DESIGNS FOR REGRESSION PROBLEMS WITH CORRELATED ERRORS

By Jerome Sacks[1] and Donald Ylvisaker[2]

Northwestern University and University of Washington

1. Introduction. The regression model which will underlie the discussion in this paper is as follows: An observation taken at a point t in the closed, bounded interval $[a, b]$ has the form

$$Y(t) = \sum_{i=1}^{k} \beta_i f_i(t) + X(t) \tag{1.1}$$

where $f_1, \cdots, f_k$ are given regression functions; $\beta_1, \cdots, \beta_k$ are unknown parameters; and the error, $X(t)$, is a random variable with $EX(t) = 0$ and $EX^2(t) < \infty$. The model is not quite specified yet since, if more than one observation is taken, say at t_1 and t_2 with both t_1 and t_2 in $[a, b]$, we must say something about the joint behavior of the random variables $X(t_1)$ and $X(t_2)$. Additionally, something should be said about the possibility of "repeating" an observation, i.e., whether two or more observations may be taken at the same t.

The most thoroughly discussed model has been the one which assumes uncorrelated errors with constant variance and repeatable observations. In this case, if n observations are taken at m distinct points $t_1, \cdots, t_m$ with μ_i observations taken at t_i, $\mu_1 + \cdots + \mu_m = n$, the questions of best linear unbiased estimation of $\beta_1, \cdots, \beta_k$, or of some set of linear combinations of the β_i's, have answers which have been known for some time. In recent years, the corresponding design problems, involving optimum choice of the t_i's and μ_i's, have received extensive attention, notably in the papers of Kiefer and Wolfowitz [14] and [15], and Kiefer [10], [11], [12] and [13]. In this paper, we are concerned with such an n observation design problem when the error process has a smooth correlation structure and observations are not repeatable. Our perspective is most easily understood by viewing the error process X as a time series and the experimenter as sampling in time. In this stochastic process context, where an infinite observation set is considered feasible, linear estimation of the β_i's is well understood, due especially to the efforts of Grenander [3], Hájek [4], [5] and [6], and Parzen [17], [18] and [19] (this theory subsumes, of course, the standard estimates for finite observation sets). Within this framework, the design problem can be looked at also from a regret point of view, comparing n-observation designs with observation over all of [0, 1]; this is of interest since the estimators based on interval observation are frequently unknown or, if known, may be difficult to compute; the estimators based on n observations require the inversion of an $n \times n$ matrix and this may be more feasible. The best es-

Received 10 February 1965; revised 18 August 1965.

[1] Research sponsored by NSF grant G-24500.

[2] Research sponsored by NSF grant G-25211.

timators depend heavily on the form the correlation takes and we will assume it is known up to a multiplicative constant. In the area of design for comparable situations, we are aware only of the work of Hoel [8] and [9]; there is little overlap in these papers with what we do.

The non-repeatability of observations is natural enough in time series design problems, but it is not necessarily incompatible with uncorrelated errors. Elfving [2] treats a problem possessing both of these elements, for example. In either event the elimination of repeated observations introduces a different aspect into the design problem. For example, suppose that the errors are uncorrelated with constant variance, that $k = 1$ and that f_1 is continuous on $[a, b]$ with $|f_1|$ having a unique maximum at t_0, say. If observations *are* repeatable, the design which minimizes the variance of the best linear estimate of β_1 is, trivially, that which calls for n observations at t_0; if observations are *non*-repeatable there is no optimum (in the sense just described) design. The difficulty, in this latter instance, is that the optimum, if it existed, would be obtained by maximizing $\sum_{i=1}^{n} f_1^2(t_i)$ on the non-compact set $a \leqq t_1 < \cdots < t_n \leqq b$. This phenomenon is typical of problems with $k > 1$ as well. In certain instances, the presence of correlated errors will circumvent the non-compactness above and we turn now to the model of our specific interest.

The design problem with correlated errors exhibits considerable complication when $k = 1$ (unlike the uncorrelated, repeatable case) so that we take $k = 1$ throughout this paper. The cases $k > 1$ will be considered in a subsequent article. For convenience, we replace f_1 by f, β_1 by β and $[a, b]$ by $[0, 1]$. For non-repeatable observations, an n-point design is an n-tuple $\{t_1, \cdots, t_n\}$ with $0 \leqq t_1 < \cdots < t_n \leqq 1$ and we denote by D_n the class of all n-point designs. If $T \varepsilon D_n$, we let V_T be the variance of the best linear estimate (BLE) of β based on observations taken according to T. The criterion of optimality we shall use is the natural one for $k = 1$, viz., T^* is optimum if V_T is minimized (over D_n) by T^*.

In order to understand some of the complications that arise, suppose $X(\cdot)$ is a Brownian motion with, therefore, the covariance function $R(s, t) = EX(s)X(t) = \min(s, t)$ for s, t in $[0, 1]$. For $T \varepsilon D_n$, it is known (see Section 4) that

$$V_T^{-1} = f^2(t_1)/t_1 + \sum_{i=1}^{n-1} [(f(t_{i+1}) - f(t_i))^2/(t_{i+1} - t_i)] \tag{1.2}$$

where, if $t_1 = 0$ and $f(0) = 0$, the first term on the right hand side of (1.2) is 0, and, if $t_1 = 0$ and $f(0) \neq 0$, we define $V_T^{-1} = +\infty$. Clearly, if $f(0) \neq 0$, no design is better than taking one observation at 0; we may ignore this trivial case and assume $f(0) = 0$. If the derivative of f fails to exist in an unpleasant enough way (e.g., if f has a discontinuity at some t_0 in $(0, 1)$ or if $f(t) = |t - t_0|^p + C$ with $0 < p < \frac{1}{2}$ near some t_0 in $(0, 1)$) then the supremum of the right hand side of (1.2) is ∞ for $n > 1$ and clearly there is no optimum design. This is reminiscent of the uncorrelated non-repeatable observation model but, unlike that case, optimum designs exist for a large class of functions. Indeed, it will be seen, for example, that if f has a continuous derivative on $[0, 1]$, optimum designs exist for this error process. The question raised by this is: Given the covariance function R, for what functions f does there exist an optimum design?

In Section 2, the existence question will be treated under the assumption that f is in the reproducing kernel Hilbert space $\mathfrak{F}$ of functions on [0, 1] which is associated with the covariance function R (we are no longer assuming $X(\cdot)$ is Brownian motion). That this space is an appropriate place for the regression function follows from the work of Parzen who first exploited the general theory of these spaces in connection with time series regression analysis. Here we develop regression results concerning $\mathfrak{F}$ in a manner somewhat different than Parzen's in order to emphasize design considerations. Subsequently, we show that for a class of R's an optimum design exists for each n if $f \in \mathfrak{F}$ (optimum designs may exist for $f \notin \mathfrak{F}$ as well, note $f(t) \equiv 1$ in the Brownian motion case as a trivial example).

Returning to the Brownian motion example, suppose $f \in \mathfrak{F}$ so that, for each n, an optimum design exists (Section 4). The problem of calculating these designs confronts us. Examination of (1.2) reveals that such a calculation is not a simple task; in fact, we are only able to calculate all exact n-point designs for simple piecewise linear functions and for $f(t) = t^2$, where the optimum n-point design is given by $t_i = i/n$, $i = 1, \cdots, n$. The difficulty in computation leads us, in Section 3, to circumvent the problem of finding *exactly* optimum designs by proposing an asymptotic "solution," which may then be used to calculate approximately optimum n-point designs.

What we do is motivated by noticing that $\min_{T \varepsilon D_n} V_T \to \|f\|^{-2} = V_0$ for $f \in \mathfrak{F}$. It is then natural to say that a sequence of designs $\{T_n ; n \geqq 1\}$, $T_n \in D_n$, is asymptotically optimum if

$$\lim_{n \to \infty} [(V_{T_n} - V_0)/(\min_{T \varepsilon D_n} V_T - V_0)] = 1.$$

For some R's and some f's, we are able to obtain such sequences $\{T_n\}$. For example, when $R(s, t) = \min(s, t)$ and f has a continuous second derivative, we find that the design $T_n^* = (t_{1n}^*, \cdots, t_{nn}^*)$ with

$$\int_0^{t^*_{jn}} [f''(s)]^{2/3} \, ds = (j/n) \int_0^1 [f''(s)]^{2/3} \, ds, \qquad j = 1, \cdots, n, \tag{1.3}$$

is the nth element in an asymptotically optimum sequence. It is comforting to note that, in the case $f(t) = t^2$, T_n^* as just defined, agrees with the exactly optimum n-point design. (The reader may note that the T_n^* of (3.38) and the T_n^* of (1.3) differ; this may be reconciled by taking T_{n+1}^* in (3.38) and throwing away the observation at 0 since for Brownian motion an observation at 0 provides no information when $f(0) = 0$.)

For the kernel $R(s, t) = \min(s, t)$, we are not the first to give (1.3) as an approximate solution to the problem of maximizing the right hand side of (1.2). Särndal ([20] and [21]) in treating other problems, has given this same result under more stringent conditions on f. His argument invokes the existence of unique optimum designs and this last is achieved by assuming f to be five times continuously differentiable and convex (or concave).

Asymptotically optimum sequences are obtained generally under some rather strong restrictions on R and we discuss, particularly in Remark 3.2, why the result seems hard to obtain for a wider class of kernels. A sidelight of the development in Section 3 is contained in Lemma 3.1 where we obtain an estimate of $V_T^{-1} - V_0^{-1}$ for $T \varepsilon D_n$, viz., that under milder restrictions than stated in Theorem 3.1, $V_T^{-1} - V_0^{-1} \leqq K \sup_{0 \leqq i \leqq n} (t_{i+1} - t_i)^2$ when $T = \{t_1, \cdots, t_n\}$, $t_0 = 0$, $t_{n+1} = 1$ (the value of K can be seen in Lemma 3.1). This result enables us to find an integer n_0 so that taking more than n_0 observations would lead to a negligible gain in the variance of the BLE and, as such, may serve as a useful guide in helping the experimenter decide on the number of observations to be taken.

The final section of this paper is given over to examples and remarks. We intend there to exhibit the design problem in situations most amenable to the types of calculations which are of interest. In particular, we defer until then the (straightforward) verification of some statements from prior discussion and we give examples of asymptotically optimum designs.

2. Existence of optimum designs. In this section, the design problem is placed in the context of the reproducing kernel space $\mathfrak{F}$ of functions on [0, 1] associated with the covariance kernel R. Parzen has identified this function space with the class of regression functions which are natural in the presence of the underlying correlation structure. We present a slightly different treatment of the appropriate facts than those previously given so as to demonstrate the role of $\mathfrak{F}$ in design questions. Specifically, we show that the class, $\mathfrak{F}_n$, of regression functions on [0, 1] for which n-point designs produce minimum variances bounded away from zero, forms a Banach space with a suitable norm, and that $\mathfrak{F}_n$ shrinks to $\mathfrak{F}$ as $n \rightarrow \infty$.

Then, supposing the regression function f to be in $\mathfrak{F}$, we study the problem of minimizing minimum variance in the class of n-point designs. In order that the minimum be achieved, it is generally necessary to consider "boundary" designs to the set of possible designs and in certain circumstances, the minimum will actually occur at a boundary point and thereby not correspond to a physically realizable design. However, as we shall see below, this unpleasantness will not occur under a further assumption on R.

Suppose now that $Y(t)$ is given by (1.1) for $t \varepsilon [0, 1]$, with $k = 1$ and $EX(s)X(t) = R(s, t)$. Our basic assumptions concerning this model are that the kernel R thus defined on $[0, 1] \times [0, 1]$ forms a non-singular matrix when restricted to $T \times T$ for any finite set T and that the process X is continuous in the mean on [0, 1]. (The first assumption is for convenience only, the second is a standard "smoothing" assumption.)

We begin by considering finite observation sets drawn from the interval [0, 1]. Suppose $T = \{t_1, \cdots, t_n\}$ is such a set with $0 \leqq t_1 < t_2 < \cdots < t_n \leqq 1$. Let Y_T be the observation vector, $Y_T = (Y(t_1), \cdots, Y(t_n))'$, f_T be the restriction to T of the regression function, $f_T = (f(t_1), \cdots, f(t_n))'$, and R_T be the restriction to $T \times T$ of the covariance kernel, $R_T = (R(t_i, t_j))$. For minimum variance linear unbiased estimation in this context, one wants an n-vector c which satisfies

$E_\beta(c'Y_T) = c'f_T\beta = \beta$ and, subject to this condition, which minimizes $\sigma_\beta^2(c'Y_T) = c'R_Tc$. This minimization problem is solved by a choice of $c = R_T^{-1}f_T/f_T'R_T^{-1}f_T$ and the corresponding estimate has variance $(f_T'R_T^{-1}f_T)^{-1}$. Of course, if the set T is enlarged through adjoining new points, this minimum variance does not increase. We note here that, due to the positive definiteness of R_T, the collection of all n-vectors becomes a Hilbert space under the norm $\| \quad \|_T$ given by $\|u\|_T^2 = u'R_T^{-1}u$.

Let $D_n = \{T \mid T = \{t_1, \cdots, t_n\}, 0 \leqq t_1 < \cdots < t_n \leqq 1\}$. For f an arbitrary real valued function on [0, 1], consider the functional $\sup_{T\varepsilon D_n} \|f_T\|_T$ where f_T is again the restriction of f to T. (Since no difficulty should arise, we drop this last T when considering the norm and write $\|f\|_T$.) Now, let

$$\mathfrak{F}_n = \{f \text{ on } [0, 1] \mid \sup_{T\varepsilon D_n} \|f\|_T < \infty\}. \tag{2.1}$$

$\mathfrak{F}_n$ is the class of regression functions whose n-point design minimum variances are bounded away from zero.

THEOREM 2.1. *$\mathfrak{F}_n$ is a Banach space under the norm $\| \quad \|_n$ defined by $\|f\|_n = \sup_{T\varepsilon D_n} \|f\|_T$.*

PROOF. The fact that for each $T \varepsilon D_n$, $\| \quad \|_T$ is a norm, shows that $\|af\|_n = |a| \, \|f\|_n$ and that if $\|f\|_n = 0$, then $\|f\|_T = 0$ for all T, $f_T = 0$ for all T and $f \equiv 0$. Secondly, $\|f + g\|_n = \sup_{T\varepsilon D_n} (\|f + g\|_T) \leqq \sup_{T\varepsilon D_n} (\|f\|_T + \|g\|_T) \leqq \|f\|_n + \|g\|_n$. Thus $\| \quad \|_n$ is a valid norm and we need only show completeness. Suppose then that $\|f_k - f_m\|_n \to 0$ as $k, m \to \infty$. For each $T \varepsilon D_n$, $\|f_k - f_m\|_T \to 0$ and therefore there exists a vector f^T such that $\|f_k - f^T\|_T \to 0$ as $k \to \infty$. In particular, $(f_k)_T \to f^T$ in a pointwise sense. If S and T are both in D_n, $S \cap T \neq \varnothing$, we have $\|f_k - f^T\|_T \geqq \|f_k - f^T\|_{S\cap T}$ and $\|f_k - f^S\|_S \geqq \|f_k - f^S\|_{S\cap T}$ from which it follows that f^S and f^T agree on $S \cap T$. Consequently, there is a function f on [0, 1] which satisfies $(f)_T = f^T$ for all $T \varepsilon D_n$. Now let N be so large that $\|f_k - f_m\|_n < \epsilon$ for $k > N$, $m > N$. Then $\|f_k - f\|_T \leqq \|f_k - f_m\|_T + \|f_m - f\|_T < \epsilon + \|f_m - f\|_T$ for $k > N$, $m > N$ and all $T \varepsilon D_n$. Choosing, for each $T \varepsilon D_n$, m so large that $\|f_m - f\|_T < \epsilon$, we have $\|f_k - f\|_T < 2\epsilon$ when $k > N$ for all T. Hence $\|f_k - f\|_n \to 0$ as $k \to \infty$.

By virtue of an earlier remark, the norms $\| \quad \|_n$ are non-decreasing in n so that the spaces $\mathfrak{F}_n$ satisfy $\mathfrak{F}_n \supset \mathfrak{F}_{n+1}$ for all n. Let

$$\mathfrak{F} = \{f \text{ on } [0, 1] \mid \lim_{n\to\infty} \|f\|_n < \infty\}. \tag{2.2}$$

$\mathfrak{F}$ is the class of regression functions for which all finite observation sets have associated minimum variances bounded away from zero.

THEOREM 2.2. *$\mathfrak{F}$ is a Hilbert space under the norm $\| \quad \|$ defined by $\|f\| = \lim_n \|f\|_n$.*

PROOF. Clearly $\| \quad \|$ is a norm on $\mathfrak{F}$. To show completeness, suppose $\|f_k - f_m\| \to 0$ as $k, m \to \infty$. Since $\|f_k - f_m\|_n \leqq \|f_k - f_m\|$, there is for each n a function f^n for which $\|f_k - f^n\|_n \to 0$ as $k \to \infty$, this by Theorem 2.1. Inasmuch as this also implies that $f_k \to f^n$ pointwise, we have the existence of a single function $f \varepsilon \bigcap_n \mathfrak{F}_n$ for which $\|f_k - f\|_n \to 0$ as $k \to \infty$, any n. Now

$$\|f_k - f\|_n \leqq \|f_k - f_m\|_n + \|f_m - f\|_n \leqq \|f_k - f_m\| + \|f_m - f\|_n .$$

If N is so large that $\|f_k - f_m\| < \epsilon$ for $k > N$, $m > N$, then $\|f_k - f\|_n < \epsilon + \|f_m - f\|_n$. Choosing, for each n, m so large that $\|f_m - f\|_n < \epsilon$, we have $\|f_k - f\|_n < 2\epsilon$ for all n. Hence $\|f_k - f\| \to 0$ as $k \to \infty$. It remains to show that the parallelogram law holds. We note first that if $f, g \varepsilon \mathfrak{F}$,

$$\lim_n \sup_{T \varepsilon D_n} (\|f\|_T + \|g\|_T) = \lim_n \sup_{T \varepsilon D_n} \|f\|_T + \lim_n \sup_{T \varepsilon D_n} \|g\|_T = \|f\| + \|g\|,$$

for we can find a set $T \varepsilon D_n$, n sufficiently large so that $\|f\|_T > \|f\| - \epsilon$ and a set $T^* \varepsilon D_m$, m sufficiently large so that $\|g\|_{T^*} > \|g\| - \epsilon$. Then $\|f\|_{T \cup T^*} + \|g\|_{T \cup T^*} > \|f\| + \|g\| - 2\epsilon$. With this we note that

$$\begin{aligned}\|f + g\|^2 + \|f - g\|^2 &= \lim_n \sup_{T \varepsilon D_n} \|f + g\|_T^2 + \lim_n \sup_{T \varepsilon D_n} \|f - g\|_T^2 \\ &= \lim_n \sup_{T \varepsilon D_n} (\|f + g\|_T^2 + \|f - g\|_T^2) \\ &= \lim_n \sup_{T \varepsilon D_n} (2\|f\|_T^2 + 2\|g\|_T^2) \\ &= 2 \lim_n \sup_{T \varepsilon D_n} \|f\|_T^2 + 2 \lim_n \sup_{T \varepsilon D_n} \|g\|_T^2 \\ &= 2\|f\|^2 + 2\|g\|^2.\end{aligned}$$

The Hilbert space $\mathfrak{F}$ can be readily identified as the reproducing kernel space of functions of [0, 1] associated with R. For this, we need only verify the two defining properties of the kernel space, [1]:

(2.3) $$\text{for each } t \varepsilon [0, 1], R(\cdot, t) \varepsilon \mathfrak{F},$$

(2.4) $$\text{for each } f \varepsilon \mathfrak{F} \text{ and each } t \varepsilon [0, 1), (f, R(\cdot, t)) = f(t).$$

It is easy to see that (2.3) is satisfied, for $\|R(\cdot, t)\|_T = R(t, t)$ when $t \varepsilon T$ and $\|R(\cdot, t)\|_T \leqq R(t, t)$ when $t \not\varepsilon T$. Secondly, if $f \varepsilon \mathfrak{F}$ and $t \varepsilon T \varepsilon D_n$, $\|f + R(\cdot, t)\|_T^2 = \|f\|_T^2 + R(t, t) + 2f(t)$. Since, for $T \varepsilon D_{n-1}$, we have $\|f + R(\cdot, t)\|_{T \cup \{t\}}^2 = \|f\|_{T \cup \{t\}}^2 + R(t, t) + 2f(t) \geqq \|f\|_T^2 + R(t, t) + 2f(t)$ and

$$\sup_{T \varepsilon D_{n-1}} \|f + R(\cdot, t)\|_{T \cup \{t\}}^2 \leqq \|f + R(\cdot, t)\|_n^2 ,$$

it follows that $\|f + R(\cdot, t)\|_n^2 \geqq \|f\|_{n-1}^2 + R(t, t) + 2f(t)$. Similarly,

$$\|f + R(\cdot, t)\|_{n-1}^2 \leqq \|f\|_n^2 + R(t, t) + 2f(t),$$

and (2.4) is then concluded.

With regard to the estimation of the parameter β, we see that there are unbiased estimates of the form $\sum_{j=1}^n c_j Y(t_j)$ which have variances not exceeding $\|f\|_n^{-2} + 1/n$, $\|f\|_n^{-2}$ being taken as zero if $f \not\varepsilon \mathfrak{F}_n$. Let $Z_n = \sum_{j=1}^n c_{jn} Y(t_{jn})$, $n = 1, 2, \cdots$, be any such sequence of estimators.

THEOREM 2.3. *There is a random variable Z such that for each β, $E_\beta(Z_n - Z)^2 \to 0$ and Z is an unbiased estimator satisfying $\sigma_\beta^2(Z) = \|f\|^{-2}$, $\|f\|^{-2}$ being taken as zero if $f \not\varepsilon \mathfrak{F}$.*

PROOF. Consider the estimator $(Z_n + Z_{n+k})/2$ for β, based on at most $2n + k$ points.

$$\sigma_\beta^2[(Z_n + Z_{n+k})/2] = \sigma_\beta^2(Z_n)/4 + \sigma_\beta^2(Z_{n+k})/4 + \operatorname{cov}_\beta (Z_n, Z_{n+k})/2 \geqq \|f\|_{2n+k}^{-2}.$$

Thus

$$\begin{aligned} 2 \operatorname{cov}_\beta (Z_n, Z_{n+k}) &\geqq 4\|f\|_{2n+k}^{-2} - \sigma_\beta^2(Z_n) - \sigma_\beta^2(Z_{n+k}) \\ &\geqq 4\|f\|_{2n+k}^{-2} - \|f\|_n^{-2} - \|f\|_{n+k}^{-2} - 1/n - 1/(n+k). \end{aligned}$$

This yields

$$\begin{aligned} E_\beta(Z_n - Z_{n+k})^2 &= \sigma_\beta^2(Z_n) + \sigma_\beta^2(Z_{n+k}) - 2 \operatorname{cov}_\beta (Z_n, Z_{n+k}) \\ &\leqq 2\|f\|_n^{-2} + 2\|f\|_{n+k}^{-2} + 2/n + 2/(n+k) - 4\|f\|_{2n+k}^{-2}, \end{aligned}$$

which approaches 0 as $n, k \to \infty$. This completes the proof.

We conclude this development with some supplemental facts concerning $\mathfrak{F}$ and the random variable Z of Theorem 2.3. This material is standard and we omit any proofs. The continuity in mean imposed on $X(\cdot)$ is equivalent to, for each $t \varepsilon [0,1]$,

$$(2.5) \quad R(s, s) + R(t, t) - 2R(s, t) = \|R(\cdot, s) - R(\cdot, t)\|^2 \to 0 \quad \text{as} \quad s \to t.$$

Using (2.4) it follows that $\mathfrak{F}$ consists of continuous functions and is therefore a separable Hilbert space. $\mathfrak{F}$ is isomorphic to the L_2 subspace spanned by the random variables $\{X(t), t \varepsilon [0, 1]\}$ according to a mapping induced by

$$(2.6) \qquad \Psi[X(t)] = R(\cdot, t), \; t \varepsilon [0, 1].$$

If now $\{f_k\}$ is chosen to be a complete orthonormal family in $\mathfrak{F}$, then $f \varepsilon \mathfrak{F}$ may be written as $\sum_k a_k f_k$ and $X(\cdot)$ may be represented as $X(t) = \sum_k X_k f_k(t)$ with $\{X_k\}$ an orthonormal sequence of zero mean random variables. This enables us to define (Y, f) as $\sum_k (\beta a_k + X_k) a_k$ and we can thereby identify the estimator Z of Theorem 2.3 as $(Y, f)/\|f\|^2$. We turn to the question of optimum design.

An optimum design in D_n is a T^* which maximizes $\|f\|_T$ over D_n. If $f \varepsilon \mathfrak{F}$ we may rewrite $\|f\|_T$ as $\|P_T f\|$ where P_T is the projection operator defined on $\mathfrak{F}$ to the subspace spanned by the functions $R(\cdot, t)$, $t \varepsilon T$ (such subspaces are henceforth denoted by $V[R(\cdot, t), t \varepsilon T]$). To see that this replacement is valid, it suffices to note the isomorphism between $V[R(\cdot, t), t \varepsilon T]$ and the Hilbert space of n-vectors under the norm $\| \;\; \|_T$ which is induced by setting, for $t \varepsilon T$, $\Psi_T[R(\cdot, t)] = R(\cdot, t)_T$ (this last is the restriction of $R(\cdot, t)$ to T).

Consider now the family $\{P_T, T \varepsilon D_n\}$ of projections operating on $\mathfrak{F}$. It will be seen that the continuity in mean condition imposed on $X(\cdot)$ will insure that as $T \varepsilon D_n$ approaches $S \varepsilon D_n$, $P_T f \to P_S f$ for all $f \varepsilon \mathfrak{F}$. Thus $\|P_T f\|$ is a continuous function on D_n. In order that $\sup_{T \varepsilon D_n} \|P_T f\|$ be achieved, it is necessary to inspect the behaviour of $\|P_T f\|$ near the boundary of D_n. To this end, let $\bar{D}_n = \{T \mid T = \{t_1, t_2, \cdots, t_n\}, 0 \leqq t_1 \leqq t_2 \leqq \cdots \leqq t_n \leqq 1\}$. Sets $S \varepsilon \bar{D}_n - D_n$

may be given two interpretations as n-point designs: Observations are taken at the r distinct points of S, $r < n$, or, S represents the limiting case of n observations taken at $t_1 < t_2 < \cdots < t_n$ with $|s_i - t_i| < \epsilon$, $i = 1, 2, \cdots, n$. We do not know *a priori* that the second interpretation makes sense, that is, whether there is a single limiting case. Boundary designs can easily be optimum when given the second interpretation and when R has sufficient derivatives on the diagonal. For example, it can be shown that if $R(s, t) = \exp[-(s - t)^2/2]$ and $f(t) = t \exp(-t^2/2)$, then $\sup_{T \varepsilon D_2} \|P_T f\| = 1 = \|f\|$ and the supremum occurs not in D_2 but rather as $\epsilon \to 0$ in $T_\epsilon = \{0, \epsilon\}$. There is but a single limiting case in this example and a decrease in variance is achieved over the design which takes one observation at 0.

The difficulties inherent in such situations can be avoided by adding a further assumption on R. For $S \varepsilon D_n$, consider the subspaces of $\mathfrak{F}$ defined by $I_\epsilon(S) = V[R(\cdot, t), t \varepsilon \bigcup_{s \varepsilon S} (s - \epsilon, s + \epsilon)]$ and $I(S) = \bigcap_{\epsilon > 0} I_\epsilon(S)$. Inasmuch as each of the subspaces $I_\epsilon(S)$ contains $V[R(\cdot, t), t \varepsilon S]$, $I(S)$ does also. We are going to assume that R satisfies the

Definition. R is said to have a simple present if in $\mathfrak{F}$, $I(S) = V[R(\cdot, t), t \varepsilon S]$ for each $S \varepsilon D_n$, all n.

There is an equivalent (through use of (2.6)) definition phrased in terms of the error process $X(\cdot)$, and a simple present there, for example, disallows quadratic mean derivatives. When $X(\cdot)$ is stationary, a sufficient condition has been given for this behaviour (cf. [22]).

With these preliminaries, we are in a position to prove

Theorem 2.4. *If R has a simple present and if $\{T_k\}$ is any sequence from D_n with limit S, $P_{T_k} f \to P_S f$ for all $f \varepsilon \mathfrak{F}$.*

Proof. Since a weakly convergent subsequence can be extracted from the sequence $\{P_{T_k}\}$, we relabel it $\{P_{T_k}\}$ and suppose $P_{T_k} \to_w A$. If f is in the range of A, say $Ag = f$, then

$$\|P_{I_\epsilon(S)} f\|^2 = (P_{I_\epsilon(S)} f, f) = \lim_{k \to \infty} (P_{I_\epsilon(S)} f, P_{T_k} g) = \lim_{k \to \infty} (f, P_{T_k} g) = \|f\|^2;$$

that is, $f \varepsilon I_\epsilon(S)$ for all ϵ or, what is the same $f \varepsilon I(S) = V[R(\cdot, t), t \varepsilon S]$. Consider first the action of A on a function g which is orthogonal to $I(S)$. We have seen that g must be orthogonal to Ag and from $\|P_{T_k} g\|^2 = (g, P_{T_k} g) \to (g, Ag) = 0$ it follows that $P_{T_k} g$ actually converges in the usual sense to 0. Hence $Ag = 0$. Now let $s \varepsilon S$ and suppose t_k is drawn from T_k in such a way that the sequence $\{t_k\}$ converges to s. Using (2.5), we find

$$\|P_{T_k} R(\cdot, s)\|^2 \geqq \|P_{\{t_k\}} R(\cdot, s)\|^2 = R^2(s, t_k)/R(t_k, t_k) \to R(s, s)$$
$$= \|R(\cdot, s)\|^2 \quad \text{as} \quad k \to \infty.$$

Consequently, $\|R(\cdot, s) - P_{T_k} R(\cdot, s)\|^2 = \|R(\cdot, s)\|^2 - \|P_{T_k} R(\cdot, s)\|^2 \to 0$ as $k \to \infty$ and hence for all $f \varepsilon V[R(\cdot, s), s \varepsilon S] = I(S)$ we have $P_{T_k} f \to Af = f$. Clearly then, $A = P_S$ and the theorem is proved.

When R has a simple present, Theorem 2.4 insures that $\|P_T f\|$ is a continuous function in T over $\bar{D}_n$ for any $f \varepsilon \mathfrak{F}$, hence it achieves its maximum on $\bar{D}_n$. Further, if $S \varepsilon \bar{D}_n$ has $r < n$ distinct points and the maximum is achieved at S, any r-point design at the distinct points of S is optimum as is any n-point design containing it.

REMARK. An analogue to Theorem 2.4 may be given which applies to a much broader class of kernels R. If for a given R and all finite sets S, $I(S)$ is sufficiently well behaved that it contains only derivative functions $R^{(j)}(\cdot, s)$, $j = 0, 1, 2, \cdots, N(s)$, (these functions are the images under the mapping Ψ given in (2.6) of the quadratic mean derivatives of $X(\cdot)$ at s) then projections can be assigned to sets $S \varepsilon \bar{D}_n$ in such a way as to extend the continuity of the functions $\|P_T f\|$ to $\bar{D}_n$. However, the range of the boundary projection at S will generally contain $V[R(\cdot, t), t \varepsilon S]$ properly and if f is sufficiently close to derivative functions (in the sense of projection) optimum designs will occur on the boundary. For the example given prior to Theorem 2.4, the optimum design corresponds to projection onto the subspace spanned by $R(\cdot, 0)$ and $R^{(1)}(\cdot, 0)$.

3. Asymptotic optimality. In this section we derive our main asymptotic results for the problem of estimating the coefficient of a single regression function with observations restricted to be taken in the interval [0, 1]. Thus, our model will be

$$Y(t) = \beta f(t) + X(t); \qquad t \varepsilon [0, 1], \tag{3.1}$$

where β is unknown, $X(t)$ is a covariance process with mean 0 and covariance function $R(s, t)$, and f is a continuous function on [0, 1]. For the present, we will further restrict f to be of the form

$$f(t) = \int_0^1 R(s, t)\varphi(s)\, ds, \tag{3.2}$$

where φ is continuous on [0, 1] so that, in particular, f is in the reproducing kernel space $\mathfrak{F}$. For the arguments that follow it will be essential that $f \varepsilon \mathfrak{F}$ but, as will be noted in Remark 3.3 below, some of the arguments and results, including Theorem 3.1, will hold for somewhat more general f's in $\mathfrak{F}$ than exemplified in (3.2)

For any $f \varepsilon \mathfrak{F}$ (not necessarily of the form described in (3.2)) we will say that $\{T_n ; n \geqq 1\}$ is an asymptotically optimal sequence of designs for estimating β if

$$\lim_{n\to\infty} [(\|f\|^2 - \|P_{T_n} f\|^2)/(\|f\|^2 - \sup_{T\varepsilon D_n}\|P_T f\|^2)] = 1. \tag{3.3}$$

This definition is equivalent to

$$\lim_{n\to\infty} [(\operatorname{Var} \hat{\beta}_{T_n} - \operatorname{Var} \hat{\beta})/(\inf_{T\varepsilon D_n} \operatorname{Var} \hat{\beta}_T - \operatorname{Var} \hat{\beta})] = 1 \tag{3.4}$$

where $\hat{\beta}_T$ is the best linear estimate of β with observations taken in T and $\hat{\beta}$ is the BLE of β when observations are taken over the entire interval [0, 1].

In order to find such sequences $\{T_n\}$ we examine the asymptotic behavior of the denominator of (3.3). In the course of the development there will be some results which are of independent interest. These will be remarked on. We begin by listing the assumptions we require; a discussion of the assumptions follows their presentation.

ASSUMPTION A. R is continuous on the entire unit square and has continuous derivatives up to order two at every (s, t) in the complement of the diagonal in the unit square, i.e., at every (s, t) with $s \neq t$. At the diagonal we assume that R has all right and left derivatives up to order two.

Thus, for example,

$$\lim_{t \downarrow t_0, s \uparrow t_0} (\partial^2 R/(\partial t \partial s))(s, t) = \lim_{t \downarrow t_0, s \uparrow t_0} R_{21}(s, t) = R_{21}^{+-}(t_0, t_0)$$

is assumed to exist for all $t_0 \varepsilon [0, 1]$ (note that we have introduced the notation R_{21} to denote differentiation with respect to the second variable followed by differentiation with respect to the first variable).

Assumption A requires little discussion except to note that when f is of the form (3.2) and Assumption A holds then f is twice differentiable and in fact

$$f'(t) = \int_0^t R_2^+(s, t) \varphi(s) \, ds + \int_t^1 R_2^-(s, t)\varphi(s) \, ds \tag{3.5}$$

and

$$f''(t) = [R_2^+(t, t) - R_2^-(t, t)]\varphi(t) + \int_0^1 R_{22}^{++}(s, t)\varphi(s) \, ds. \tag{3.6}$$

ASSUMPTION B. Let $\alpha(t) = R_2^-(t, t) - R_2^+(t, t)$ for $t \varepsilon (0, 1)$. α is continuous on $(0, 1)$, $\inf_{0<t<1} \alpha(t) = \alpha_0 > 0$, and $\sup_{0<t<1} \alpha(t) = \alpha_1 < \infty$ so that α may be extended to a strictly positive continuous function on $[0, 1]$.

Examples which satisfy Assumptions A and B are easy to give. One class is with $R(s, t) = u(\min(s, t))v(\max(s, t))$ with $u'(t)v(t) - v'(t)u(t)$ never zero and u and v both having two continuous derivatives. Another class of examples is given by the stationary covariance functions of the form $1 - a|t - s| + \Psi(t - s)$ where $0 < a \leqq 1$ and Ψ has two continuous derivatives.

For all stationary covariance functions α is, of course, constant and because R is a covariance function α can never be negative. This last fact is generally true when $R(t, t) \equiv c$.

If $\alpha(t) \equiv 0$ and Assumption A is satisfied then R is differentiable everywhere in the unit square. We have not been able to extend our methods to such cases. Some of the difficulty may be seen in the example $R(s, t) = \exp[-(s - t)^2/2]$ (which is discussed above Theorem 2.4 in Section 2) where exactly optimum designs may not exist and asymptotically optimum designs will not possess the "nice" structure described in Theorem 3.1.

ASSUMPTION C. For each $t \varepsilon [0, 1]$ the function $R_{22}^{++}(\cdot, t) \varepsilon \mathfrak{F}$, and $\sup_{0 \leqq t \leqq 1} \|R_{22}^{++}(\cdot, t)\| < \infty$.

Assumption C obviously requires further explanation. Since every function in $\mathfrak{F}$ is continuous, C implies that $R_{22}^{++}(\cdot, t)$ is a continuous function for each t. Therefore, in particular, $R_{22}^{++}(t, t) = \lim_{s \downarrow t} R_{22}^{++}(s, t) = \lim_{s \downarrow t} R_{22}^{--}(s, t) = R_{22}^{--}(t, t)$ and this occurs despite the fact, due to Assumption B, that $R_{22}(\cdot, t)$ does not exist at t. It appears then that C is a rather restrictive assumption in the presence of B. Before turning to examples which satisfy C, as well as A and B, let us note here that the role of C is, primarily, to prove Lemma 3.2 below. In Remark 3.2 we discuss the difficulty in verifying the conclusion of Lemma 3.2 and why adequate conditions (other than C) are not easy to obtain.

As for examples which satisfy the three assumptions the easiest to see is $R(s, t) = \min(s, t)$ since $R_{22}^{++}(s, t) \equiv 0$. Another simple class of examples is given by $R(s, t) = 1 - \lambda |t - s|$ if $|t - s| \leqq 1/\lambda$ and $R(s, t) = 0$ otherwise (here $0 < \lambda \leqq 1$). $R_{22}^{++}(s, t) \equiv 0$ and thus Assumption C is trivially true.

Somewhat less apparent is the class of convex stationary covariance functions of the form

$$R(s, t) = \int_0^{1/|t-s|} \{1 - \lambda |t - s|\} p(\lambda)\, d\lambda \tag{3.7}$$

where p is a probability density satisfying

$$\lim_{\lambda\to\infty} \lambda^3 p(\lambda) = c < \infty, \tag{3.8}$$

$$\int_a^\infty [\lambda p'(\lambda) + 3p(\lambda)]^2 \lambda^6\, d\lambda < \infty, \tag{3.9}$$

for some a (we are assuming that the derivative p' of p exists). Assumption A is easy to verify. Assumption B is simple to check because $\alpha(0) = 2 \int_0^\infty \lambda p(\lambda)\, d\lambda$, which is finite by virtue of (3.8). That Assumption C is satisfied is not so obvious but follows from either Hájek [7] or Ylvisaker [23] where the norm structure of reproducing kernel spaces derived from covariance functions of the type (3.7) is discussed. In particular (3.8) guarantees that $R_{22}^{++}(0, 0)$ exists and (3.9) is equivalent to having $|\partial R_{22}^{++}(\cdot, t)/\partial s|^2$ integrable over some interval $[0, 1/a]$ which will guarantee that $R_{22}^{++}(\cdot, t)$ lies in the reproducing kernel space generated by (3.7). A subclass of these examples is given by

$$R(s, t) = \int_0^\infty \exp(-\lambda |t - s|)\, dP(\lambda)$$

with $\int_0^\infty \lambda^3\, dP(\lambda) < \infty$.

For the remainder of this section we will use the following notation: When $\{T_n ; n \geqq 1\}$ is a sequence of designs with $T_n \varepsilon D_n$, and $T_n = \{t_{1n}, \cdots, t_{nn}\}$ with $0 \leqq t_{1n} < t_{2n} < \cdots < t_{nn} \leqq 1$ we will put $t_{0n} = 0$, $t_{n+1,n} = 1$ and $d_{in} = t_{i+1,n} - t_{in}$ for $i = 0, \cdots, n$. In the proofs we will drop the index n in t_{in} and d_{in} when doing so causes no confusion.

LEMMA 3.1. *Let f be as in* (3.2) *and let Assumptions A and B be satisfied. Let* $\varphi_1 = \sup_{0\leqq t\leqq 1} |\varphi(t)|$, $K_1 = \sup_{0<s,t<1} |R_{21}^{-+}(s, t)|$, $K_2 = \sup_{0\leqq s,t\leqq 1} |R_{22}^{++}(s, t)|$. *Then, if $\{T_n ; n \geqq 1\}$ is a sequence of designs with $T_n \varepsilon D_n$*

$$\|f - P_{T_n} f\|^2 \leqq K \sup_{0\leqq j\leqq n} d_{jn}^2 \tag{3.10}$$

where, if $t_{1n} = 0$ *or* $t_{nn} = 1$, $K = \frac{1}{12}\varphi_1^2(8\alpha_1 + 3K_1 + 4K_2)$ *and if* $0 < t_{1n}$ *and* $t_{nn} < 1$, $K = \frac{1}{12}\varphi_1^2(20\alpha_1 + 3K_1 + 4K_2)$.

PROOF. Suppose, to begin with, that $t_1 = 0$ and $t_n < 1$. Let $\gamma_i = \int_{t_i}^{t_{i+1}} \varphi(s)\, ds$ for $i = 1, \cdots, n$ and put $g_n(t) = \sum_{i=1}^n R(t_i, t)\gamma_i$. Then

$$\|f - P_{T_n} f\|^2 \leqq \|f - g_n\|^2 = \int_0^1 [f(t) - g_n(t)]\varphi(t)\, dt - \sum_{i=1}^n [f(t_i) - g_n(t_i)]\gamma_i$$
$$= \sum_{i=1}^n \{\int_{t_i}^{t_{i+1}} [f(t) - g_n(t)]\varphi(t)\, dt - [f(t_i) - g_n(t_i)]\gamma_i\}.$$

Because of Assumption A we know that f is twice differentiable everywhere and g_n is twice differentiable except on T_n at each point of which, however, it has right and left derivatives.

Let $g'_{ni}(t_i) = g_n'^{+}(t_i)$, $g''_{ni}(t_{i+1}) = g_n''^{-}(t_{i+1})$, $g''_{ni}(t_i) = g_n''^{+}(t_i)$, and let $g''_{ni}(t) = g_n''(t)$ if $t \varepsilon (t_i, t_{i+1})$. Then, for $t \varepsilon (t_i, t_{i+1})$,

$$(3.11) \quad f(t) - g_n(t) = f(t_i) - g_n(t_i) + [f'(t_i) - g'_{ni}(t_i)](t - t_i) + [f''(\theta_t) - g_n''(\theta_t)](t - t_i)^2/2$$

where $\theta_t \varepsilon (t_i, t_{i+1})$. Now

$$(3.12) \quad g'_{ni}(t_i) = \textstyle\sum_{j=1}^n R_2^+(t_j, t_i)\gamma_j$$

and

$$(3.13) \quad f'(t_i) = \int_0^1 R_2^+(s, t_i)\varphi(s)\, ds = \textstyle\sum_{j\neq i} \int_{t_j}^{t_{j+1}} R_2^+(s, t_i)\varphi(s)\, ds + \int_{t_i}^{t_{i+1}} R_2^+(s, t_i)\varphi(s)\, ds.$$

When $j \neq i$,

$$(3.14) \quad \int_{t_j}^{t_{j+1}} R_2^+(s, t_i)\varphi(s)\, ds = R_2(t_j, t_i)\gamma_j + \int_{t_j}^{t_{j+1}} R_{21}(\rho_{sj}, t_i)(s - t_j)\varphi(s)\, ds$$

where $\rho_{sj} \varepsilon (t_j, t_{j+1})$, while

$$(3.15) \quad \begin{aligned} \int_{t_i}^{t_{i+1}} R_2^+(s, t_i)\varphi(s)\, ds &= \int_{t_i}^{t_{i+1}} R_2^-(s, t_i)\varphi(s)\, ds \\ &= \int_{t_i}^{t_{i+1}} R_{21}^{-+}(\rho_{si}, t_i)(s - t_i)\varphi(s)\, ds \\ &\quad + R_2^-(t_i, t_i)\gamma_i . \end{aligned}$$

Putting (3 .12), (3.13), (3.14), and (3.15) together, we get

$$(3.16) \quad \begin{aligned} |f'(t_i) - g'_{ni}(t_i)| &\leqq \alpha(t_i)\gamma_i + K_1\varphi_1 \textstyle\sum_{j=1}^n \int_{t_j}^{t_{j+1}} (t - t_j)\, dt \\ &\leqq \alpha_1\varphi_1 d_i + (K_1\varphi_1/2) \textstyle\sum_{i=1}^n d_i^2. \end{aligned}$$

To estimate $f''(\theta_t) - g''_{ni}(\theta_t)$ observe that from (3.6)

$$(3.17) \quad |f''(\theta)| \leqq \alpha_1\varphi_1 + K_2\varphi_1$$

for all θ and that

$$(3.18) \quad |g''_{ni}(\theta)| = |\textstyle\sum_{j=1}^n R_{22}^{++}(t_j, t_i)\gamma_j| \leqq K_2\varphi_1$$

for all $\theta \varepsilon [t_i, t_{i+1}]$ so that

$$(3.19) \quad |f''(\theta_t) - g''_{ni}(\theta_t)| \leqq (\alpha_1 + 2K_2)\varphi_1 .$$

(3.11), (3.16), and (3.19) yield

$$\begin{aligned} &\left|\int_{t_i}^{t_{i+1}} [f(t) - g_n(t)]\varphi(t)\, dt - [f(t_i) - g_n(t_i)]\gamma_i\right| \\ &\quad\leqq (\alpha_1\varphi_1 d_i + (K_1\varphi_1/2) \textstyle\sum_1^n d_j^2) \int_{t_i}^{t_{i+1}} (t - t_i)\varphi(t)\, dt \\ &\qquad + [(\alpha_1 + 2K_2)\varphi_1/2] \int_{t_i}^{t_{i+1}} (t - t_i)\varphi(t)\, dt \\ &\quad\leqq (K_1/4)\varphi_1^2(\textstyle\sum_{j=1}^n d_j^2)\, d_i^2 + \varphi_1^2[(2\alpha_1 + K_2)/3]\, d_i^3, \end{aligned}$$

which, used in (3 .10), yields

(3.20) $\|f - P_{T_n}f\|^2 \leqq (K_1\varphi_1^2/4)(\sum_1^n d_j^2)^2 + [(2\alpha_1 + K_2)/3]\varphi_1^2 \sum_1^n d_j^3.$

The conclusion of Lemma 3.1 for the case when $t_{1n} = 0$, $t_{nn} < 1$ follows immediately.

In the case $t_{1n} > 0$ and $t_{nn} = 1$ a similar argument will work by expanding $f(t) - g_n(t)$ for $t \varepsilon (t_i, t_{i+1})$, in a Taylor series around the point t_{i+1} (in the above we expanded in a Taylor series around t_i). When $t_{in} = 0$ and $t_{nn} = 1$ we have $\|f - P_{T_n}f\|^2 \leqq \|f - P_{T_{n-1}}f\|^2$ where $T_{n-1} = \{t_{1n}, \cdots, t_{n-1,n}\}$ and we can apply the result for $n - 1$ in the case $t_1 = 0$ and $t_{n-1} < 1$. When $t_{1n} > 0$ and $t_{nn} < 1$ an analogous argument can be carried out as follows: let $\gamma_1 = \int_0^{t_2} \varphi(t)\,dt$ and $\gamma_2, \cdots, \gamma_n$ be defined as above. To estimate $\int_0^{t_2} (f - g_n)\varphi$ expand $f - g_n$ around t_1 and estimate $\int_{t_i}^{t_{i+1}} (f - g_n)\varphi\,dt$ for $i \geqq 2$ as in the above. The details will only slightly differ and will yield the result stated. The mild complications we run into in this last case are due to the need of estimating $\int (f - g_n)\varphi$ on $n + 1$ intervals with the use of a linear combination of $R(t_1, \cdot), \cdots, R(t_n, \cdot)$.

REMARK 3.1. Considering only the case where $t_{1n} = 0$ it is clear from (3 .20) that there are designs for which $\|f - P_{T_n}f\|^2 \leqq K/n^2$; in particular, take $T_n = \{0, 1/(n-1), \cdots, (n-1)/n\}$. (In fact, since $\sum_{j=1}^n d_j^2 \geqq \sum_1^n d_j \cdot 1/n = 1/n$ and $\sum_1^n d_j^3 \geqq (\sum d_j)^3 1/n^2 = 1/n^2$, this choice of $\{T_n\}$ minimizes the right hand side of (3.20).) Thus $\inf_{T_n \varepsilon D_n} \|f - P_{T_n}f\|^2 \leqq K/n^2$ is a consequence of Lemma 3.1. Although the constant K may be improved, the order of magnitude $1/n^2$ cannot, in general, be improved for the types of covariance functions and f's we are concerned with. This will be seen in the theorem below.

Lemma 3.1 besides being of use in the development of this section has independent interest in that it provides us with some idea of how many observations one needs to take before additional observations yield trivial gain. Thus, if we measure the (relative) gain in taking more than n observations by $\gamma_n = (\operatorname{Var} \hat{\beta}_n - \operatorname{Var} \hat{\beta})/\operatorname{Var} \hat{\beta}$, then Lemma 3.1 provides an upper bound on γ_n, namely, $\gamma_n \leqq K/n^2(\|f\|^2 - K/n^2)$. This use is heightened by the fact that Lemma 3.1 does not require Assumption C.

LEMMA 3.2. *Let f be as in* (3.2) *and suppose that Assumptions A, B, and C are satisfied. Let $\{T_n\}$ be a sequence of designs with $T_n \varepsilon D_n$ and such that* $\sup_{0 \leqq j \leqq n} d_{jn} \to 0$ *as $n \to \infty$. Let $g_n = P_{T_n}f$ so that $g_n(t) = \sum_{i=1}^n R(t_{in}, t)m_{in}$ (the m_{in} being determined by the fact that $g_n(t_{in}) = f(t_{in})$ for $i = 1, \cdots, n$). Then*

(3.21) (a) $m_{1n} \to 0$, $m_{nn} \to 0$,

and

(b) $m_{in} = \varphi(t_{in})(d_{i-1,n} + d_{in})/2 + a_{in}\, d_{i-1,n} + b_{in}\, d_{in}$

$(i = 2, \cdots, n-1)$ *where* $\sup_{2 \leqq i \leqq n-1} \{|a_{in}| + |b_{in}|\} \to 0$ as $n \to \infty$.

It follows from (3.21) *that*

(3.22) $$\sup_{n \geqq 1} \sum_{i=1}^n |m_{in}| < \infty$$

and that $F_n(t) = \sum_{i \mid t_i \leq t} m_{in}$ *defines a sequence* $\{F_n\}$ *of functions of uniformly bounded variation which converges weakly to* F *which is defined by* $F(t) = \int_0^t \varphi(s)\, ds$.

PROOF. First observe that, due to Assumptions A and B, whenever $t_i \varepsilon (0, 1)$,

$$g_n'^-(t_i) - g_n'^+(t_i) = \alpha(t_i)m_i, \qquad \alpha(t_i) > 0. \tag{3.23}$$

From Assumption C (and the comments following Assumption C) we can talk about a function $R_{22}(s, t)$ which is defined for all (s, t) in the unit square and agrees with the second partial derivative of R whenever the latter exists. Thus $g_n''^+(t_i) = g_n''^-(t_i) = \sum_{j=1}^n R_{22}(t_j, t_i)m_j$; $i = 1, \cdots, n$. It follows that for $i = 1, \cdots, n-1$,

$$\begin{aligned} f(t_{i+1}) &- g_n(t_{i+1}) \\ &= f(t_i) - g_n(t_i) + [f'(t_i) - g_n'^+(t_i)]\, d_i + [f''(\sigma_i) - g_n''(\sigma_i)]\, d_i^2/2, \end{aligned}$$

where $\sigma_i \varepsilon (t_i, t_{i+1})$. From this we conclude

$$f'(t_i) - g_n'^+(t_i) = -(d_i/2)[f''(\sigma_i) - g_n''(\sigma_i)]; \quad i = 1, \cdots, n-1. \tag{3.24}$$

Similarly

$$f'(t_i) - g_n'^-(t_i) = (d_{i-1}/2)[f''(\theta_i) - g_n''(\theta_i)]; \; i = 2, \cdots, n, \tag{3.25}$$

where $\theta_i \varepsilon (t_{i-1}, t_i)$. From (3.6) and Assumption C we obtain

$$f''(t) - g_n''(t) = -\alpha(t)\varphi(t) + \langle R_{22}(\cdot, t), f - g_n \rangle, \tag{3.26}$$

where $\langle\, , \rangle$ denotes the scalar (or inner) product in $\mathfrak{F}$. Subtracting (3.25) from (3.24) and using (3.23) and (3.26) we obtain, for $i = 2, \cdots, n-1$

$$\begin{aligned} \alpha(t_i)m_i &= \alpha(\theta_i)\varphi(\theta_i)\, d_{i-1}/2 + \alpha(\sigma_i)\varphi(\sigma_i)\, d_i/2 \\ &\quad - \langle R_{22}(\cdot, \theta_i), f - g_n \rangle\, d_{i-1}/2 - \langle R_{22}(\cdot, \sigma_i), f - g_n \rangle\, d_i/2. \end{aligned} \tag{3.27}$$

Using the continuity of φ on $[0, 1]$ ((3.2)), the continuity and positivity of α on $[0, 1]$ (Assumption B), the boundedness of $\|R_{22}(\cdot, t)\|$ (Assumption C), and the result that $\|f - g_n\| \to 0$ as $\sup d_{jn} \to 0$ (Lemma 3.1) we obtain (b) of (3.21). To obtain m_1 and m_n after obtaining $m_2, \cdots, m_{n-1}$ is easy since

$$m_1R(t_1, x) + m_nR(t_n, x) = f(x) - \sum_{j=2}^{n-1} R(t_j, x)m_j \tag{3.28}$$

for $x = t_1, t_n$. From (3.21) (b), the continuity of φ and R, we obtain

$$\sum_{j=2}^{n-1} R(t_j, x)m_j = \int_0^1 \varphi(t)R(t, x)\, dt + o(1) = f(x) + o(1), \tag{3.29}$$

where the $o(1)$ term goes to 0 as $n \to \infty$. Solving (3.28) for m_1 and m_n and using (3.29) yields (3.21) (a). The rest of Lemma 3.2 follows in obvious fashion from (3.21).

REMARK 3.2. It is relevant for the proof of Lemma 3.3 that we point out that (3.27) is valid without $\sup d_{jn} \to 0$.

The part of Lemma 3.2 which is most relevant for us is (3.22). With this last fact it would be unnecessary to have Assumption C. Although this doesn't seem apparent from the proof below of Theorem 3.1 we point out here that we could with some modification alternatively prove Theorem 3.1 with Assumption C replaced by (3.22). The difficulty we have is finding adequate conditions for the validity of (3.22). That the question of the validity of (3.22) seems to be nontrivial may be seen in the following: Consider the integral equation of the first kind

$$f(t) = \int_0^1 R(s, t)\varphi(s)\, ds, \qquad t \varepsilon [0, 1], \tag{3.30}$$

where f is given, R satisfies Assumptions A and B (and is a positive definite function, i.e., is a covariance function), and φ is unknown. To solve (3.30) numerically we might proceed by taking n points $t_1, \cdots, t_n$ and solving for $m_1, \cdots, m_n$ in

$$f(t_i) = \sum_{j=1}^{n} R(t_j, t_i)m_j. \tag{3.31}$$

The question of how $m_1, \cdots, m_n$ is related to φ is critical in deciding on the utility of (3.31) for numerical approximations of φ. The literature seems to be silent on numerical solutions of (3.30) except to remark on the difficulties of obtaining such. Now, if (3.22) were true, so that $\{F_n\}$, as defined following (3.22), would be of uniformly bounded variation, it would follow from

$$\int_0^1 R(s, t)\, d(F - F_n) = \|f - P_{T_n}f\|^2 \rightarrow 0,$$

that $F_n \rightarrow F$ weakly and, consequently, (3.31) could be used to obtain an approximate solution of (3.30) in an obvious sense. With Assumption C we have the somewhat stronger conclusion (and more convenient approximation) contained in (3.21) although the "error term" in the approximation is elusive, since the best we can do here is to give bounds on the error term which depend on φ (of course, the same difficulty occurs in estimating the difference between F_n and F).

LEMMA 3.3. *Suppose Assumptions A, B, and C are satisfied and that f is given by* (3.2) *where φ is such that $\{t \mid \varphi(t) = 0\}$ does not contain an interval of positive length. If $\{T_n\}$ is a sequence of designs with $T_n \varepsilon D_n$ then $\|f - P_{T_n}f\| \rightarrow 0$ if and only if*

$$\sup_{0 \leqq j \leqq n} d_{jn} \rightarrow 0.$$

PROOF. The "if" part is a consequence of Lemma 3.1, so we need only concern ourselves with the "only if" part.

If $\sup_j d_{jn}$ does not approach 0 then there is a positive number ϵ, a number $c\epsilon(0, 1 - \epsilon)$ such that $\varphi(c) \neq 0$, and a sequence $\{n_k\}$ such that $T_{n_k} \subset [0, c] \cup [c + \epsilon, 1]$ (thus $\sup_j d_{jn_k} \geqq \epsilon$). We might as well assume that $\{n_k\}$ is the sequence of all positive integers. Put $C = [0, c] \cup [c + \epsilon, 1]$. Then, since $T_n \subset C$ for all n we have $\|f - P_{T_n}f\|^2 = \|f - P_C f\|^2 + \|P_C f - P_{T_n}f\|^2 \geqq \|f - P_C f\|^2$ which implies that $\|f - P_C f\|^2 = 0$, i.e., that $f \varepsilon \mathfrak{F}(C)$. Let $C_\eta = [0, c] \cup [c + \eta, 1]$ for $0 \leqq \eta \leqq \epsilon$. Then $f \varepsilon \mathfrak{F}(C_\eta)$ for all $0 \leqq \eta \leqq \epsilon$.

Fix η. We can find a sequence $\{S_n\}$ of designs with $S_n \varepsilon D_n$ for $n \geqq 4$ such that $S_n \subset C_\eta$, $S_n \subset S_{n+1}$, $\{0, c, c + \eta, 1\} \subset S_n$, and $\|f - P_{S_n}f\|^2 \rightarrow 0$. In addition, we can choose S_n so that, putting $S_n = \{\sigma_{1n}, \cdots, \sigma_{nn}\}$, $\delta_{in} = \sigma_{i+1,n} - \sigma_{in}$, $\sigma_{i^*n} = \eta$ (i^* actually depends on n but we suppress it), $\sup_{i \neq i^*} \delta_{in} \to 0$. In fact, we can choose $\{S_n\}$ so that, for each n, there is a ν_n with $T_n \subset S_{n+k}$ for all $k \geqq \nu_n$. Since (3.27) is valid under the present hypotheses (see the first sentence of Remark 3.2) we have, for $i \neq 1, i^*, i^* + 1, n$,

$$\mu_i = \varphi(\sigma_i)(\delta_{i-1} + \delta_i)/2 + o(1)[\delta_{i-1} + \delta_i]$$

where $f(\sigma_i) = \sum_{j=1}^{n} R(\sigma_j, \sigma_i)\mu_j$, $i = 1, \cdots, n$. Thus, for all $x \varepsilon [0, 1]$,

$$\lim_{n\to\infty} \sum_{i=2}^{i^*-1} \mu_i R(\sigma_i, x) = \int_0^c R(\sigma, x)\varphi(\sigma)\, d\sigma, \tag{3.32}$$

$$\lim_{n\to\infty} \sum_{i=i^*+2}^{n-1} \mu_i R(\sigma_i, x) = \int_{c+\eta}^1 R(\sigma, x)\varphi(\sigma)\, d\sigma.$$

Now

$$\mu_1 R(0, x) + \mu_{i^*} R(c, x) + \mu_{i^*+1} R(c + \eta, x) + \mu_n R(1, x) = f(x) - [\textstyle\sum_{j=1}^{i^*-1} \mu_j R(\sigma_j, x) + \sum_{i^*+2}^{n-1} \mu_j R(\sigma_j, x)] \tag{3.33}$$

for $x \varepsilon S_n$. From (3.27) we can conclude that μ_{i^*n} and $\mu_{i^*+1,n}$ are bounded in n, and, with the aid of (3.33), we also have that μ_{1n} and μ_{nn} are bounded in n. Hence, we can extract a subsequence along which $\mu_{in}, \mu_{i^*n}, \mu_{i^*+1,n}, \mu_{nn}$ converge respectively to limits $\lambda_0, \lambda_c, \lambda_{c+\eta}, \lambda_1$. The subsequence may as well be taken to be the original sequence. Since $P_{S_n}f(x) \to f(x)$ for each $x \varepsilon [0,1]$ (this because $\|f - P_{S_n}f\| \to 0$), we can use (3.32), (3.2), and the existence of $\lambda_0, \lambda_c, \lambda_{c+\eta}, \lambda_1$ to conclude that

$$f(x) - P_{S_n}f(x) = f(x) - \textstyle\sum_{j=1}^{n} \mu_j R(\sigma_j, x) \to$$
$$\int_c^{c+\eta} R(\sigma, x)\varphi(\sigma)\, d\sigma - \lambda_0 R(0, x) - \lambda_c R(c, x) - \lambda_{c+\eta} R(c + \eta, x) - \lambda_1 R(1, x)$$

and, therefore.

$$\lambda_0 R(0, x) + \lambda_c R(c, x) + \lambda_{c+\eta} R(c + \eta, x) + \lambda_1 R(1, x) = \int_c^{c+\eta} R(\sigma, x)\varphi(\sigma)\, d\sigma \tag{3.34}$$

The differentiability of the right side of (3.34) at c and $c+\eta$ and the non-differentiability (Assumption B) of $R(c, \cdot)$ at c and $R(c + \eta, \cdot)$ at $c + \eta$ implies that $\lambda_c = \lambda_{c+\eta} = 0$. Hence, for all $x \varepsilon [0, 1]$,

$$\lambda_0 R(0, x) + \lambda_1 R(1, x) = \int_c^{c+\eta} R(\sigma, x)\varphi(\sigma)\, d\sigma. \tag{3.35}$$

The dependence of λ_0 and λ_1 on η can be determined by putting $x = 0, 1$ in (3.35) and solving the two linear equations in λ_0 and λ_1. This yields

$$\lim_{\eta\to 0} (\lambda_0(\eta)/\eta) = \varphi(c)p_0, \qquad \lim_{\eta\to 0} (\lambda_1(\eta)/\eta) = \varphi(c)p_1$$

(p_0 and p_1 are easily calculated).

This, used in (3.35) after dividing (3 .35) by η and letting $\eta \to 0$, gives, for all $x \varepsilon [0, 1]$, $\varphi(c)p_0R(0, x) + \varphi(c)p_1R(1, x) = R(c, x)\varphi(c)$. Since $\varphi(c)$ is not 0, $R(c, \cdot)$ is not differentiable at c, and $R(0, \cdot)$ and $R(1, \cdot)$ are differentiable at c, we have a contradiction and Lemma 3.3 is proved.

THEOREM 3.1. *Let Assumptions A, B, and C be satisfied and suppose that for* $t \in [0, 1]$,

$$f(t) = \int_0^1 R(s, t)\varphi(s)\, ds \tag{3.36}$$

with φ *continuous on* [0, 1]. *Then*

$$\lim_{n\to\infty} n^2 \inf_{T \in D_n} \|f - P_T f\|^2 = \tfrac{1}{12}\{\int_0^1 [\alpha(t)\varphi^2(t)]^{\frac{1}{3}}\, dt\}^3. \tag{3.37}$$

Furthermore, the sequence $\{T_n^*\}$ *defined by*

$$\int_0^{t^*_{in}} [\alpha(t)\varphi^2(t)]^{1/3}\, dt = [(i - 1)/(n - 1)] \int_0^1 [\alpha(t)\varphi^2(t)]^{1/3}\, dt, \tag{3.38}$$

$i = 1, \cdots, n$, *where* t^*_{in} *is the smallest solution to* (3.38) (*ambiguity will occur when* φ *is* 0 *on an interval*) *is an asymptotically optimal sequence of designs.*

PROOF. Suppose first that φ is never 0 on [0, 1]. Our first step will be to show that, if $\{T_n ; n \geqq 1\}$ is a sequence of designs with $t_{1n} = 0$, $t_{nn} = 1$ for all $n \geqq 2$, and with $\sup_j d_{jn} \to 0$ then,

$$\liminf_{n\to\infty} n^2\|f - P_{T_n}f\|^2 \geqq \gamma^3/12 \tag{3.39}$$

where γ is the value of the term in braces on the right side of (3.37). If we do so, and if $\{T_n\}$ is *any* sequence of designs with $\sup_j d_{jn} \to 0$ then, by adjoining 0 and 1 to T_n (when they aren't present), we form a sequence $\{S_n\}$ with $S_n \in D_{n+2}$ and satisfying (3.39). Since $\|f - P_{S_n}f\|^2 \leqq \|f - P_{T_n}f\|^2$ we conclude that (3.39) holds for all $\{T_n\}$ with $\sup_j d_{jn} \to 0$. With the aid of Lemma 3.3 (which enables us to ignore $\{T_n\}$ for which $\sup_j d_{jn}$ does not go to 0) we find that

$$\liminf_{n\to\infty} n^2 \inf_{T \in D_n} \|f - P_T f\|^2 \geqq \gamma^3/12. \tag{3.40}$$

Our second step will be to take the sequence $\{T_n^*\}$ given by (3.38) and show that

$$\lim_{n\to\infty} n^2\|f - P_{T_n^*}f\|^2 = \gamma^3/12. \tag{3.41}$$

(3.40) and (3.41) then yield (3.37) and the asymptotic optimality of $\{T_n^*\}$ follows from (3.37) and (3.41). The third part of the proof will be to remove the restriction that φ is never 0. We now turn to the first part, namely, the proof of (3.39).

Let $g_n = P_{T_n}f$. Then

$$\|f - g_n\|^2 = \int_0^1 [f(t) - g_n(t)]\varphi(t)\, dt = \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} (f - g_n)\varphi\, dt. \tag{3.42}$$

Since $f(t_i) = g_n(t_i)$ for $i = 1, \cdots, n$, and since (3.24) and (3.26) in the proof of Lemma 3.2 are valid, we can expand $f(t) - g_n(t)$, for $t \in (t_i, t_{i+1})$ in a Taylor series and obtain

$$\begin{aligned} f(t) - g_n(t) &= (-d_i/2)[f''(\sigma_i) - g_n''(\sigma_i)](t - t_i) \\ &\quad + [f''(\sigma_t) - g_n''(\sigma_t)](t - t_i)^2/2 \\ &= (d_i/2)\alpha(\sigma_i)\varphi(\sigma_i)(t - t_i) - \alpha(\sigma_t)\varphi(\sigma_t)(t - t_i)^2/2 \\ &\quad - (d_i/2)\langle R_{22}(\cdot, \sigma_i), f - g_n\rangle(t - t_i) \\ &\quad + \langle R_{22}(\cdot, \sigma_t), f - g_n\rangle(t - t_i)^2/2 \end{aligned} \tag{3.43}$$

where σ_t and σ_i are in (t_i, t_{i+1}). For $i = 1, \cdots, n-1$ put

$$\begin{aligned} A_{1i} &= \int_{t_i}^{t_{i+1}} (t - t_i)[\varphi(t) - \varphi(\sigma_i)]\, dt, \\ A_{2i} &= \int_{t_i}^{t_{i+1}} [\varphi(t) - \varphi(\sigma_i)][(t - t_i)^2/2]\, dt, \\ A_{3i} &= \int_{t_i}^{t_{i+1}} [\alpha(\sigma_t)\varphi(\sigma_t) - \alpha(\sigma_i)\varphi(\sigma_i)]\varphi(t)[(t - t_i)^2/2]\, dt, \\ A_{4i} &= (d_i/2)\langle R_{22}(\cdot, \sigma_i), f - g_n\rangle \int_{t_i}^{t_{i+1}} \varphi(t)(t - t_i)\, dt, \\ A_{5i} &= \int_{t_i}^{t_{i+1}} \langle R_{22}(\cdot, \sigma_t), f - g_n\rangle[(t - t_i)^2/2]\varphi(t)\, dt, \\ B_{1i} &= \sup_{t_i \leqq s, t \leqq t_{i+1}} |\varphi(t) - \varphi(s)|, \\ B_{2i} &= \sup_{t_i \leqq s, t \leqq t_{i+1}} |\alpha(s)\varphi(s) - \alpha(t)\varphi(t)|. \end{aligned} \tag{3.44}$$

All these quantities depend on n, of course. From the continuity of φ and α on $[0, 1]$ we have, as $n \to \infty$,

$$\sup_{1 \leqq i \leqq n-1} B_{1i} = o(1), \qquad \sup_{1 \leqq i \leqq n-1} B_{2i} = o(1). \tag{3.45}$$

An obvious calculation shows

$$|A_{1i}| \leqq B_{1i}\, d_i^2/2, \quad |A_{2i}| \leqq B_{1i}\, d_i^3/6, \quad |A_{3i}| \leqq B_{2i}\varphi_1\, d_i^3/6 \tag{3.46}$$

where $\varphi_1 = \sup_{0 \leqq t \leqq 1} |\varphi(t)|$. Assumption C and Lemma 3.1 yield

$$|A_{4i}| + |A_{5i}| \leqq K_0\, d_i^3(\sup_j d_j) \tag{3.47}$$

where K_0 is some constant. Using (3.44) in (3.43) and then using (3.46) and (3.47) we obtain

$$\begin{aligned} \int_{t_i}^{t_{i+1}} [f(t) - g_n(t)]\varphi(t)\, dt &= \alpha(\sigma_i)\varphi^2(\sigma_i)d_i^3/4 + \alpha(\sigma_i)\varphi(\sigma_i)\, d_i A_{1i} \\ &\quad - \alpha(\sigma_i)\varphi(\sigma_i)A_{2i} - \alpha(\sigma_i)\varphi^2(\sigma_i)d_i^3/6 - A_{3i} \\ &\quad - A_{4i} + A_{5i} \\ &\geqq \alpha(\sigma_i)\varphi^2(\sigma_i)\, d_i^3/12 \\ &\quad - d_i^3[K_1B_{1i} + K_2B_{2i} + K_0 \sup_j d_j] \end{aligned} \tag{3.48}$$

for appropriate constants K_1 and K_2. Let

$$\rho_n = \sup_{1 \leqq i \leqq n-1} [K_1B_{1i} + K_2B_{2i} + K_0 \sup_j d_j].$$

From (3.45) and the fact that $\sup d_j \to 0$ we know that $\rho_n \to 0$, and, since α and φ are continuous and never 0, we have, for all n large enough,

$$\alpha(\sigma_i)\varphi^2(\sigma_i)/12 - \rho_n \geqq 0 \quad \text{for } i = 1, \cdots, n-1. \tag{3.49}$$

Using (3.49) in (3.48), referring to (3.42), then using a Hölder inequality and the Riemann integrability of $\alpha\varphi^2$ we obtain, for n large enough,

$$\begin{aligned} \|f - g_n\|^2 &\geqq \sum_{i=1}^{n-1} [\alpha(\sigma_i)\varphi^2(\sigma_i)/12 - \rho_n]\, d_i^3 \\ &\geqq [1/(n-1)^2]\{\sum_{i=1}^{n-1} [\alpha(\sigma_i)\varphi^2(\sigma_i)/12 - \rho_n]^{1/3}\, d_i\}^3 \\ &= [1/12(n-1)^2]\{\int_0^1 [\alpha(t)\varphi^2(t)]^{1/3}\, dt\}^3 + o(1/n^2) \\ &= [1/12\, n^2]\gamma^3 + o(1/n^2). \end{aligned} \tag{3.50}$$

Thus (3.39) is established.

We turn now to the asymptotic behavior of $\|f - P_{T_n^*}f\|^2$ where T_n^* is given in (3.38). From (3.38) and the mean value theorem it is easy to verify that $\gamma/(n-1) = \int_{t_i^*}^{t_{i+1}^*} [\alpha\varphi^2]^{1/3}\, dt = d_i^*[\alpha(\theta_i)\varphi^2(\theta_i)]^{1/3}$ for some $\theta_i \varepsilon (t_i^*, t_{i+1}^*)$. A calculation similar to that leading to (3.48) (in fact, we could use the first part of (3.48) with σ_i replaced by any $\sigma \varepsilon (t_i, t_{i+1})$ and with the A's suitably modified although still satisfying (3.46) and (3.47)) yields

$$(3.51)\quad \|f - P_{T_n^*}f\|^2 \leqq (\gamma^3/12) \sum_{i=1}^{n-1} [1/(n-1)^3] + o(1/n^2) = \gamma^3/12n^2 + o(1/n^2).$$

(3.51) and (3.39) yield (3.41) and thus, as noted following (3.41), yield (3.37).

The validity of (3.37) when $Z = \{t \mid \varphi(t) = 0\}$ is not empty requires further argument because (3.49) may not be valid for all i when φ has zeroes, and the subsequent use of the Hölder inequality may not make sense. However, we can proceed as follows: Since φ is continuous, Z is compact. Hence, for each $\epsilon > 0$, there are open intervals $I_j = (a_j, b_j)$, $1 \leqq j \leqq k_\epsilon$ whose closures are disjoint and such that $A_\epsilon = \cup I_j \supset Z$ and $\sum \mu(I_j) \leqq \mu(Z) + \epsilon$ (μ is Lebesgue measure). If $T = \{\tau_1, \cdots, \tau_r\}$ let $m(T) = \max_{1 \leqq j \leqq r-1} (\tau_{j+1} - \tau_j)$. Let $b_0 = 0$, $a_{k_\epsilon+1} = 1$. Let $T_{nj} = T_n \cap [b_j, a_{j+1}]$, $j = 0, \cdots, k_\epsilon$. Let $\hat{T}_{nj} = T_{nj} \cup \{b_j, a_{j+1}\}$. We can show, as in Lemma 3.3, that $\|f - P_{T_n}f\|^2 \to 0$ if, and only if, for each choice of ϵ, $\{I_j\}$ satisfying the above, $\max_{0 \leqq j \leqq k_\epsilon} m(\hat{T}_{nj}) \to 0$. This agrees with Lemma 3.3 when Z contains no intervals. Let us suppose then that we are dealing with such a sequence $\{T_n\}$ and let us suppose that $\epsilon > 0$, $\{I_j\}$ etc. are as above and, in addition, that $|\varphi(t)| < \epsilon$ on A_ϵ. Since $[0, 1] - A_\epsilon = H_\epsilon$ is compact, we have φ bounded away from 0 on H_ϵ. We note here that $\gamma_\epsilon = \int_{H_\epsilon} [\alpha(t)\varphi^2(t)]^{1/3}\, dt \to \gamma$ as $\epsilon \to 0$ (γ is defined in (3.39)). Let us adjoin to the design T_n all the endpoints of the intervals $I_1, \cdots, I_{k_\epsilon}$ and, in addition, let us adjoin at most $[\mu(I_j)n]$ ($[x]$ is the greatest integer less than x) points from each I_j so that, in the new design so formed (call it S_n), any point in $S_n \cap I_j$ is at most $1/n$ from its immediate predecessor and its immediate successor. Let $J_n = \{i \mid s_i \varepsilon H_\epsilon, s_{i+1} \varepsilon H_\epsilon, s_i$ is not a left endpoint of one of the I_j's$\}$. Now $\nu_n = \text{Card } J_n \leqq \text{Card } (T_n \cap H_\epsilon) + k_\epsilon \leqq n + k_\epsilon$. Let $h_n = P_{s_n}f$. Then, for $i \varepsilon J_n$, (3.48) and (3.49) are valid with h_n replacing g_n, S_n replacing T_n, so that, for n large enough,

$$(3.52)\quad \begin{aligned} \sum_{i\varepsilon J_n} \int_{s_i}^{s_{i+1}} (f - h_n)\varphi\, dt &\geqq \sum_{i\varepsilon J_n} [\alpha(\sigma_i)\varphi^2(\sigma_i)/12 - \rho_n]\, d_i^3 \\ &\geqq (1/12\cdot\nu_n^2)\{\textstyle\int_{H_\epsilon} [\alpha(t)\varphi^2(t)]^{1/3}\, dt\}^3[1 + o(1)] \\ &= (1/12\cdot\nu_n^2)\gamma_\epsilon^3[1 + o(1)] \\ &\geqq (1/12\cdot(n + k_\epsilon)^2]\gamma_\epsilon^3[1 + o(1)]. \end{aligned}$$

Now, if s_1 and s_2 are two successive points in S_n with either s_1 or s_2 in A_ϵ or s_1 a left endpoint of one of the I_j's, we have $s_2 - s_1 \leqq 1/n$ and, from the first part of (3.48),

$$\left|\int_{s_1}^{s_2} [f(t) - h_n(t)]\varphi(t)\, dt\right| \leqq r_\epsilon\, (1/n^3)$$

where $r_\epsilon \to 0$ as $\epsilon \to 0$. There are at most $\sum_{j=1}^{k_\epsilon} ([\mu(I_j)n] + 1) \leqq k_\epsilon + (\mu(Z) + \epsilon)n$ such pairs s_1, s_2, so that

$$(3.53) \qquad |\int_{A_\epsilon} (f - h_n)\varphi\, dt| \leqq r_\epsilon(\mu(Z) + \epsilon)/n^2 + r_\epsilon k_\epsilon/n^3.$$

Combining (3.52) and (3.53) we obtain

$$(3.54) \qquad n^2\|f - P_{S_n}f\|^2 \geqq (\gamma_\epsilon^3/12)[1 + o(1)] - r_\epsilon(\mu(Z) + \epsilon)$$

and, since $T_n \subset S_n$, $\liminf_{n\to\infty} n^2\|f - P_{T_n}f\|^2 \geqq \gamma^3/12$. Thus

$$\liminf_{n\to\infty} n^2 \inf_{T\epsilon D_n} \|f - P_T f\|^2 \geqq \gamma^3/12.$$

Since (3.51) is valid whether or not φ has zeroes, we are then able to conclude that (3.37) holds and that $\{T_n^*\}$ is asymptotically optimal. This concludes the proof of Theorem 3.1.

REMARK 3.3. Theorem 3.1 may be extended to f's which are of the form

$$(3.55) \qquad f(t) = \int_0^1 R(s, t)\varphi(s)\, ds + \sum_{k=1}^{L} \mu_k R(a_k, t)$$

where the μ_k's are not zero and the a_k's are all in [0, 1]. φ is still assumed to be continuous. In this case, to obtain asymptotically optimal designs we should adjoin, to the designs defined by (3.38), the points $a_1, \cdots, a_L$. Often, as in the examples discussed in Section 4 (in particular, see (4.9) et seq.), $L = 2$ and $a_1 = 0$, $a_2 = 1$, so that nothing has to be adjoined to the designs of (3.38) to obtain asymptotically optimal designs.

That there is something to be done in order to extend the results to f's of the form (3.55) is due to the non-differentiability of f at $a_1, \cdots, a_L$. However, f does possess right and left derivatives at these points and this makes it possible to carry out the extension. We will omit the details since, for the most part, they are repetitions of earlier arguments. We first note that Lemma 3.1 is valid for f of the type (3.36) and that Lemma 3 .3 is also valid. Lemma 3.2, as it stands, is not true, but can easily be modified to hold. Now if $\{T_n\}$ is a sequence of designs containing $a_1, \cdots, a_L$ for all $n \geqq L$ and if we put $f(t) = \int_0^1 R(s, t)\, dF(s)$, $P_{T_n}f(t) = \int_0^1 R(s, t)\, dF_n$ (the definition of F is obvious from (3.55) and F_n is the appropriate discrete signed measure whose support is T_n) then

$$\|f - P_{T_n}f\|^2 = \int_0^1 (f - P_{T_n}f)\, d(F - F_n) = \int_0^1 (f - P_{T_n}f)\, dF = \int_0^1 (f - P_{T_n}f)\varphi\, dt$$

where the last two equalities result from $f - P_{T_n}f$ being 0 on T_n. Proceeding as we did in the proof leading to (3.48) except that we use right derivatives at t_i we obtain (3.39) for sequences $\{T_n\}$ of the kind considered here. For other sequences $\{T_n\}$ we adjoin $\{a_1, \cdots, a_L\}$ and argue as we did following (3.39). (3.51) can be established without difficulty once we have adjoined $a_1, \cdots, a_L$ to the designs and thereby removed the effect of having $\sum \mu_k R(a_k, t)$, and having done so we are finished with the proof of the extension.

REMARK 3.4. If we imposed further restrictions on φ and α we could obtain a better estimate of the second order term in the approximation $\inf_T \|f - P_T f\|^2 = \gamma^3 n^{-2}/12 + o(n^{-2})$. In fact, if φ and α are continuously differentiable then the

$o(n^{-2})$ term can be replaced by $O(n^{-3})$ as can be seen via (3.48). (3.48) could also be used to obtain upper bounds on $\|f - P_T f\|^2$ which would improve upon those found in Lemma 3.1. However, the bound in Lemma 3.1 is valid even when Assumption C does not hold.

REMARK 3.5. It is easy to verify that we need not restrict ourselves to the interval [0, 1] and that Theorem 3.1 holds without change provided we alter (3.38), when the interval is $[a, b]$, by substituting a for 0 and b for 1.

4. Miscellaneous remarks. This section is devoted to miscellaneous remarks and examples which will elaborate on points raised in the previous discussion and, at the same time, exhibit concrete cases.

The major difficulties encountered in applying the results of Section 3 are in obtaining explicit information about $\mathfrak{F}$ and the norm on $\mathfrak{F}$. The reproducing kernel spaces which, at present, are the most easily handled from a computational point of view, are those which correspond to covariance kernels having the form $R(s, t) = u(s)v(t)$, $s < t$. Here the norms $\| \quad \|_T$ and $\| \quad \|$ are easily found. Let us therefore take $R(s, t) = u(s)v(t)$, $s < t$, where u and v are twice continuously differentiable functions on [0, 1] which satisfy

$$u(t)v(s) - u(s)v(t) > 0, \qquad s < t, \tag{4.1}$$

$$u'(t)v(t) - u(t)v'(t) > 0, \qquad t \in [0, 1]. \tag{4.2}$$

The matrices R_T, $T \in D_n$, are shown to be positive definite by virtue of the positivity of u and v together with (4.1).

For f a function on [0, 1] and $T = \{t_1, \cdots, t_n\} \in D_n$ we have

$$\|f\|_T^2 = \sum_{k=1}^{n-1} [(f(t_k)v(t_{k+1}) - f(t_{k+1})v(t_k))^2 / v(t_k)v(t_{k+1})(u(t_{k+1})v(t_k) - u(t_k)v(t_{k+1}))] + f^2(t_1)/u(t_1)v(t_1) \tag{4.3}$$

and if f/v is absolutely continuous we can write

$$\|f\|_T^2 = \sum_{k=1}^{n-1} [(\int_{t_k}^{t_{k+1}} (f/v)')^2 / \int_{t_k}^{t_{k+1}} (u/v)'] + f^2(t_1)/u(t_1)v(t_1).$$

That $\|f\|_T^2$ is actually $f_T' R_T^{-1} f_T$ may be verified by checking that $(f, R(\cdot, t))_T = f(t)$ for $t \in T$. In a similar fashion, one shows that

$$\|f\|^2 = \int_0^1 [(f/v)'^2/(u/v)'] + f^2(0)/u(0)v(0). \tag{4.4}$$

$\mathfrak{F}$, therefore, is the class of functions f for which the integral in (4.4) is finite and f/v is absolutely continuous.

The simplicity of the norm structure for these examples allows one to verify directly the simple present condition of Section 2. Let $\theta = (\epsilon_1, \cdots, \epsilon_n, \delta_1, \cdots, \delta_n)$ where ϵ_i, δ_i are all positive and let $I_\theta(S) = V[R(\cdot, t), t \in \bigcup_k (s_k - \epsilon_k, s_k + \delta_k)$, $S = \{s_1, \cdots, s_n\} \in D_n]$ where the intervals $(s_k - \epsilon_k, s_k + \delta_k)$ are taken to be disjoint and contained in [0, 1]. It is easy to verify that

$$\|P_{I_\theta(S)} f\|^2 = \sum_{k=1}^{n} \int_{s_k-\epsilon_k}^{s_k+\delta_k} [(f/v)'^2/(u/v)'] + f^2(s_1 - \epsilon_1)/u(s_1 - \epsilon_1)v(s_1 - \epsilon_1) + \sum_{k=1}^{n-1} [(\int_{s_k+\delta_k}^{s_{k+1}-\epsilon_{k+1}} (f/v)')^2 / \int_{s_k+\delta_k}^{s_{k+1}-\epsilon_{k+1}} (u/v)']. \tag{4.5}$$

Now the simple present condition is equivalent to requiring $\|P_{I_\theta(S)} f\|^2 = \|f\|^2$

for all θ only if $f \varepsilon V[R(\cdot, s), s \varepsilon S]$. Looking at the nth term in the first summation on the right side of (4.5) it is clear that if we fix all ϵ_i's and δ_i's except for δ_n which we let vary over $(0, 1 - s_n]$ that we obtain $(f/v)'^2 \equiv 0$ on $(s_n, 1]$ or $f = cv$ on $(s_n, 1]$. Similarly, on $[0, s_1)$

$$f^2(s_1 - \epsilon_1)/u(s_1 - \epsilon_1)v(s_1 - \epsilon_1) - f^2(0)/u(0)v(0) + c^* = \int_0^{s_1-\epsilon_1}[(f/v)'^2/(u/v)']$$

or $f = c'u$ on $[0, s_1)$. Finally, on an interior interval (s_k, s_{k+1}),

$$\left(\int_{s_k}^{s_{k+1}} (f/v)'\right)^2 \Big/ \int_{s_k}^{s_{k+1}} (u/v)' = \int_{s_k}^{s_{k+1}} [(f/v)'^2/(u/v)'].$$

This implies by the Cauchy-Schwartz inequality that $(f/v)' = c''(u/v)'$ and therefore $f = c_k u + c_k' v$ on (s_k, s_{k+1}). On the other hand, it is easy to see that $V[R(\cdot, s), s \varepsilon S]$ contains all continuous functions of this form.

Turning to asymptotically optimal designs, we see first that the integral equation (3.2) may be solved for these covariance kernels. Indeed, let f satisfy, for φ continuous,

$$(4.6) \quad f(t) = \int_0^1 R(s, t)\varphi(s)\, ds = v(t) \int_0^t u(s)\varphi(s)\, ds + u(t) \int_t^1 v(s)\varphi(s)\, ds.$$

Then, f is twice continuously differentiable and

$$(4.7) \quad f'(t) = v'(t) \int_0^t u(s)\varphi(s)\, ds + u'(t) \int_t^1 v(s)\varphi(s)\, ds,$$

$$(4.8) \quad f''(t) = v''(t) \int_0^t u(s)\varphi(s)\, ds + u''(t) \int_t^1 v(s)\varphi(s)\, ds + \varphi(t)(u(t)v'(t) - u'(t)v(t)).$$

Solving (4.6) and (4.7) for $\int_0^t u(s)\varphi(s)\, ds$ and $\int_t^1 v(s)\varphi(s)\, ds$ and inserting these in (4.8) gives

$$(4.9) \quad \varphi = -\{f'' - v''[(fu' - f'u)/(u'v - uv')] + u''[(fv' - f'v)/(u'v - v'u)]\}/(u'v - uv').$$

For $R(s, t) = \min(s, t)$, (4.9) reduces to $-f''$. If f is twice continuously differentiable then it may be seen that

$$\begin{aligned} f(t) = {} & \int_0^1 R(s, t)\varphi(s)\, ds \\ & + \{[f(0)u'(0) - u(0)f'(0)]/u(0)[u'(0)v(0) - u(0)v'(0)]\}R(0, t) \\ & + \{[f'(1)v(1) - f(1)v'(1)]/v(1)[u'(1)v(1) - u(1)v'(1)]\}R(1, t), \end{aligned}$$

where φ is given by (4.9). The presence of the "end terms" reflects the need for having the extension of Theorem 3.1 described in Remark 3.3.

The explicit computation of asymptotically optimal designs depends, of course, on our ability to integrate the function $(u'(t)v(t) - u(t)v'(t))^{1/3}(\varphi(t))^{2/3}$ in a sufficiently closed form. Even for the Brownian motion kernel, where this function becomes $(f'')^{2/3}$, this computation may be awkward to perform. The simplest functions to treat in the Brownian motion case are the power functions t^γ, with $\gamma > \frac{1}{2}$, where the nth design T_n^* for t^γ is easily calculated to be $t_{jn}^* = (j/n)^{3/(2\gamma-1)}$, $j = 1, 2, \cdots, n$.

For kernels not of the form $R(s, t) = u(s)v(t)$, $s < t$, the amount of known kernel space information is relatively meager. The evaluation of $\| \quad \|_T$, for example, involves the inversion of R_T, and when this is possible it must be viewed as a fortunate accident. The only example of which we are presently aware is (essentially) $R(s, t) = 1 - |s - t|$ on [0, 1]. Here, for

$$T = \{t_1, \cdots, t_n\} \varepsilon D_n,$$

$$\|f\|_T^2 = \sum_{i=1}^{n-1} [(f(t_{i+1}) - f(t_i))^2/(t_{i+1} - t_i)] + (f(t_1) + f(t_n))^2/2(2 - t_n + t_1),$$

which indicates a close parallel with the Brownian motion case. The evaluation of $\| \quad \|$ is more feasible (cf. [8]). Of prime importance for obtaining the designs of Section 3, however, is the solution of the integral equation (3.2); the norm of f can be determined from this solution.

A class of kernels for which (3.2) has tractable solutions for important special cases is given by $R_n(s, t) = (1 - |s - t|)^n$, s, t in [0, 1], for $n = 1, 2, \cdots$. R_n is completely monotone of order $n + 1$ and in fact, generates all such stationary covariance kernels in a suitable sense [16]. These kernels satisfy Assumptions A, B, and C in Section 3 as can be seen by observing that R_n is of the form (3.7) (when $n \geqq 2$, $p(\lambda) = \lambda^{-3}(1 - \lambda^{-1})^{n-2}n(n-1)$ for $\lambda > 1$ and $p(\lambda) = 0$ otherwise) and $p(\lambda)$ satisfies (3.8) and (3.9)).

For $n = 1$ and f twice continuously differentiable, (3.2) has the solution $\varphi = -f''/2$ with end terms involving $R(\cdot, 0)$ and $R(\cdot, 1)$. This can be extended to $R(s, t) = (1 - \lambda|s - t|)^+$, s, t in [0, 1], for $\lambda > 1$. Here φ is a linear combination of the values of f'' at the points $t + r\lambda^{-1}$ with r an integer and the end terms involve $R(\cdot, r\lambda^{-1})$ and $R(\cdot, 1 - r\lambda^{-1})$, r integral.

For $n \geqq 2$ and f sufficiently smooth, (3.2) can be transformed to a linear differential equation with constant coefficients. For example, if n is odd and φ is $(n - 1)$ times differentiable,

$$(4.10) \quad -(1/2n)\left(\int_0^1 R_n(s, t)\varphi(s)\, ds\right)^{n+1}(t) = \sum_{k=0}^{(n-1)/2} [(n-1)!/(2k)!]\varphi^{(2k)}(t).$$

There are $(n - 1)$ linearly independent solutions to (4.10) set equal to zero for each of which $\int_0^1 R_n(s, t)\varphi(s)\, ds$ is a polynomial of degree at most n. These polynomials together with $R_n(t, 0) = (1 - t)^n$ and $R_n(t, 1) = t^n$ are linearly independent so that (3.2) can be solved for all polynomials of degree n. More generally, one must solve the nonhomogeneous equation (4.10) set equal to h, say. For n even the situation is analogous. As a special case, if $n = 2$ and if f has two continuous derivatives, we obtain

$$\begin{aligned} f(t) = {} & \int_0^1 R_2(s, t)[-f''(s)/4 + \tfrac{3}{5}\int_0^1 f + \tfrac{3}{10}(f(0) + f(1))]\, ds \\ & + [\tfrac{3}{10}\int_0^1 f + \tfrac{2}{5}f(0) - \tfrac{1}{10}f(1) - f'(0)/4]R_2(t, 0) \\ & + [\tfrac{3}{10}\int_0^1 f + \tfrac{2}{5}f(1) - \tfrac{1}{10}f(0) + f'(1)/4]R_2(t, 1). \end{aligned}$$

REFERENCES

[1] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

[2] ELFVING, G. (1956). Selection of nonrepeatable observations for estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.*, **1** 69–75. Univ. of California Press.

[3] GRENANDER, ULF (1950). Stochastic processes and statistical inference. *Ark. für Mat.* **1** 195–277.

[4] HÁJEK, J. (1960). On a simple linear model in Gaussian processes. *Trans. Second Prague Conf. Inf. Theor.* 185–197.

[5] HÁJEK, J. (1961). On linear estimation theory for an infinite number of observations. *Teor. Veroyatnost. i Primenen.* **6** 166–177 (English translation).

[6] HÁJEK, J. (1962). On linear statistical problems in stochastic processes. *Czech. Math. J.* **12** 404–444.

[7] HÁJEK, J. (1962). An inequality concerning random linear functionals on a linear space with a random norm and its statistical application. *Czech. Math. J.* **12** 486–491.

[8] HOEL, PAUL G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29** 1134–1145.

[9] HOEL, PAUL G. (1961). Asymptotic efficiency in polynomial estimation. *Ann. Math. Statist.* **32** 1042–1047.

[10] KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319.

[11] KIEFER, J. (1961). Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp. Math. Statist. and Prob.* **1** 381–405.

[12] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[13] KIEFER, J. (1962). Two more criteria equivalent to D-optimality of designs. *Ann. Math. Statist.* **33** 792–796.

[14] KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

[15] KIEFER, J. AND WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366.

[16] LÉVY, PAUL (1962). Extensions d'un théorème de D. Dugué et M. Girault. *Z. Wahrscheinlishkeitstheorie und Verw. Gebiete* **1** 159–173.

[17] PARZEN, EMANUEL (1959). Statistical inference on time series by Hilbert space methods, I. Tech. Report No. 23 (NR-042-993). Appl. Math. and Statist. Lab., Stanford Univ.

[18] PARZEN, EMANUEL (1961). Regression analysis of continuous parameter time series. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 469–489.

[19] PARZEN, EMANUEL (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[20] SÄRNDAL, CARL-ERIK (1962). On a maximizing problem with several applications in statistical theory. *Ark. für Mat.* **4** 385–392.

[21] SÄRNDAL, CARL-ERIK (1962). *Information from censored samples.* Almquist and Wiksell, Stockholm.

[22] TUTUBALIN, V. N. AND FREIDLIN, M. I. (1962). On the structure of the infinitesimal σ-algebra of a Gaussian process. *Teor. Veroyatnost. i Primenen.* **7** 196–199 (English translation).

[23] YLVISAKER, N. DONALD (1964). Lower bounds for minimum covariance matrices in time series regression problems. *Ann. Math. Statist.* **35** 362–368.