convex set. J. Optimization Theory and Applications. to appear.

Spruill, M. C. (1980). Optimal designs for second order processes with general linear means. Ann. Statist. to appear in Vol. 8 no. 3.

Spruill, Carl and Studden, W. J. (1978). Optimum designs when the observations are second order processes. J. Multivariate Anal. **8** 153-172.

Spruill, M. C. and Studden, W. J. (1979). A Kiefer-Wolfowitz theorem in a stochastic process setting. Ann. Statist. **7**, 1329-1332.

Studden, W. J. (1971). Elfving's theorem and optimal designs for quadratic loss. Ann. Math. Statist. **42**, 1613-1621.

Studden, William J. and Tsay, Jia-Yeong (1976). Remez's procedure for finding optimal designs. Ann. Statist. 4, 1271-1279.

Taylor, A. E. (1958). Introduction to Functional Analysis. Wiley, New York.

Tustin, A. (1947). The nature of the operator's response in manual control and its implications for controller design. J. Inst. Elec. Eng. **94** 190-202.

Viort, B., (1972a). On the dispersion matrix of the maximum likelihood estimators of the parameters of a dynamic model. Technical Report No. 313 Department of Statistics, University of Wisconsin.

Viort, B., (1972b). D-optimal designs for dynamic models, theory. Technical Report No. 314 Department of Statistics, University of Wisconsin.

Viort, B., (1972c). D-optimal designs for dynamic models, applications. Technical Report No. 316 Department of Statistics, Univeristy of Wisconsin.

Wahba, G., (1971). On the regression design problem of Sacks and Ylvisaker. Ann. Math. Statist. **42**, 1035-1053.

Wahba, G., (1974). Regression design for some equivalence classes of kernels. Ann. Statist. **2**, 925-934.

Wahba, Grace (1976). On the optimal choice of nodes in the collocation-projection method for solving linear operator equations. J. Approximation Theory. **16** 175-186.

Wahba, Grace (1979). Parameter estimation in linear dynamic systems. Tech. Rept. No. 547, Dept. of Statistics, Univ. of Wisconsin, Madison, Wisconsin.