The Annals of Applied Probability
1998, Vol. 8, No. 4, 1184–1215

ESTIMATING FUNCTIONS EVALUATED BY SIMULATION: A BAYESIAN/ANALYTIC APPROACH

BY J. R. KOEHLER, A. A. PUHALSKII AND B. SIMON

University of Colorado, University of Colorado and Institute for Problems in Information Transmission, and University of Colorado

Consider a function f: $B \to R$, where B is a compact subset of R^m and consider a "simulation" used to estimate $f(x)$, $x \in B$ with the following properties: the simulation can switch from one $x \in B$ to another in zero time, and a simulation at x lasting t units of time yields a random variable with mean $f(x)$ and variance $v(x)/t$. With such a simulation we can divide T units of time into as many separate simulations as we like. Therefore, in principle we can design an "experiment" that spends $\tau(A)$ units of time simulating points in each $A \in \mathscr{B}$, where $\mathscr{B}$ is the Borel σ-field on B and τ is an arbitrary finite measure on $(B, \mathscr{B})$. We call a design specified by a measure τ a "generalized design." We propose an approximation for f based on the data from a generalized design. When τ is discrete, the approximation, $\hat{f}$, reduces to a "Kriging"-like estimator. We study discrete designs in detail, including asymptotics (as the length of the simulation increases) and a numerical procedure for finding optimal n-point designs based on a Bayesian interpretation of $\hat{f}$. Our main results, however, concern properties of generalized designs. In particular, we give conditions for integrals of $\hat{f}$ to be consistent estimates of the corresponding integrals of f. These conditions are satisfied for a large class of functions, f, even when $v(x)$ is not known exactly. If f is continuous and τ has a density, then consistent estimation of $f(x)$, $x \in B$ is also possible. Finally, we use the Bayesian interpretation of $\hat{f}$ to derive a variational problem satisfied by globally optimal designs. The variational problem always has a solution and we describe a sequence of n-point designs that approach (with respect to weak convergence) the set of globally optimal designs. Optimal designs are calculated for some generic examples. Our numerical studies strongly suggest that optimal designs have smooth densities.

1. Introduction. Let $f(x)$, $x \in B$ be an unknown function defined on a compact region $B \subset R^m$. We can think of f as a function of m parameters associated with some stochastic model. Consider an idealized simulation (a simulation "meta-model") with the following properties:

(P1) A simulation with the parameters set to $x \in B$ that runs for $t > 0$ units of time yields a random variable, $Y_t(x)$, with mean $f(x)$ and variance $v(x)/t$.
(P2) The simulation can switch from one setting of the parameters to another setting in zero time.

Received October 1995; revised December 1997.
AMS 1991 *subject classifications*. 62K05, 60G60, 60F25.
Key words and phrases. Kriging, experimental designs.

Neither (P1) nor (P2) will hold exactly in any real simulation. However, for Monte Carlo and finite horizon simulations they are typically good approximations, and (P1) and (P2) hold asymptotically in many other cases of interest, for example, regenerative simulations. In fact, very often $Y_t(x)$ will be asymptotically unbiased and normally distributed [Iglehart (1978)]. In any case, (P1) and (P2) serve as a useful meta-model when analyzing problems where multiple simulations are necessary.

Our goal is to estimate $f(x)$, $x \in B$ while constrained to T units of (simulation) time. Property (P2) allows us to run as many separate simulations as we like (at various $x \in B$) during our T unit budget. However, (P1) implies that short simulations have large variances and are therefore not worth very much individually. On the other hand, if f is continuous, simulations at points in close proximity will tend to reinforce each other for estimating f in that neighborhood.

We address the problem of designing efficient experiments for estimating f using T units of simulation time. Normally a design is specified by a choice of points, $x_i \in B$, $i = 1, 2, \ldots, n$ to be simulated, with associated times, $t_i > 0$, where $\sum_{i=1}^{n} t_i = T$. We consider a much larger class of designs which we call "generalized designs." Roughly speaking, a generalized design is specified by a finite measure τ on $(B, \mathscr{B})$, where $\tau(A)$ is interpreted as the amount of time invested simulating points in the set $A \in \mathscr{B}$, and $\mathscr{B}$ is the σ-field of Borel subsets of B. Generalized designs were used by Kiefer and Wolfowitz (1959) as a class of designs that one searches through for an optimal design in regression experiments, and the idea has been utilized elsewhere [see, e.g., Pukelsheim, (1993)]. Kiefer and Wolfowitz found that optimal designs were discrete measures for regressions. Our estimate of f is not a regression and numerical experiments (including those summarized in Section 5) strongly suggest that optimal designs in our context are usually measures with densities.

Unless τ is a discrete measure, a generalized design is not "implementable" in the normal sense. However, in principle (P2) allows us to implement (or at least closely approximate) designs corresponding to any finite measure. We will refer to τ as the "simulation time measure," and define a related measure μ by

$$\mu(A) = \int_A v(x)^{-1} \tau(dx), \qquad A \in \mathscr{B}, \tag{1.1}$$

where $v(x) > 0$, $x \in \mathscr{B}$. From (P1) we see that $v(x)$ can be interpreted as the variance of a unit length simulation at x, so $v(x)^{-1}$ is the asymptotic efficiency constant of a simulation at x [Hamersley and Hanscomb (1964)]. We will refer to μ as the "simulation efficiency measure."

Once we have chosen a simulation design, there remains the problem of constructing an approximation for f based on the resulting data. We choose to consider an approximation, $\hat{f}$, that can be interpreted in two ways.

There is a deterministic interpretation where $\hat{f}$ is simply a smoothing of the simulation data. The "smoothing kernel" is uniquely specified by a (given) symmetric positive definite function, $\rho(x, y)$ [via (3.4)]. Analytic properties of $\hat{f}$ are best understood from this perspective.

We can also consider f to be a realization of a random function, Z, with zero mean and covariance function given by $\rho(x, y)$. The best linear unbiased predictor for Z based on the simulation data is often referred to as a "Kriging" estimate [see Ripley (1981)]. The Kriging estimate, $\hat{Z}$, is the same as $\hat{f}$. We can explicitly calculate the expected value of a wide range of (objective) functions involving Z and $\hat{Z}$, and therefore search for a design that optimizes the objective function. If we think of Z as being a "prior" for f, then this approach can be considered Bayesian. Since f is not known in advance, the Kriging (or Bayesian) perspective seems appropriate for formalizing optimal design criteria.

It is very common in the "design and analysis of computer experiments" literature to assume that an unknown function is a sample from a Gaussian (or other) random field [Currin, Mitchell, Morris and Ylvisaker (1991), Koehler and Owen (1996), Sacks, Schiller and Welch (1989) and Sacks, Welch, Mitchell and Wynn (1989)]. A very flexible technique used for estimating unknown functions is "universal Kriging," which allows the random function Z to have a nonzero mean. A linear model for $E(Z)$ is assumed, and the weights associated with the linear model are estimated as part of the process of constructing the universal Kriging estimate. One can also assume a parametric form for the covariance function and use one that matches (e.g., via maximum likelihood) the observed data. The analysis of a universal Kriging-like estimator in our context is very much harder than the present analysis, due primarily to the presence of the unknown regression coefficients. Although certain "halfway measures" are possible, we choose to analyze the case where $E(Z) = 0$ and $\rho(x, y)$ is chosen in advance. We feel justified in this choice for the following reasons. Theorem 3 shows that integrals of $\hat{f}$ are consistent whenever $f(x)$ and $\rho(x, y)$ jointly satisfy a fairly mild condition, (3.10), so no added complexity is necessary to assure consistency. From a Bayesian perspective, one feels that the random function, Z, should be a good "prior" for f. If a zero mean Gaussian process with a predetermined covariance structure does not seem to be a good prior for f, one can run a short "pilot study" (independent of the main simulation study) to construct a rough (Kriging) estimate of f. The pilot study estimate $\hat{f}_0$ is then subtracted from f to obtain a new function $g = f - \hat{f}_0$ that can be modeled as a zero mean Gaussian process, $Z_0 = Z - \hat{Z}_0$, whose covariance structure is easily calculated from Corollary 2.1. Our method can then be applied to g, resulting in an estimate $\hat{g}$ from the main simulation study. Finally, the estimate for f is $\hat{f} = \hat{g} + \hat{f}_0$.

Ritter, Wasilkowski and Wozniakowski (1995) obtain sharp bounds on the minimal L^2 error for linear estimation of integrals and approximations of a multivariate random function based on n observations. They assume noiseless observations and a covariance function that satisfies a certain condition. They do not attempt to find optimal n-point designs. Plaskota (1992) analyzes the case where the (scalar) function to be estimated is an n-times integrated Wiener process. Noisy data (independent and with constant variance) is included in his formulation. He obtains exact asymptotics as the number of observations increase and finds that an evenly spaced grid is "almost" optimal. The validity of the bounds and asymptotics obtained in these papers (and

others that they reference) depends on the unknown function being a sample path from a particular random field. In our context it is usually difficult to justify the assumption that f is a sample of a random function. In fact, there is no real randomness in the observations besides the uncertainty of the simulations. Our results differ from previous work since the asymptotic analysis of our estimator does not rely in any way on a (Bayesian) assumption that the unknown function comes from a particular random field. The Bayesian framework apparently allows for stronger theorems to be proved (e.g., pointwise convergence vs. convergence of integrals). However, we believe our underlying model is less objectionable and our results are strong enough for most applications.

Of course, since we do not know f beforehand it is tempting to treat it as a random function. Furthermore, there are results that seem to justify that approach. Stein (1988) has shown that under reasonable conditions, misspecified covariance functions do not significantly hinder the estimation of a random function, and we show here that integrals of f can be estimated consistently whenever f and ρ jointly satisfy (3.10). Nevertheless we believe our results better demonstrate the applicability of Kriging-like estimators by avoiding the explicit assumption that f is a random function. On the other hand, we are free to take a Bayesian perspective when it is appropriate. We argue that the Bayesian approach is appropriate in the optimal design problem since presumably the same optimal design is used for a large class of functions. The only place we explicitly assume that f is a random function is when we look for optimal designs.

In the next section we construct $\hat{f}$ in the case where τ is a discrete measure, and discuss its relationship with the Kriging estimator, $\hat{Z}$. We discuss the consistency of $\hat{f}$ and show that $\hat{f}$ is consistent at points where τ has atoms. We show how a Bayesian interpretation of $\hat{f}$ leads to a formulation of an optimal design problem.

In Section 3 we define $\hat{f}$ for a generalized simulation design and show that it has an integral representation, analogous to the discrete case. We also show that it reduces to the approximation described in Section 2 when τ is discrete. The construction of $\hat{f}(x)$ involves a function $a(x, y)$, which satisfies a certain integral equation. We show that $\hat{f}(x)$ is well defined for generalized designs by showing that the integral equation has a unique solution. The function $a(x, y)$ is also shown to have a probabilistic interpretation in terms of the Gaussian process Z. The main result in Section 3 shows that if a certain integral equation involving f, (3.10), has a solution, then $\int_B \phi\hat{f}\, d\mu$ is a consistent estimator for $\int_B \phi f\, d\mu$ for any $L_2(\mu)$ function ϕ. In fact the integrals are consistent even if the variance function $v(x)$ is not known exactly. The integral equation condition is satisfied by a large class of functions [dense in $L_2(\mu)$], further motivating $\hat{f}$ as an estimator for functions evaluated by simulation. We are unable to prove pointwise convergence $\hat{f}(x) \to f(x)$ in general. However, in many cases of interest, consistency of integrals allows us to obtain convergent estimates of $f(x)$. (See Remark 3.4, and Example 3 in Section 5.) In Section 3 we also show that if μ_n converges weakly to μ (or if τ_n converges

weakly to τ and v is continuous) then $\int_B \phi \hat{f}_{\mu_n} d\mu_n$ converges in distribution to $\int_B \phi \hat{f}_{\mu} d\mu$.

In Section 4 we discuss globally optimal designs. The optimal design problem is based on the Bayesian interpretation of $\hat{f}$ in analogy with the development in Section 2. Optimal designs can be found by solving a certain variational problem. We show that a solution is guaranteed to exist for the globally optimal design. The optimal n-point designs found in Section 2 are shown to approach (with respect to weak convergence) globally optimal designs, as the atoms become dense in B.

The proofs of our major theorems are in Appendix A. In Appendix B we describe a numerical procedure for finding optimal discrete designs (on a given set of points) based on Newton's method. In Section 5 we use the procedure to approximate the globally optimal designs for some generic examples by optimizing on a fairly dense grid of points. Optimal designs appear to be measures with smooth densities in all our examples, although we have not been able to establish this property formally.

2. Discrete simulation designs. Normally the only "implementable" simulation designs are discrete; that is, τ is a discrete measure specified by

$$\tau(A) = \sum_{i=1}^{n} t_i 1(x_i \in A), \qquad A \in \mathscr{B},$$

where the x_i's are the (distinct) points in B to be simulated and the t_i's are the corresponding simulation times. Let

$$Y = \left(Y_{t_1}(x_1), Y_{t_2}(x_2), \ldots, Y_{t_n}(x_n)\right)', \tag{2.1}$$

be the data from the simulations (where "$'$" denotes transpose). Property (P1) implies that we can write

$$Y_{t_i}(x_i) = f(x_i) + \alpha_i \chi_i,$$

where χ_i, $i = 1, 2, \ldots, n$ are random variables with zero means, unit variances and covariances

$$c_{ij} = E(\chi_i \chi_j)$$

and

$$\alpha_i^2 = \frac{v(x_i)}{t_i}.$$

There are many ways to estimate $f(x)$, $x \in B$ based on Y, the standard, perhaps, being regression. We propose a different kind of approximation. Let $\rho(x, y)$, $x, y \in B$ be a bounded, positive definite, symmetric function. Define

$$\gamma(x) = \left(\rho(x, x_1), \rho(x, x_2), \ldots, \rho(x, x_n)\right)' \tag{2.2}$$

and

$$\Gamma = R + S, \tag{2.3}$$

where

$$R = \begin{bmatrix} \rho(x_1, x_1) & \rho(x_1, x_2) & \cdots & \rho(x_1, x_n) \\ \rho(x_2, x_1) & \rho(x_2, x_2) & \cdots & \rho(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(x_n, x_1) & \rho(x_n, x_2) & \cdots & \rho(x_n, x_n) \end{bmatrix}$$

and

$$S = \begin{bmatrix} \alpha_1^2 & \alpha_1\alpha_2 c_{12} & \cdots & \alpha_1\alpha_n c_{1n} \\ \alpha_2\alpha_1 c_{21} & \alpha_2^2 & \cdots & \alpha_2\alpha_n c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n\alpha_1 c_{n1} & \alpha_n\alpha_2 c_{n2} & \cdots & \alpha_n^2 \end{bmatrix}.$$

Note that $S_{ij} = \mathrm{Cov}(Y_{t_i}(x_i), Y_{t_j}(x_j))$. Very often S will be a diagonal matrix since the individual simulations are independent. However, the simulation designer may choose to use the same stream of random numbers for each simulation, use likelihood ratios to obtain all the estimates from the same simulation, or any of a number of other variance reduction or efficiency enhancing techniques that can cause the simulations to be dependent.

Our approximation for f based on Y is

$$\hat{f}(x) = \gamma(x)'\Gamma^{-1}Y, \qquad x \in B. \tag{2.4}$$

For each $x \in B$, $\hat{f}(x)$ is a linear combination of the elements of Y. The motivation for (2.4) comes from a Bayesian approach to estimating $f(x)$ from Y, which will be described shortly. Most of the important properties of $\hat{f}(x)$ can be derived directly from (2.4) [or (3.3) and (3.6), which generalize (2.4)] without utilizing the Bayesian interpretation. We can write

$$Y = F + \varepsilon \tag{2.5}$$

where

$$F = \big(f(x_1), f(x_2), \ldots, f(x_n)\big)' \tag{2.6}$$

and

$$\varepsilon = (\alpha_1\chi_1, \alpha_2\chi_2, \ldots, \alpha_n\chi_n)'. \tag{2.7}$$

The mean squared error of $\hat{f}(y)$ is therefore

$$E\big[\big(f(y) - \hat{f}(y)\big)^2\big] = \delta^2(y) + \sigma^2(y),$$

where

$$\delta(y) = f(y) - \gamma(y)'\Gamma^{-1}F \tag{2.8}$$

and

$$\sigma^2(y) = \gamma(y)'\Gamma^{-1}S\Gamma^{-1}\gamma(y). \tag{2.9}$$

We will refer to $\delta(y)$ and $\sigma^2(y)$ as the bias and variance of $\hat{f}(y)$, respectively.

Now, consider a class of simulation designs indexed by $T > 0$, specified by simulation time measures $\tau_T = T\tau$ (which are simulations exactly like τ, but allowed to run for T times as long). Let $\hat{f}_T$ be the estimate of f obtained from τ_T. Let $\delta_T(y)$ and $\sigma_T^2(y)$ be the corresponding bias and variance terms. From (2.8) and (2.9) we have

$$\delta_T(y) = f(y) - \gamma(y)'\Gamma_T^{-1}F$$

and

$$\sigma_T^2(y) = \frac{1}{T}\,\gamma(y)'\Gamma_T^{-1}S\Gamma_T^{-1}\gamma(y),$$

where

$$\Gamma_T = R + \frac{1}{T}S. \tag{2.10}$$

Since ρ is positive definite, R is guaranteed to be invertible. Using (2.10) we write

$$\Gamma_T^{-1} = R^{-1} - \frac{1}{T}\,R^{-1}SR^{-1} + \frac{1}{T^2}\,R^{-1}SR^{-1}S\Gamma_T^{-1}. \tag{2.11}$$

It follows that

$$\lim_{T\to\infty} T\delta_T(x_i) = \alpha_i^2(R^{-1}F)_i, \qquad i = 1, 2, \ldots, n, \tag{2.12}$$

and

$$\lim_{T\to\infty} T\sigma_T^2(y) = \gamma(y)'R^{-1}SR^{-1}\gamma(y), \qquad y \in B. \tag{2.13}$$

If $y \in \{x_1, x_2, \ldots, x_n\}$ then (2.13) simplifies to $T\sigma_T^2(x_i) \to \alpha_i^2$. From (2.12) and (2.13) we see that $\hat{f}_T(x_i)$ is a consistent estimator for $f(x_i)$, but from (2.4) and (2.5) it follows that

$$\hat{f}_T(y) \to_p \gamma(y)'R^{-1}F$$

(where $\to_p$ is convergence in probability) which, in general, is not equal to $f(y)$ unless $y \in \{x_1, x_2, \ldots, x_n\}$.

Readers familiar with linear prediction might have noticed that (2.4) looks like a "Kriging" estimate [Ripley (1981)]. Indeed, suppose $Z(x)$, $x \in B$ is a Gaussian process with zero mean and covariance function, ρ. Since f is an unknown function we can adopt a Bayesian perspective and assign Z as a prior for f. We then use the simulation data to construct $\hat{Z}$, the mean of the posterior process, which can be used to estimate Z. The Gaussian process, Z, is a good prior for f if "typical" sample paths of Z look qualitatively like f. See Currin, Mitchell, Morris and Ylvisaker (1991) for a discussion about properties of sample paths of Gaussian processes with various covariance functions.

From the Bayesian perspective the results of the simulation are

$$\tilde{Y} = \zeta + \varepsilon,$$

where

$$\zeta = \big(Z(x_1), Z(x_2), \ldots, Z(x_n)\big)'$$

and ε is given by (2.7). The mean of the posterior process based on $\tilde{Y}$ is

$$\hat{Z}(x) = E(Z(x) \mid \tilde{Y}) = \gamma(x)'\Gamma^{-1}\tilde{Y}. \tag{2.14}$$

If $\tilde{Y}$ is Gaussian, then $\hat{Z}$ is also the best linear unbiased predictor for Z [see Ripley (1981)]. Clearly

$$E(\hat{Z}) = 0, \tag{2.15}$$

and since $\mathrm{Cov}(\tilde{Y}) = \Gamma$, the covariance structure for $\hat{Z}$ is found to be

$$E\big(\hat{Z}(x)\hat{Z}(y)\big) = \gamma(x)'\Gamma^{-1}\gamma(y). \tag{2.16}$$

Using the variance decomposition formula we can write

$$\rho(x, x) \equiv \mathrm{Var}\big(Z(x)\big) = \mathrm{Var}\big(E(Z(x)|\tilde{Y})\big) + E\big(\mathrm{Var}(Z(x)|\tilde{Y})\big). \tag{2.17}$$

It is well known [see Koehler and Owen (1996), Sacks, Schiller and Welch (1989)] that $\mathrm{Var}(Z(x)|\tilde{Y})$ only depends on the design points $\{x_1, x_2, \ldots, x_n\}$ and is therefore a constant with respect to expectation. Considering (2.14) and (2.16), we obtain

$$\mathrm{Var}(Z(x)|\tilde{Y}) = \rho(x, x) - \gamma(x)'\Gamma^{-1}\gamma(x). \tag{2.18}$$

REMARK 2.1. If we run the design τ over and over, constructing $\hat{Z}$ each time, we will see that the empirical distribution of $\hat{Z}$ does not match (2.15) and (2.16). This is because f is not really a random function (i.e., $Z \equiv f$ each time). The simulation "knows" f in the sense that $E(Y) = F$. The simulation yields $\hat{f}$, whose correct bias and variance are given by (2.8) and (2.9).

Suppose we are interested in finding a design that optimizes some measure of the quality of $\hat{f}$. For example, we might want to do the following:

$$\underset{\mathscr{D}_T^n}{\text{minimize}}\ E\left[\left(\int_B \phi(x)\hat{f}(x)\,dx - \int_B \phi(x)f(x)\,dx\right)^2\right], \tag{2.19}$$

where $\phi(x)$, $x \in B$ is some bounded Borel function and $\mathscr{D}_T^n$ is the class of discrete measures τ with support on n points and $\tau(B) = T$. [This optimality criterion is also used by Ritter, Wasilkowski and Wozniakowski (1995).] Clearly, the optimal design depends on f, which is unknown. It is therefore impossible to find an optimal design in advance.

In this context, the Bayesian perspective seems appropriate as long as Z is a reasonable "prior" for f. We substitute the random function Z for f in (2.19), which yields

$$\begin{aligned} &E\left[\left(\int_B \phi(x)\hat{Z}(x)\,dx - \int_B \phi(x)Z(x)\,dx\right)^2\right] \\ &\quad = \int_B\int_B \phi(x)\phi(y)\rho(x, y)\,dy\,dx - \psi'\Gamma^{-1}\psi, \end{aligned}$$

where

$$\psi = \int_B \phi(x)\gamma(x)\,dx.$$

Since ϕ and ρ do not depend on the design, the optimization problem becomes

$$\underset{\mathscr{D}_T^n}{\text{maximize}}\ \psi'\Gamma^{-1}\psi. \tag{2.20}$$

Problem (2.20) will always have a solution since $\mathscr{D}_T^n$ is compact. However, since property (P2) allows us to use designs from $\mathscr{D}_T = \bigcup_{n=1}^{\infty} \mathscr{D}_T^n$ it is natural in our context to

$$\underset{\mathscr{D}_T}{\text{maximize}}\ \psi'\Gamma^{-1}\psi. \tag{2.21}$$

Unfortunately, it is not clear that a solution to (2.21) exists. In fact, our numerical studies (including the examples in Section 5) strongly suggest that optimal designs are not discrete. [Equations (B.10) and (B.11) may also be relevant here.]

REMARK 2.2. In many applications, the purpose of estimating f is to find its roots or optimal points as opposed to a linear functional. Our theorems in Section 3 prove convergence of linear functionals, while pointwise convergence would clearly be preferable for rootfinding and optimization. On the other hand, Kriging estimators can be used to find roots and optimal points in adaptive searches [Simon (1997)].

3. Estimating *f* from a generalized design. This section formalizes the concept of a generalized design and studies some of its properties. As above, B is a compact subset of R^m with Borel σ-field $\mathscr{B}$, τ is a nonnegative finite measure on $(B, \mathscr{B})$ and $\rho(x, y)$, $x, y \in B$ is a bounded, positive definite, symmetric function. The function $v(x)$, $x \in B$ is assumed to be a positive Borel function.

The motivation for introducing generalized designs is the desire to get hold of the situation where the number of simulation points is large. Our numerical studies (e.g., Section 5) imply that good designs must use large numbers of points. It follows from (2.5) that the random part of the data from a discrete simulation design can be specified by the (discrete) stochastic measure

$$\nu(A) = \sum_{i=1}^{n} \alpha_i^{-1}\chi_i 1(x_i \in A), \qquad A \in \mathscr{B}. \tag{3.1}$$

The data from a generalized design is characterized analogously.

We say a simulation has an (orthogonal) generalized design if it yields data whose randomness is characterized by an orthogonal stochastic measure ν [Shiryaev (1984)] where $\nu(A)$ has zero mean and

$$E(\nu(A)^2) = \mu(A), \qquad A \in \mathscr{B}. \tag{3.2}$$

If in addition, $\nu(A)$ is normally distributed for all $A \in \mathscr{B}$ we will call the design "Gaussian." In this section we require that ν be an orthogonal stochastic measure.

REMARK 3.1. In order to strictly generalize the material in the preceding section we would need to consider nonorthogonal stochastic measures as a generalization of (3.1). However, we restrict ourselves in this section to orthogonal generalized designs, and when we refer to a "generalized design" it is understood that ν is an orthogonal stochastic measure.

Let $f(x)$, $x \in B$ be a real valued function. The approximation for f based on the data from a generalized design depends on both the design (specified by ν) and on ρ. Define

$$\hat{f}(x) = \int_B a(y, x) f(y) \mu(dy) + \int_B a(y, x) \nu(dy), \qquad x \in B, \tag{3.3}$$

where $a(y, x)$, $y, x \in B$ is the solution of the equation

$$a(z, x) = \rho(z, x) - \int_B \rho(z, y) a(y, x) \mu(dy), \qquad z, x \in B. \tag{3.4}$$

By (3.2), the measure ν is a.s. absolutely continuous with respect to the measure μ. Let

$$\lambda(y) = \frac{d\nu}{d\mu}(y) \tag{3.5}$$

be the Radon–Nikodym derivative. Then (3.3) takes the form

$$\hat{f}(x) = \int_B a(y, x)(f(y) + \lambda(y)) \mu(dy). \tag{3.6}$$

We call $f(y) + \lambda(y)$ a "generalized observation." We now show that our definition of $\hat{f}$ is consistent with the material in the previous section. Let μ be a discrete measure with atoms at x_i, $i = 1, 2, \ldots, n$. That is,

$$\mu(A) = \sum_{i=1}^{n} \alpha_i^{-2} 1(x_i \in A), \qquad A \in \mathscr{B}.$$

Define

$$\eta_i(x) = \alpha_i^{-2} a(x_i, x), \qquad x \in B, \ i = 1, 2, \ldots, n$$

and

$$\eta(x) = \big(\eta_1(x), \eta_2(x), \ldots, \eta_n(x)\big)'.$$

Letting $z = x_i$, $i = 1, 2, \ldots, n$ in (3.4) yields

$$\eta(x) = \Gamma^{-1} \gamma(x),$$

where $\gamma(x)$ and Γ are given by (2.2) and (2.3). Also, in this case,

$$\nu(A) = \sum_{i=1}^{n} \alpha_i^{-1} \chi_i 1(x_i \in A), \qquad A \in \mathscr{B},$$

where χ_i, $i = 1, 2, \ldots, n$ are independent random variables with zero means and unit variances. Substituting into (3.3) yields (2.4).

The following theorem shows that $\hat{f}$ is well defined, which boils down to proving that (3.4) has a unique solution. We also state some properties of $a(\cdot, \cdot)$. We denote the Hilbert space of square integrable functions on B with respect to μ by $L_2(\mu)$.

THEOREM 1. *For every $x \in B$ there exists a unique $a(\cdot, x) \in L_2(\mu)$ satisfying* (3.4). *The function $a(z, x)$ is, moreover, bounded and symmetric in its two variables. If ρ is continuous in both variables, then $a(z, x)$ is continuous in both variables.*

The proof is given in Appendix A.

The function $a(x, y)$ which satisfies (3.4) has a probabilistic interpretation given by Theorem 2 whose proof is also in Appendix A.

THEOREM 2. *Let $Z(x)$, $x \in B$ be a Gaussian process, independent of ν, with $E(Z) = 0$ and $E(Z(x)Z(y)) = \rho(x, y)$. Define*

$$\hat{Z}(x) = \int_B a(x, y)Z(y)\mu(dy) + \int_B a(x, y)\nu(dy).$$

Then

$$a(x, y) = E\Big[\big(Z(x) - \hat{Z}(x)\big)\big(Z(y) - \hat{Z}(y)\big)\Big].$$

Using (2.16) and (A.9) from the proof of Theorem 2, we have the following useful result.

COROLLARY 3.1. *When τ is discrete and S is diagonal* (*independent simulations*),

$$a(x, y) = \rho(x, y) - \gamma(x)'\Gamma^{-1}\gamma(y). \tag{3.7}$$

In Section 2 we showed that $\hat{f}$ consistently estimates f at the atoms of τ as the length of the simulation increases. We now establish similar properties of generalized designs. As in Section 2 we will be assuming that the simulation is governed by a parameter, T, so that the simulation time measure corresponding to T is $T\tau$, where τ is a fixed measure. Likewise, the simulation efficiency measure corresponding to T is $T\mu$. Functions corresponding to simulations of length T will be subscripted with a T. According to (3.3) and (3.4), the estimate of f based on a simulation of T units is

$$\hat{f}_T(x) = \int_B Ta_T(y, x)f(y)\mu(dy) + \int_B \sqrt{T}a_T(y, x)\nu(dy), \qquad x \in B, \tag{3.8}$$

where $a_T(x, y)$ satisfies

$$a_T(z, x) = \rho(z, x) - T\int_B \rho(z, y)a_T(y, x)\mu(dy), \qquad z, x \in B. \tag{3.9}$$

Our main convergence result is the following theorem.

THEOREM 3. *Assume that $f \in L_2(\mu)$ and is representable in the form*

$$f(x) = \int_B \rho(z, x)k(z)\mu(dz) \tag{3.10}$$

for some $k \in L_2(\mu)$. Then for every function $\phi \in L_2(\mu)$,

$$E\left[\left(\int_B \hat{f}_T(x)\phi(x)\mu(dx) - \int_B f(x)\phi(x)\mu(dx)\right)^2\right] \to 0,$$

as $T \to \infty$.

We prove Theorem 3 in Appendix A.

REMARK 3.2. Equation (3.10) is a Fredholm equation of the first kind, which does not have a solution $k \in L_2(\mu)$ in general. However, the set of functions, f, for which a solution exists is dense in $L_2(\mu)$ [see Tricomi (1985)]. In some cases it is possible to guarantee a solution if f is smooth enough. For example, if $B = [0, 1]$ and if f is twice differentiable with $f(0) = f'(0) = f(1) = f'(1) = 0$, then

$$\int_0^1 e^{-|x-y|}k(y)\,dy = f(x)$$

has a solution,

$$k(y) = \frac{f(y) - f''(y)}{2}.$$

Covariance functions of the form $\rho(x, y) = e^{c|x-y|}$ are common in the Kriging literature [Ripley (1981)].

REMARK 3.3. If τ (and therefore μ) is discrete, then (3.10) has a solution and Theorem 3, in accordance with the results of Section 2, shows that the values of $\hat{f}_T$ at the atoms are consistent estimates of f at those points. On the other hand, Theorem 3 is not strong enough to obtain the rates of convergence given by (2.12) and (2.13).

REMARK 3.4. We are unable to prove convergence $\hat{f}(x) \to f(x)$ if x is not an atom of τ. However, we can obtain convergent estimates in certain cases. Fix $x \in B$ and let $\Delta_n \subset B$ be a sequence of neighborhoods of x, with diameters decreasing to 0, and let m_n be the Lebesgue measure of Δ_n. Define

$$I_T^n = m_n^{-1} \int_{\Delta_n} \hat{f}_T(y)\mu(dy).$$

If f and v are continuous at x and τ has a density $t(\cdot)$ which is positive and continuous in a neighborhood of x, then there is a sequence $T_n \to \infty$ satisfying

$$E\left[\left(\frac{v(x)}{t(x)} I_{T_n}^n - f(x)\right)^2\right] \to 0.$$

Suppose (as would be typical in practice) we do not know the variance of a unit length simulation exactly. This misspecified variance is similar in spirit to the problem considered in Stein (1988). Let $\tilde{v}(x)$ be the "true" variance of a unit length simulation at x, which we assume is positive and which we estimate by $v(x)$. We define $\tilde{\mu}$ via (1.1) with $\tilde{v}(x)$ replacing $v(x)$. The measures μ and $\tilde{\mu}$ are mutually absolutely continuous with density

$$\frac{d\tilde{\mu}}{d\mu}(y) = \frac{v(y)}{\tilde{v}(y)}. \tag{3.11}$$

Since we do not know $\tilde{v}(x)$, we must use $\mu(\cdot)$ and $a(\cdot,\cdot)$ based on $v(x)$ when we estimate f. Using (1.1), (3.5) and (3.11), the estimate we obtain from τ is

$$\tilde{f}(x) = \int_B a(y,x)\big(f(y) + \tilde{\lambda}(y)\big)\mu(dy),$$

where

$$\tilde{\lambda}(y) = \frac{d\tilde{\nu}}{d\mu} = \sqrt{\frac{v(y)}{\tilde{v}(y)}}\lambda(y). \tag{3.12}$$

Likewise, from τ_T we obtain

$$\tilde{f}_T(x) = \int_B Ta_T(y,x)\big(f(y) + \tilde{\lambda}(y)\big)\mu(dy), \tag{3.13}$$

where

$$\tilde{\lambda}_T(y) = T^{-1/2}\tilde{\lambda}(y). \tag{3.14}$$

Clearly, $\tilde{f}_T(x)$ is not the same as $\hat{f}_T(x)$. However, the following result (proved in Appendix A) shows that $\tilde{f}_T(x)$ is consistent in the same sense that $\hat{f}_T(x)$ is.

COROLLARY 3.2. *Assume that $v(y)/\tilde{v}(y)$ is bounded. Then under the assumptions of Theorem* 3,

$$E\left[\left(\int_B \tilde{f}_T(x)\phi(x)\mu(dx) - \int_B f(x)\phi(x)\mu(dx)\right)^2\right] \to 0,$$

as $T \to \infty$.

We now address the issue of convergence of estimates $\hat{f}$ if the corresponding measures μ converge weakly. In particular, we show that estimates from generalized designs are limits of the estimates obtained from discrete designs. Consider a sequence of generalized simulation designs on B with covariance functions ρ_n and efficiency measures μ_n. We denote all characteristics of the nth design with subscript n. We assume, in addition to the above assumptions, that the ρ_n's are continuous and the $\nu_n(A)$, $A \in \mathscr{B}$, are normally distributed (i.e., we require Gaussian generalized designs here).

Along with the sequence of designs we consider another design with covariance function ρ and efficiency measure μ which will be the limiting one in

the following Theorem 4. As with the designs in the sequence, the orthogonal measure ν which corresponds to μ is Gaussian.

In all the designs the same function, $f(x),\ x \in B$, is estimated. As above, we are fixing some Borel function ϕ which is assumed to be bounded. Let

$$\langle \hat{f}_n, \phi \rangle_n = \int_B \hat{f}_n(x)\phi(x)\mu_n(dx)$$

and

$$\langle \hat{f}, \phi \rangle = \int_B \hat{f}(x)\phi(x)\mu(dx).$$

Denote by $\to_w$ weak convergence of measures defined on $(B, \mathscr{B})$, and by $\to_d$, convergence in distribution on the real line [Billingsley (1968)].

THEOREM 4. *Assume that as* $n \to \infty$, $\sup_{x,\, y \in B} |\rho_n(x, y) - \rho(x, y)| \to 0$ *and* $\mu_n \to_w \mu$. *Then*

$$\langle \hat{f}_n, \phi \rangle_n \to_d \langle \hat{f}, \phi \rangle.$$

The proof is in Appendix A.

4. Globally optimal designs. Let $\mathscr{M}_T$ be the set of measures on $(B, \mathscr{B})$ satisfying $\tau(B) = T$ and let $\mathscr{M} = \bigcup_{T>0} \mathscr{M}_T$. Let $\mathscr{C}$ be the class of functions $g: B \to R$ satisfying $|g(x)| \le 1,\ x \in B$, and with Lipschitz constant no larger than one. We make $\mathscr{M}$ a metric space by using

$$d(\tau_1, \tau_2) = \sup_{g \in \mathscr{C}} \left| \int_B g(x)\tau_1(dx) - \int_B g(x)\tau_2(dx) \right|$$

as a distance function. In the topology on $\mathscr{M}$ induced by the metric d (the "weak topology"), the mode of convergence is weak convergence, that is, $d(\tau_n, \tau) \to 0$ if and only if $\tau_n \to_w \tau$ [see, e.g., Dudley (1989)]. In the weak topology, $\mathscr{M}_T$ is a compact subset of $\mathscr{M}$, and the set of discrete measures $\mathscr{D}_T \subset \mathscr{M}_T$ is dense in $\mathscr{M}_T$, but not closed.

Let ϕ: $B \to R$ be a (fixed) bounded Borel function. Ideally, we would like to find $\tau \in \mathscr{M}_T$ that minimizes the expected mean squared error of $\int_B \phi(x)\hat{f}(x)\,dx$. However, the optimal measure depends on f (which is unknown) and therefore cannot be determined in advance. In Section 2 we formulated the optimal design problem for $\mathscr{D}_T$ by using the Bayesian interpretation of $\hat{f}$. The same reasoning applies in the general case. For $\tau \in \mathscr{M}$, let

$$Q(\tau) = E\left[\left(\int_B \phi(x)\hat{Z}(x)\,dx - \int_B \phi(x)Z(x)\,dx \right)^2 \right], \tag{4.1}$$

where Z is a Gaussian process with $E(Z(x)) = 0$, $E(Z(x)Z(y)) = \rho(x, y)$ and let

$$\hat{Z}(x) = \int_B a(u, x)Z(u)\mu(du) + \int_B a(u, x)\nu(du), \tag{4.2}$$

where ν is an orthogonal stochastic measure satisfying $E(\nu(A)) = 0$, $E(\nu(A)^2) = \mu(A)$ and ν is independent of Z. If we believe that Z is a reasonable prior for f, then designs that minimize $Q(\tau)$ can be considered to be optimal.

By Theorem 2, (4.1) reduces to

$$Q(\tau) = \int_B \int_B \phi(x)\phi(y) E\left(\left(\hat{Z}(x) - Z(x)\right)\left(\hat{Z}(y) - Z(y)\right)\right) dy\, dx$$
$$= \int_B \int_B \phi(x)\phi(y) a(x, y)\, dy\, dx.$$

Optimal designs must therefore satisfy the variational problem,

$$\text{minimize}_{\tau \in \mathscr{M}_T} Q(\tau). \tag{4.3}$$

We have not found a nontrivial example where (4.3) can be solved exactly, so numerical approaches are necessary. We now justify an approach to approximating optimal designs via discrete designs. For the remainder of this section we assume that $v(x)$ is continuous.

LEMMA 5.1. *If $v(x)$ is continuous and ρ is continuous, then the map Q: $\mathscr{M} \to \mathscr{R}$ is continuous in the weak topology.*

PROOF. The metric was chosen because $d(\tau_n, \tau) \to 0$ if and only if $\tau_n \to_w \tau$. Since $v(x)$ is positive and continuous, it follows from (1.1) that $\tau_n \to_w \tau$ implies $\mu_n \to_w \mu$. It therefore suffices to show that $\mu_n \to_w \mu$ implies $Q(\tau_n) \to Q(\tau)$. From Lemma 4.1 in Appendix A, $a_n(x, y) \to a(x, y)$ uniformly, where $a_n(x, y)$ and $a(x, y)$ satisfy (3.4) for μ_n and μ, respectively. Since ϕ is bounded, the lemma follows. □

Let

$$Q_T^* = \inf_{\tau \in \mathscr{M}_T} Q(\tau),$$

and let $\mathscr{A}_T \subset \mathscr{M}_T$ be the set of optimal designs. That is,

$$\mathscr{A}_T = \left\{\tau \in \mathscr{M}_T \colon Q(\tau) = Q_T^*\right\}.$$

Since $\mathscr{M}_T$ is compact in the weak topology, Lemma 5.1 implies Theorem 5.

THEOREM 5. *The variational problem* (4.3) *has a solution. That is, $\mathscr{A}_T \neq \varnothing$.*

Since $\mathscr{D}_T$ is dense in $\mathscr{M}_T$, for any $\tau^* \in \mathscr{A}_T$ there is a sequence $\tau_n \to_w \tau^*$, where $\tau_n \in \mathscr{D}_T$, and by Lemma 5.1, $Q(\tau_n) \to Q_T^*$. However, there is no guarantee that there exists $\tau_n \in \mathscr{D}_T$ with $Q(\tau_n) = Q_T^*$. The following lemma shows that nearly optimal designs must be close to $\mathscr{A}_T$. Let $d(\tau_n, \mathscr{A}_T) \equiv \inf_{\tau \in \mathscr{A}_T} d(\tau_n, \tau)$.

LEMMA 6.1. *If $Q(\tau_n) \to Q_T^*$, then $d(\tau_n, \mathscr{A}_T) \to 0$.*

PROOF. From Lemma 5.1, Q is continuous, so $\mathscr{A}_T$ is closed. Suppose there is a sequence $\{\tilde{\tau}_n\}$, $n = 1, 2, \ldots$ satisfying $Q(\tilde{\tau}_n) \to Q_T^*$ and $d(\tilde{\tau}_n, \mathscr{A}_T) > \varepsilon$ for all n, where $\varepsilon > 0$. Then there is an open set, $\mathscr{O} \subset \mathscr{M}_T$, with $\mathscr{A}_T \subset \mathscr{O}$, such that $\tilde{\tau}_n \subset \mathscr{M}_T \backslash \mathscr{O}$, $n = 1, 2, \ldots$. But $\mathscr{M}_T \backslash \mathscr{O}$ is compact, so there exists $\tilde{\tau}^* \in \mathscr{M}_T \backslash \mathscr{O}$ satisfying $Q(\tilde{\tau}^*) = \inf_{\tau \in \mathscr{M}_T \backslash \mathscr{O}} Q(\tau) \leq \inf_n Q(\tilde{\tau}_n) = Q_T^*$. But this contradicts the definition of $\mathscr{A}_T$. □

Let $\mathscr{D}_T^n \subset \mathscr{D}_T$ be the set of discrete measures with n (or fewer) atoms, and let $\tau^* \in \mathscr{A}_T$. There is a sequence $\tau_n \in \mathscr{D}_T^n$, $n = 1, 2, \ldots$, with $\tau_n \to_w \tau^*$, so by Lemma 6.1, $Q(\tau_n) \to Q_T^*$. Let τ_n^* be optimal in $\mathscr{D}_T^n$. Then $Q(\tau_n^*) \leq Q(\tau_n)$, so $Q(\tau_n^*) \to Q_T^*$. By Lemma 6.1, for large enough n, τ_n^* is very close (with respect to d) to $\mathscr{A}_T$. In order to find, τ_n^*, one must optimize over all possible $\{x_1, x_2, \ldots, x_n\} \subset B$ and $t_i \geq 0$, $i = 1, 2, \ldots, n$. Theorem 6 shows that one can find nearly optimal discrete designs by fixing $\{x_1, x_2, \ldots, x_n\} \subset B$ (the "denser" the better) and optimizing over $\{t_1, t_2, \ldots, t_n\}$ only.

THEOREM 6. *Let $S_n = \{x_{n1}, x_{n2}, \ldots, x_{nn}\} \subset B$ be a sequence of finite subsets of B satisfying $\sup_{x \in B} \min_{i \leq n} |x - x_{ni}| \to 0$ as $n \to \infty$ (i.e., S_n becomes dense in B), and let $\mathscr{D}_T(S_n) \subset \mathscr{D}_T^n$ be measures with support on S_n. Let $\tilde{\tau}_n^*$ be optimal in $\mathscr{D}_T(S_n)$. Then $d(\tilde{\tau}_n^*, \mathscr{A}_T) \to 0$ and $Q(\tilde{\tau}_n^*) \to Q_T^*$.*

PROOF. Let $\tau^* \in \mathscr{A}_T$. We first show that there exists a sequence $\tau_n \in \mathscr{D}_T(S_n)$, $n = 1, 2, \ldots$ with $\tau_n \to_w \tau^*$. Define

$$U_{ni} = \left\{ y \in B \colon |y - x_{ni}| \leq \min_{j \neq i} |y - x_{nj}| \right\} \Big\backslash \bigcup_{j<i} U_{nj}$$

to be the points in B closest to x_{ni}. The sets U_{ni}, $i = 1, 2, \ldots, n$ form a disjoint partition of B, so we can define $\tau_n \in \mathscr{D}_T(S_n)$ by

$$\tau_n\big(\{x_{ni}\}\big) = \tau^*(U_{ni}).$$

Fix $\varepsilon > 0$ and let $g \colon B \to R$ be continuous. Since B is compact, g is uniformly continuous, and since by hypothesis $\max_i \operatorname{diam}(U_{ni}) \to 0$, we can find n large enough so that g varies by no more than ε/T on any of the U_{ni}'s. Thus,

$$\left| \int_B g(x) \tau_n(dx) - \int_B g(x) \tau^*(dx) \right| < \varepsilon.$$

Since g and ε are arbitrary, $\tau_n \to_w \tau^*$. Since Q is continuous, $Q(\tau_n) \to Q(\tau^*) = Q_T^*$, and since $Q_T^* \leq Q(\tau_n^*) \leq Q(\tau_n)$ we have $Q(\tau_n^*) \to Q_T^*$. Finally, $d(\tau_n^*, \mathscr{A}_T) \to 0$ follows from Lemma 6.1. □

REMARK 4.1. Theorem 6 implies that if we select n points in B evenly spaced or chosen from some "low dispersion" sequence, then the optimal design with support on that set will closely resemble an optimal design if n is large enough. If the optimal design, τ^*, is unique then $\tau_n^* \to_w \tau^*$.

5. Numerical examples. In this section we illustrate our method by calculating optimal designs and related quantities for some "generic" examples. We know of no (nondegenerate) examples where the variational problem (4.3) can be solved analytically. Fortunately, it is possible to calculate the optimal discrete design on a given finite set of points numerically. Theorem 6 implies that if the points are dense enough, the resulting design is close (with respect to weak convergence) to a globally optimal design. In Appendix B we provide the details for a numerical solution of the optimization problem (2.20) using Newton's method.

EXAMPLE 1. Perhaps the simplest example is estimating functions on the unit interval $B = [0, 1]$ with constant simulation variance $v(x) = 1$. In this case the optimal design, τ^*, minimizes

$$Q(\tau) = \int_0^1 \int_0^1 \phi(x)\phi(y)a(x, y)\,dy\,dx,$$

where $a(x, y)$ satisfies

$$a(x, y) = \rho(x, y) - \int_0^1 \rho(x, z)a(z, y)\tau(dz).$$

Even in this simple case we are unable to find $a(x, y)$ (let alone τ^*) explicitly for any positive definite covariance function, $\rho(x, y)$. Figures 1a and 1b show $a(x, y)$ [determined numerically via (3.7)] when τ is Lebesgue measure and 100 times Lebesgue measure, respectively, and $\rho(x, y) = \exp(-(x - y)^2)$.

We now find τ^* numerically when $\phi(x) = 1$ for two different covariance functions. Let

$$\{x_1, x_2, \ldots, x_{n+1}\} = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}.$$

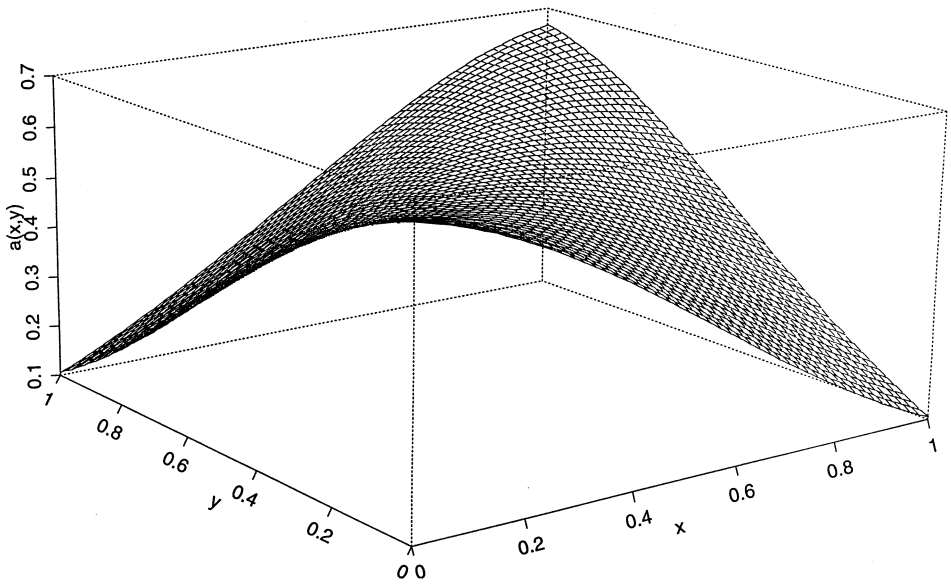
We then search for $\{t_1, t_2, \ldots, t_{n+1}\}$ that optimizes (B.1), (B.2) and (B.3).

In Figures 2a and 2b we plot the "density" of τ^* based on $n + 1 = 51$ points for various values of T when $\rho(x, y) = \exp(-10|x - y|)$ and $\rho(x, y) = \exp(-10(x - y)^2)$, respectively. We cannot be certain, based on the numerical calculation, that τ^* has a density, although the results seem to suggest that it does.

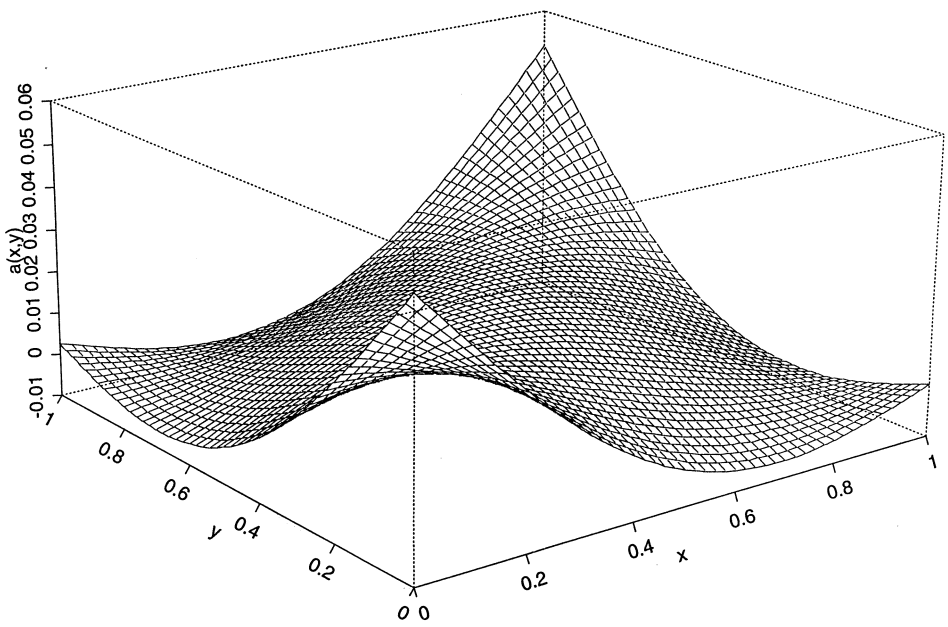
EXAMPLE 2. We next consider the simplest two-dimensional case. Let $B = [0, 1]^2$, $v(x) = 1$ and $\phi(x) = 1$. Figure 3 shows the "density" of τ^* based on $n = 961$ points (31×31 grid) when $T = 100$ and $\rho(x, y) = \exp(-\|x - y\|^2)$, where $\|\cdot\|$ is Euclidian distance.

EXAMPLE 3. Returning to $B = [0, 1]$, we wish to estimate $f(0.5)$ when $v(x) = 1 + x$. Since only integrals of $\hat{f}$ are provably consistent, we choose

$$\phi(x) = \begin{cases} 50, & \text{if } 0.49 < x < 0.51, \\ 0, & \text{otherwise.} \end{cases}$$

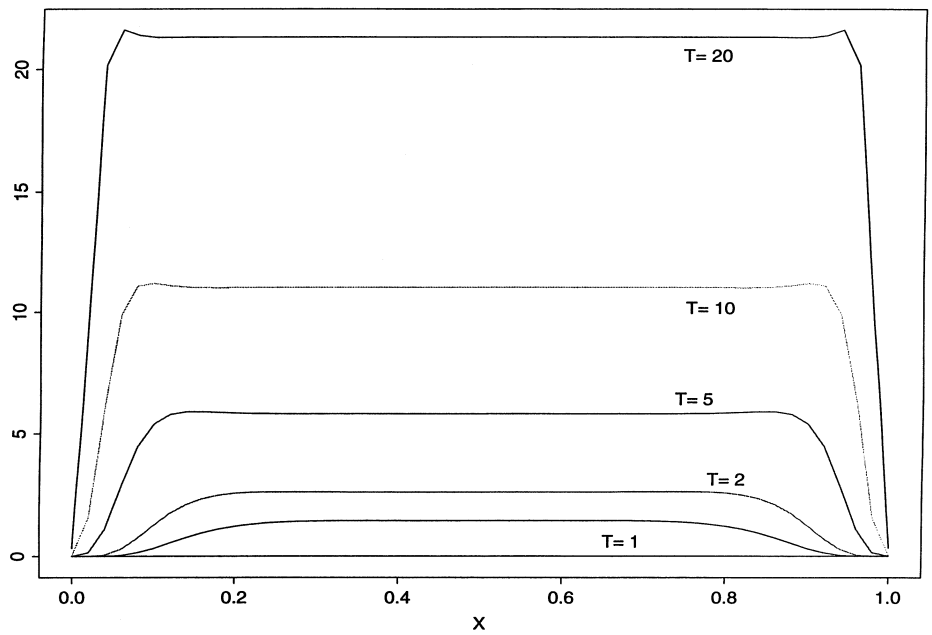

a(x,y)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
0.8
0.6
y
0.4
0.2
0 0
0.2
0.4
x
0.6
0.8
1

(a)

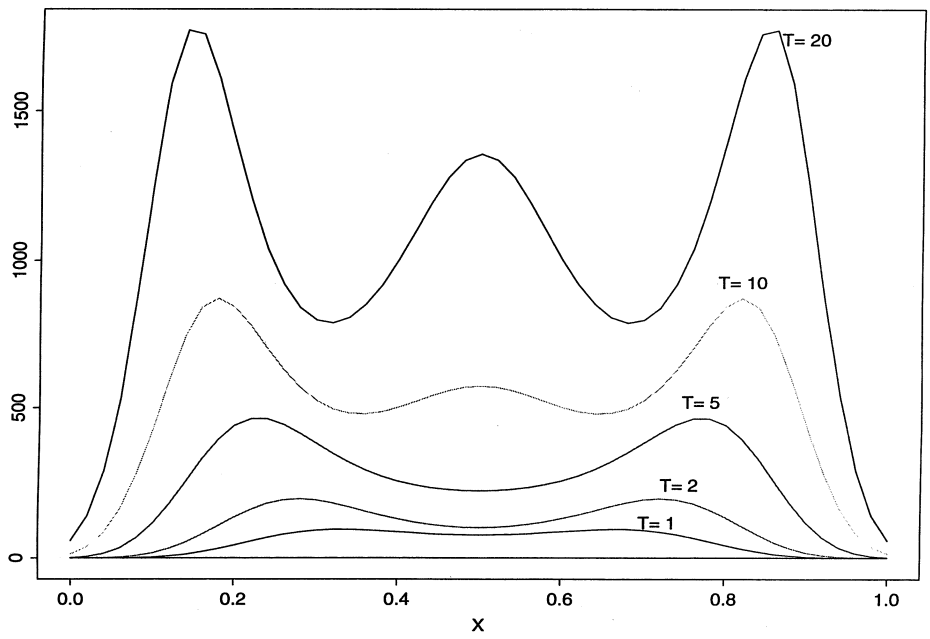

a(x,y)
-0.01
0
0.01
0.02
0.03
0.04
0.05
0.06
-1
0.8
0.6
y
0.4
0.2
0 0
0.2
0.4
x
0.6
0.8
1

(b)

FIG. 1.


T= 20
T= 10
T= 5
T= 2
T= 1
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
X

(a)


T= 20
T= 10
T= 5
T= 2
T= 1
0
500
1000
1500
0.0
0.2
0.4
0.6
0.8
1.0
X

(b)

FIG. 2.

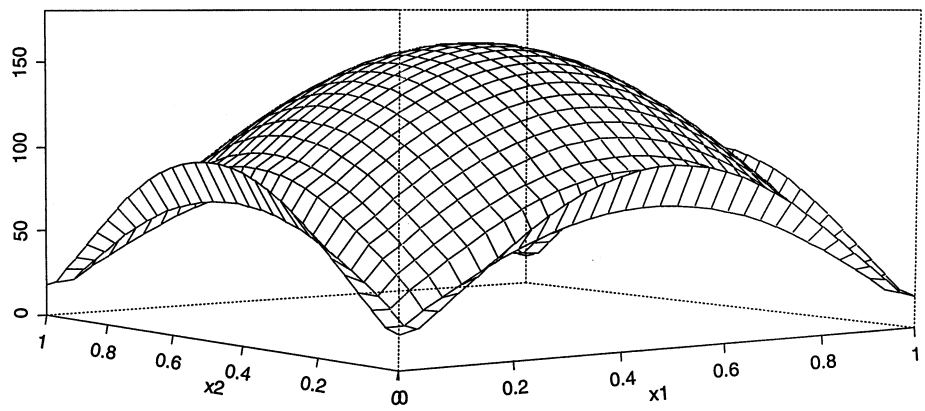


FIG. 3.

We are therefore searching for a design that minimizes

$$E\left[\left(50\int_{0.49}^{0.51}\hat{f}(x)\,dx - 50\int_{0.49}^{0.51}f(x)\,dx\right)^2\right] \approx E\left[\left(\hat{f}(0.5) - f(0.5)\right)^2\right].$$

Figure 4 shows the "density" of τ^* for various values of K when $T = 1$ and $\rho(x, y) = \exp(-K|x - y|)$.

The shapes of the optimal densities are interesting. When K is very small, the Gaussian process $Z(x)$, $0 \le x \le 1$ is essentially a constant [see Currin, Mitchell, Morris and Ylvisaker (1991)]. In that case there is no need to simulate at $x = 0.5$, so the optimal design spends most of its time at $x \approx 0$ where $v(x)$ is smallest. When $K = 0.1$, the optimal design spends virtually no time near $x = 0.5$. As K increases, the value of $Z(x)$, $x \ne 0.5$ becomes less valuable for predicting $Z(0.5)$. By the time $K = 1$, the optimal design is spending the majority of time near $x = 0.5$, although the density is skewed left due to the asymmetry in $v(x)$.

APPENDIX A

PROOF OF THEOREM 1. Consider the equation

$$\int_B \rho(z, y)g(y)\mu(dy) + g(z) = 0, \qquad z \in B. \tag{A.1}$$

We prove that (A.1) has only the trivial solution for $g(\cdot) \in L_2(\mu)$. Multiplying (A.1) by $g(z)$ and integrating with respect to $\mu(dz)$ we get

$$\int_B\int_B \rho(z, y)g(y)g(z)\mu(dy)\mu(dz) + \int_B g^2(z)\mu(dz) = 0.$$

Since ρ is positive definite, the first integral is nonnegative, and hence $g(z) = 0$ μ-a.e.

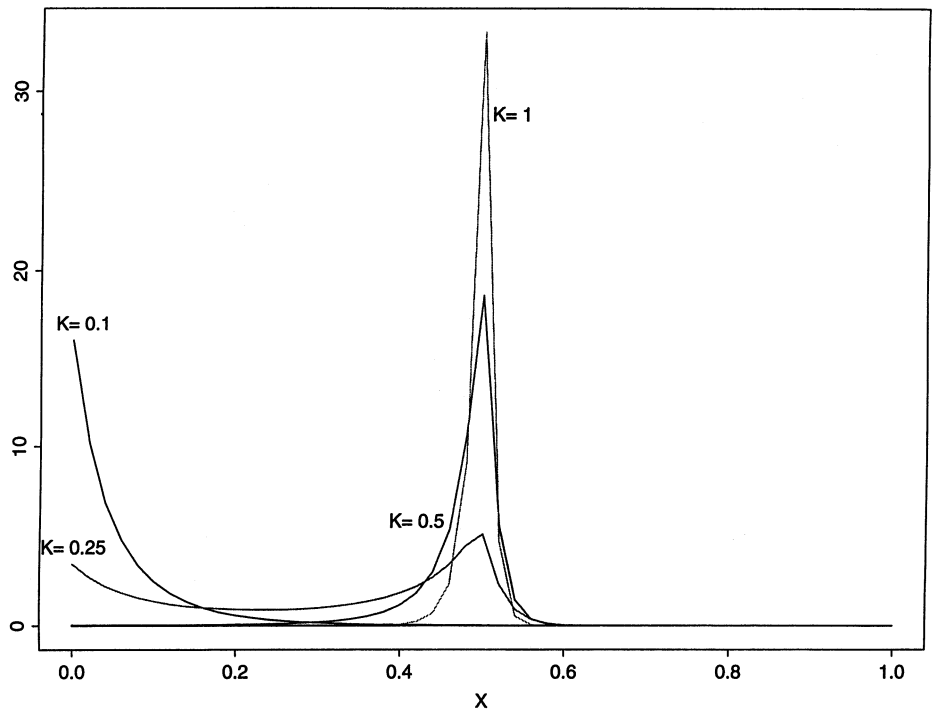


FIG. 4.

Since (A.1) has only the trivial solution, by Fredholm's alternative [Hutson and Pym (1980)], Theorem 7.3.7], (3.4) has a unique solution in $L_2(\mu)$ for every $x \in B$.

Next, by (3.4),

$$\int_B \int_B \rho(z, y)a(y, x)a(z, x)\mu(dy)\mu(dz) + \int_B a^2(z, x)\mu(dz) = \int_B a(z, x)\rho(z, x)\mu(dz),$$

so by the positive definiteness of ρ,

$$\int_B a^2(z, x)\mu(dz) \leq \int_B a(z, x)\rho(z, x)\mu(dz).$$

Applying the Cauchy–Schwarz inequality yields

$$\int_B a^2(z, x)\mu(dz) \leq \int_B \rho^2(z, x)\mu(dz). \tag{A.2}$$

Since ρ is bounded, the boundedness of $a(\cdot, \cdot)$ follows by (A.2) and (3.4).

To show that $a(\cdot,\cdot)$ is symmetric, multiply each term in (3.4) by $a(z,u)$ and integrate with respect to $\mu(dz)$. This yields

$$\begin{aligned}\int_B\int_B \rho(z,y)a(y,x)a(z,u)\mu(dy)\mu(dz) + \int_B a(z,x)a(z,u)\mu(dz)\\ = \int_B \rho(z,x)a(z,u)\mu(dz).\end{aligned} \tag{A.3}$$

The two terms on the left of (A.3) are symmetric in (x,u) [since $\rho(\cdot,\cdot)$ is symmetric], so the term on the right must be symmetric. We apply (3.4) to the right side of (A.3), obtaining

$$\int_B \rho(z,x)a(z,u)\mu(dz) = \rho(x,u) - a(x,u),$$

implying that $a(\cdot,\cdot)$ is symmetric.

Now, let ρ be continuous. The continuity of a would follow from

$$\lim_{\delta\to 0}\sup_{|z_1-z_2|<\delta}\sup_x |a(z_1,x)-a(z_2,x)| = 0 \tag{A.4}$$

and

$$\lim_{\delta\to 0}\sup_{|x_1-x_2|<\delta}\sup_z |a(z,x_1)-a(z,x_2)| = 0.$$

By symmetry, we need prove only (A.4). By (3.4) and the Cauchy–Schwarz inequality,

$$\begin{aligned}&\sup_x\bigl|a(z_1,x)-a(z_2,x)\bigr|\\ &\quad\le \sup_x |\rho(z_1,x)-\rho(z_2,x)|\\ &\qquad + \sup_y |\rho(z_1,y)-\rho(z_2,y)|\sup_x\biggl(\int_B a^2(y,x)\mu(dy)\biggr)^{1/2}\mu(B)^{1/2}.\end{aligned} \tag{A.5}$$

Since ρ is uniformly continuous on B and (A.2) holds, (A.4) is proved. The theorem is proved. □

PROOF OF THEOREM 2. First we show that the integral

$$\int_B a(x,y)Z(y)\mu(dy)$$

is well defined a.e. Since $E(Z(y)Z(y')) = \rho(y,y')$,

$$E\biggl(\int_B a(x,y)Z(y)\mu(dy)\biggr)^2 = \int_B\int_B a(x,y)a(x,y')\rho(y,y')\mu(dy)\mu(dy'),$$

which is finite by the boundedness of $a(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$. Again, using

$$E(Z(x)Z(y)) = \rho(x,y),$$

along with the fact that ν and Z are independent, yields

$$E\bigl(\hat{Z}(x)Z(y)\bigr) = \int_B a(x,u)\rho(u,y)\mu(du).$$

Applying (3.4) yields

$$E\big(\hat{Z}(x)Z(y)\big) = \rho(x, y) - a(x, y).$$

Similarly, we can write

$$\begin{aligned} E\big(\hat{Z}(x)\hat{Z}(y)\big) &= \int_B \int_B a(x, z)a(y, u)\rho(z, u)\mu(dz)\mu(du) \\ &\quad + E\int_B \int_B a(x, z)a(y, u)\nu(dz)\nu(du). \end{aligned} \tag{A.6}$$

Using (3.4) twice yields

$$\begin{aligned} &\int_B \int_B a(x, z)a(y, u)\rho(z, u)\mu(dz)\mu(du) \\ &\qquad = \rho(x, y) - a(x, y) - \int_B a(x, z)a(y, z)\mu(dz). \end{aligned} \tag{A.7}$$

Using the fact that for $g \in L_2(\mu)$,

$$E\left[\left(\int_B g(z)\nu(dz)\right)^2\right] = \int_B g(z)^2\mu(dz)$$

[Shiryaev (1984)], it follows that

$$E\int_B \int_B a(z, x)a(u, y)\nu(dz)\nu(du) = \int_B a(z, x)a(z, y)\mu(dz). \tag{A.8}$$

Combining (A.6), (A.7), (A.8) and the symmetry of $a(\cdot, \cdot)$ yields

$$E\big(\hat{Z}(x)\hat{Z}(y)\big) = \rho(x, y) - a(x, y). \tag{A.9}$$

Putting all the pieces together proves the theorem. □

PROOF OF THEOREM 3. For the proof of Theorem 3 we need to study properties of $a_T(\cdot, \cdot)$ as $T \to \infty$. For this proof, $\phi \in L_2(\mu)$ is fixed. Introduce

$$\overline{a}_T(z) = \int_B a_T(z, x)\phi(x)\mu(dx). \tag{A.10}$$

Then $\overline{a}_T(\cdot) \in L_2(\mu)$ since $a_T(\cdot, \cdot)$ is bounded by Theorem 1 and $\phi(\cdot) \in L_2(\mu)$.

Let L: $L_2(\mu) \to L_2(\mu)$ be the operator defined by

$$Lh(x) = \int_B \rho(y, x)h(y)\mu(dy), \qquad h \in L_2(\mu).$$

Note that by the assumption on ρ, L is continuous, symmetric and positive definite. Denote also

$$b_T = T\overline{a}_T \in L_2(\mu)$$

and let $\|\cdot\|$ be the norm and $\langle \cdot, \cdot \rangle$ be the inner product on $L_2(\mu)$.

LEMMA 3.1. *As* $T \to \infty$, $\langle L(b_T - \phi), b_T - \phi\rangle \to 0$.

PROOF. We first rewrite (4.2) in operator form. Multiplying through by $\phi(x)$ and integrating with respect to $\mu(dx)$, we get

$$Lb_T + \bar{a}_T = L\phi. \tag{A.11}$$

This yields

$$\langle Lb_T, b_T\rangle + \langle \bar{a}_T, b_T\rangle = \langle L\phi, b_T\rangle. \tag{A.12}$$

Since $\langle \bar{a}_T, b_T\rangle = T\langle \bar{a}_T, \bar{a}_T\rangle \geq 0$ and since L is symmetric, by (4.5) and (4.6),

$$\langle Lb_T, b_T\rangle \leq \langle L\phi, b_T\rangle = \langle Lb_T, \phi\rangle = \langle L\phi, \phi\rangle - \langle \bar{a}_T, \phi\rangle,$$

and an application of the Cauchy–Schwarz inequality yields

$$\langle Lb_T, b_T\rangle \leq \langle L\phi, \phi\rangle + \|\bar{a}_T\|\,\|\phi\|. \tag{A.13}$$

We now prove that

$$\lim_{T\to\infty} \|\bar{a}_T\| = 0. \tag{A.14}$$

By (A.12) and the definition of b_T,

$$T\langle L\bar{a}_T, \bar{a}_T\rangle + \|\bar{a}_T\|^2 = \langle L\phi, \bar{a}_T\rangle, \tag{A.15}$$

which implies, since L is positive definite, that

$$\|\bar{a}_T\|^2 \leq \langle L\phi, \bar{a}_T\rangle \leq \|L\phi\|\,\|\bar{a}_T\|.$$

Hence

$$\|\bar{a}_T\| \leq \|L\phi\|. \tag{A.16}$$

On the other hand, (A.15) obviously gives

$$\langle L\bar{a}_T, \bar{a}_T\rangle \leq \frac{1}{T}\langle L\phi, \bar{a}_T\rangle,$$

which coupled with (A.16) yields

$$\langle L\bar{a}_T, \bar{a}_T\rangle \leq \frac{1}{T}\|L\phi\|^2.$$

Next, since L is positive definite, by another application of the Cauchy–Schwarz inequality,

$$\langle L\phi, \bar{a}_T\rangle \leq \langle L\phi, \phi\rangle^{1/2}\langle L\bar{a}_T, \bar{a}_T\rangle^{1/2} \leq \langle L\phi, \phi\rangle^{1/2}\frac{\|L\phi\|}{\sqrt{T}},$$

and invoking (A.15) once again, we get

$$\|\bar{a}_T\|^2 \leq \langle L\phi, \bar{a}_T\rangle \leq \frac{1}{\sqrt{T}}\langle L\phi, \phi\rangle^{1/2}\|L\phi\|,$$

proving (A.14). In view of (A.13) and (A.14),

$$\limsup_{T\to\infty}\langle Lb_T, b_T\rangle \leq \langle L\phi, \phi\rangle. \tag{A.17}$$

Since L is positive definite there exists a symmetric operator $\hat{S}$ such that $L = \hat{S}^2$ [Hutson and Pym (1980)]. Inequality (A.17) then implies that

$$\sup_T \|\hat{S} b_T\| < \infty,$$

so the set $\{\hat{S} b_T,\ T > 0\}$ is weakly relatively compact [Hutson and Pym (1980), Theorem 6.3.7]. Let c be its weak accumulation point. Since (A.11) is easily seen to imply that, for arbitrary $\psi \in L_2(\mu)$,

$$\langle \hat{S} b_T, \hat{S}\psi \rangle + \langle \overline{a}_T, \psi \rangle = \langle \hat{S}\phi, \hat{S}\psi \rangle,$$

(A.14) necessarily implies

$$\langle c, \hat{S}\psi \rangle = \langle \hat{S}\phi, \hat{S}\psi \rangle,$$

or (use again that $\hat{S}$ is symmetric)

$$\langle \hat{S} c, \psi \rangle = \langle \hat{S}^2 \phi, \psi \rangle,$$

and since ψ is arbitrary,

$$\hat{S} c = \hat{S}^2 \phi.$$

Multiplying with $\hat{S}$ and using $L = \hat{S}^2$, we get

$$L(c - \hat{S}\phi) = 0.$$

Since L is positive definite, this implies that $c - \hat{S}\phi = 0,\ \mu$-a.e. This proves that $\hat{S} b_T$ weakly converges to $\hat{S}\phi$ as $T \to \infty$.

We now prove that the convergence is strong. Write

$$\|\hat{S} b_T - \hat{S}\phi\|^2 = \langle \hat{S} b_T, \hat{S} b_T \rangle - 2\langle \hat{S} b_T, \hat{S}\phi \rangle + \langle \hat{S}\phi, \hat{S}\phi \rangle.$$

Since $\langle \hat{S} b_T, \hat{S}\phi \rangle \to \langle \hat{S}\phi, \hat{S}\phi \rangle$, $\langle \hat{S} b_T, \hat{S} b_T \rangle = \langle L b_T, b_T \rangle$ and $\langle \hat{S}\phi, \hat{S}\phi \rangle = \langle L\phi, \phi \rangle$, we conclude by (A.17) that

$$\lim_{T \to \infty} \|\hat{S} b_T - \hat{S}\phi\|^2 = 0.$$

Since $\langle L(b_T - \phi), b_T - \phi \rangle = \|\hat{S} b_T - \hat{S}\phi\|^2$, the lemma is proved. □

COROLLARY A.1. *$L b_T$ converges weakly to $L\phi$ in $L_2(\mu)$.*

This follows from the Cauchy–Schwarz inequality: if $\psi \in L_2(\mu)$, then

$$\big|\langle L(b_T - \phi), \psi \rangle\big| \le \big\langle L(b_T - \phi), b_T - \phi \big\rangle^{1/2} \langle L\psi, \psi \rangle^{1/2}.$$

COROLLARY A.2.

$$\lim_{T \to \infty} \big\langle L(b_T - \phi), b_T \big\rangle = 0.$$

This follows since

$$\big|\big\langle L(b_T - \phi), b_T \big\rangle\big| \le \big|\big\langle L(b_T - \phi), b_T - \phi \big\rangle\big| + \big|\big\langle L(b_T - \phi), \phi \big\rangle\big|.$$

PROOF OF THEOREM 3. It is easy to see that $\hat{f}_T \in L_2(\mu)$ a.s. By (4.1), in the notation of the proof of Lemma 3.1,

$$\int_B \hat{f}_T(x)\phi(x)\mu(dx) = \langle b_T, f\rangle + \sqrt{T}\int_B \overline{a}_T(y)\nu(dy).$$

We prove the theorem by showing that

$$\lim_{T\to\infty}\langle b_T, f\rangle = \langle \phi, f\rangle \tag{A.18}$$

and

$$\lim_{T\to\infty} E\left[\sqrt{T}\int_B \overline{a}_T(y)\nu(dy)\right]^2 = 0. \tag{A.19}$$

By hypothesis, $f = Lk$, so

$$\langle b_T, f\rangle = \langle b_T, Lk\rangle = \langle Lb_T, k\rangle,$$

and the latter converges to $\langle L\phi, k\rangle = \langle \phi, Lk\rangle = \langle \phi, f\rangle$ by Corollary A.1. Limit (A.18) is proved.

Next, since ν is an orthogonal measure,

$$E\left[\sqrt{T}\int_B \overline{a}_T(y)\nu(dy)\right]^2 = T\int \overline{a}_T(y)^2\mu(dy) = T\|\overline{a}_T\|^2.$$

By (A.12), since $b_T = T\overline{a}_T$,

$$T\|\overline{a}_T\|^2 = \langle L\phi, b_T\rangle - \langle Lb_T, b_T\rangle.$$

An application of Corollary A.2 completes the proof of (A.19). □

PROOF OF COROLLARY 3.2. Using (3.12), (3.13) and (3.14) we write

$$\int_B \tilde{f}_T(x)\phi(x)\mu(dx) - \langle b_T, f\rangle = \sqrt{T}\int_B \overline{a}_T(y)\tilde{\nu}(dy).$$

Thus, by (3.11),

$$\begin{aligned} E\left[\left(\int_B \tilde{f}_T(x)\phi(x)\mu(dx) - \langle b_T, f\rangle\right)^2\right] &= T\int_B \overline{a}_T(y)^2\tilde{\mu}(dy) \\ &\le cT\int_B \overline{a}_T(y)^2\mu(dy), \end{aligned}$$

where c is a bound on $v(y)/\tilde{v}(y)$. The result follows from (A.18) and (A.19). □

PROOF OF THEOREM 4. We begin the proof with a lemma. It is more general than it is required for Theorem 4 but this generality is exploited in the proof of Theorem 5.

LEMMA 4.1. *Under the conditions of Theorem* 4,

$$\lim_{n\to\infty}\sup_{z,\,x\in B}\left|a_n(z,x) - a(z,x)\right| = 0.$$

PROOF. By Theorem 1, the functions a_n and a are elements of the space $C(B\times B)$ of continuous functions on $B\times B$ with uniform norm. We prove, first, that the sequence $\{a_n, n\ge 1\}$ is relatively compact in $C(B\times B)$. For this we check the conditions of the Arzelà–Ascoli theorem. Estimate (A.2) in the proof of Theorem 1 yields, by (3.4) applied to a_n and ρ_n,

$$\sup_{z,x}|a_n(z,x)|\le \sup_{z,x}|\rho_n(z,x)|+\sup_{z,x}\rho_n^2(z,x)\mu_n(B). \tag{A.20}$$

Since $\mu_n\to\mu$ weakly,

$$\sup_n \mu_n(B)<\infty. \tag{A.21}$$

This and the fact that $\rho_n\to\rho$ uniformly, shows that the right-hand side of (A.20) is bounded in n. Thus

$$\sup_n \sup_{z,x}|a_n(z,x)|<\infty, \tag{A.22}$$

verifying the first condition of the Arzelà–Ascoli theorem.

For the second condition, let $\delta>0$ and use the fact that the a_n's are symmetric to write

$$\sup_{\substack{|z_1-z_2|<\delta\\ |x_1-x_2|<\delta}}\bigl|a_n(z_1,x_1)-a_n(z_2,x_2)\bigr|\le 2\sup_{|z_1-z_2|<\delta}\sup_x\bigl|a_n(z_1,x)-a_n(z_2,x)\bigr|.$$

Inequality (A.5) applied to a_n and ρ_n yields

$$\begin{aligned}\sup_x\bigl|a_n(z_1,x)-a_n(z_2,x)\bigr|&\le \sup_x\bigl|\rho_n(z_1,x)-\rho_n(z_2,x)\bigr|\\ &\quad+\sup_y\bigl|\rho_n(z_1,y)-\rho_n(z_2,y)\bigr|\sup_{x,y}\bigl|a_n(x,y)\bigr|\mu_n(B).\end{aligned}$$

The $\sup_{|z_1-z_2|<\delta}$ of the right-hand side converges to 0 as $n\to\infty$ and $\delta\to 0$ by (A.21), (A.22) and since $\rho_n\to\rho$ uniformly. Thus

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\substack{|z_1-z_2|<\delta,\\ |x_1-x_2|<\delta}}\bigl|a_n(z_1,x_1)-a_n(z_2,x_2)\bigr|=0,$$

verifying the second condition of the Arzelà–Ascoli theorem.

Let $\tilde a$ be an accumulation point of $\{a_n, n\ge 1\}$ in $C(B\times B)$. That is, $a_{n'}(y,x)\to\tilde a(y,x)$ uniformly on $B\times B$ for some subsequence $\{n'\}$. We have

$$\begin{aligned}&\left|\int_B \rho_{n'}(z,y)a_{n'}(y,x)\mu_{n'}(dy)-\int_B\rho(z,y)\tilde a(y,x)\mu(dy)\right|\\ &\quad\le\int_B\bigl|\rho_{n'}(z,y)-\rho(z,y)\bigr|\,\bigl|a_{n'}(y,x)\bigr|\mu_{n'}(dy)\\ &\qquad+\int_B\bigl|\rho(z,y)\bigr|\,\bigl|a_{n'}(y,x)-\tilde a(y,x)\bigr|\mu_{n'}(dy)\\ &\qquad+\left|\int_B\rho(z,y)\tilde a(y,x)\mu_{n'}(dy)-\int_B\rho(z,y)\tilde a(y,x)\mu(dy)\right|.\end{aligned} \tag{A.23}$$

The first integral on the right of (A.23) tends to 0 uniformly in z, $x \in B$ as $n' \to \infty$ by (A.21), (A.22) and since $\rho_{n'} \to \rho$ uniformly. The second integral tends to 0 uniformly in z, $x \in B$ as $n' \to \infty$ since $a_{n'} \to \tilde{a}$ uniformly, ρ is bounded and (A.21) holds.

Consider the last integral. Since $\tilde{a}(y, x)$, $y, x \in B$ is continuous,

$$\rho(z, y)\tilde{a}(y, x)$$

is continuous in y for all $z, x \in B$. Moreover, the family $\{\rho(z, y)\tilde{a}(y, x), y \in B\}$, $z, x \in B$, of functions is equicontinuous; that is,

$$\lim_{\delta \to 0} \sup_{\substack{z \in B \\ x \in B}} \sup_{|y_1 - y_2| < \delta} \left|\rho(z, y_1)\tilde{a}(y_1, x) - \rho(z, y_2)\tilde{a}(y_2, x)\right| = 0.$$

This implies [see, e.g., Billingsley (1968), problem 8, Section 2] that since $\mu_n \to_w \mu$,

$$\int_B \rho(z, y)\tilde{a}(y, x)\mu_{n'}(dy) \to \int_B \rho(z, y)\tilde{a}(y, x)\mu(dy)$$

as $n' \to \infty$ uniformly in $z, x \in B$. Hence

$$\sup_{\substack{z \in B \\ x \in B}} \left| \int_B \rho_{n'}(z, y)a_{n'}(y, x)\mu_{n'}(dy) - \int_B \rho(z, y)\tilde{a}(y, x)\mu(dy) \right| \to 0$$

as $n' \to \infty$. Thus the left-hand side of (A.23) converges to 0 uniformly in (z, x). Since $\rho_n \to \rho$ uniformly and the a_n and ρ_n satisfy (3.4), we see that $\tilde{a}$ satisfies the equation

$$\int_B \rho(z, y)\tilde{a}(y, x)\mu(dy) + \tilde{a}(z, x) = \rho(z, x),$$

that is, $\tilde{a} = a$ by Theorem 1. The lemma is proved. □

PROOF OF THEOREM 4. Let

$$\overline{a}(y) = \int_B a(y, x)\phi(x)\mu(dx), \qquad y \in B$$

and

$$\overline{a}_n(y) = \int_B a_n(y, x)\phi(x)\mu_n(dx), \qquad y \in B.$$

By the definitions of $\overline{a}_n$, $\hat{f}_n$, $\overline{a}$ and $\hat{f}$, we have

$$\langle \hat{f}_n, \phi \rangle_n = \int_B \overline{a}_n(y)f(y)\mu_n(dy) + \int_B \overline{a}_n(y)\nu_n(dy)$$

and

$$\langle \hat{f}, \phi \rangle = \int_B \overline{a}(y)f(y)\mu(dy) + \int_B \overline{a}(y)\nu(dy).$$

Lemma 4.1 easily implies that

$$\lim_{n\to\infty} \sup_y \left|\overline{a}_n(y) - \overline{a}(y)\right| = 0, \tag{A.24}$$

and hence

$$\lim_{n\to\infty} \int_B \overline{a}_n(y) f(y) \mu_n(dy) = \int_B \overline{a}(y) f(y) \mu(dy).$$

Also, since the ν_n are orthogonal measures, for $\varepsilon > 0$,

$$\begin{aligned} P\left(\left|\int_B (\overline{a}_n(y) - \overline{a}(y))\nu_n(dy)\right| > \varepsilon\right) &\le \varepsilon^{-2} E\left[\int_B (\overline{a}_n(y) - \overline{a}(y))\nu_n(dy)\right]^2 \\ &= \varepsilon^{-2} \int_B (\overline{a}_n(y) - \overline{a}(y))^2 \mu_n(dy), \end{aligned}$$

which tends to 0 as $n \to \infty$ by (A.24) and (A.21). So the theorem would be proved if

$$\int_B \overline{a}(y)\nu_n(dy) \to_d \int_B \overline{a}(y)\nu(dy). \tag{A.25}$$

Since the ν_n are Gaussian orthogonal measures, for any $t_1, t_2, \dots, t_k$, and $A_1, A_2, \dots, A_k$ where $A_j \cap A_{j'} = \varnothing$ if $j \ne j'$,

$$E \exp\left(i \sum_{\ell=1}^k t_\ell \nu_n(A_\ell)\right) = \exp\left(-\tfrac{1}{2} \sum_{\ell=1}^k t_\ell^2 \mu_n(A_\ell)\right),$$

which implies that for any continuous function g on B,

$$E \exp\left(i \int_B g(y)\nu_n(dy)\right) = \exp\left(-\tfrac{1}{2} \int_B g^2(y)\mu_n(dy)\right),$$

and likewise,

$$E \exp\left(i \int_B g(y)\nu(dy)\right) = \exp\left(-\tfrac{1}{2} \int_B g^2(y)\mu(dy)\right).$$

Hence, since $\overline{a}$ is bounded and continuous and $\mu_n \to_w \mu$, for $t \in R$,

$$\begin{aligned} \lim_{n\to\infty} E \exp\left(it \int_B \overline{a}(y)\nu_n(dy)\right) &= \lim_{n\to\infty} \exp\left(-\tfrac{1}{2} t^2 \int_B \overline{a}(y)^2 \mu_n(dy)\right) \\ &= \exp\left(-\tfrac{1}{2} t^2 \int_B \overline{a}(y)^2 \mu(dy)\right) \\ &= E \exp\left(it \int_B \overline{a}(y)\nu(dy)\right), \end{aligned}$$

proving (A.25). The theorem is proved. □

APPENDIX B

We fix n design points $\{x_1, x_2, \ldots, x_n\} \in B$ and search for a corresponding set of simulation times $\{t_1, t_2, \ldots t_n\}$ that solves

$$\text{maximize } h(t_1, t_2, \ldots, t_n) = \psi' \Gamma^{-1} \psi, \tag{B.1}$$

$$\text{subject to } \sum_{i=1}^{n} t_i = T, \tag{B.2}$$

$$t_i \geq 0, \qquad i = 1, 2, \ldots, n. \tag{B.3}$$

The nonnegativity constraint (B.3) can be handled by defining

$$s_i^2 = t_i, \qquad i = 1, 2, \ldots, n. \tag{B.4}$$

Using Lagrange multipliers, we find that for some λ,

$$\frac{1}{s_i} \frac{\partial h}{\partial s_i} = \lambda, \qquad i = 1, 2, \ldots, n. \tag{B.5}$$

Using (B.4), we can write (B.2) and (B.5) as

$$G(s_1, s_2, \ldots, s_n) \equiv \begin{bmatrix} \frac{1}{s_1}\frac{\partial h}{\partial s_1} - \frac{1}{s_2}\frac{\partial h}{\partial s_2} \\ \frac{1}{s_1}\frac{\partial h}{\partial s_1} - \frac{1}{s_3}\frac{\partial h}{\partial s_3} \\ \vdots \\ \frac{1}{s_1}\frac{\partial h}{\partial s_1} - \frac{1}{s_n}\frac{\partial h}{\partial s_n} \\ \sum_{i=1}^{n} s_i^2 - T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \tag{B.6}$$

which can be solved by Newton's method. If S is diagonal then $\partial\Gamma/\partial t_k$ has only a single nonzero entry, $(\partial\Gamma/\partial t_k)_{kk} = -(v_k/t_k^2)$ [where $v_k \equiv v(x_k)$]. Using $\partial\Gamma^{-1}/\partial t_k = -\Gamma^{-1}(\partial\Gamma/\partial t_k)\Gamma^{-1}$, we obtain

$$\frac{\partial h}{\partial t_k} = \frac{v_k}{t_k^2}(\Gamma^{-1}\psi)_k^2, \tag{B.7}$$

$$\frac{\partial^2 h}{\partial t_k^2} = \frac{2v_k}{t_k^3}\left(\frac{v_k}{t_k}\Gamma_{kk}^{-1} - 1\right)(\Gamma^{-1}\psi)_k^2 \tag{B.8}$$

and

$$\frac{\partial^2 h}{\partial t_k \partial t_\ell} = 2\frac{v_k v_\ell}{t_k^2 t_\ell^2}\Gamma_{k\ell}^{-1}(\Gamma^{-1}\psi)_k(\Gamma^{-1}\psi)_\ell. \tag{B.9}$$

The Jacobian of $G(s_1, s_2, \ldots, s_n)$ can be obtained from (B.4) and (B.6)–(B.9) via the chain rule.

REMARK B.1. Clearly (B.7) implies that

$$\frac{\partial h}{\partial t_k} > 0. \tag{B.10}$$

Let $\hat{\Gamma}$ be the covariance matrix for $\{Y_1, Y_2, \ldots, Y_n\} \setminus \{Y_k\}$. Writing Γ as

$$\Gamma = \begin{bmatrix} \hat{\Gamma} & \gamma(x_k) \\ \gamma(x_k)' & \rho(x_k, x_k) + \dfrac{v_k}{t_k} \end{bmatrix}$$

we obtain [using the block matrix inverse formula, e.g., Searle, (1982)]

$$\Gamma_{kk}^{-1} = \left(\rho(x_k, x_k) - \gamma(x_k)' \hat{\Gamma}^{-1} \gamma(x_k) + \frac{v_k}{t_k} \right)^{-1}.$$

Since $\rho(x_k, x_k) - \gamma(x_k)' \hat{\Gamma}^{-1} \gamma(x_k)$ is the variance of $Z(x_k) - \hat{Z}(x_k)$ based on the simulation data at $\{x_1, x_2, \ldots, x_n\} \setminus \{x_k\}$, it follows that

$$\Gamma_{kk}^{-1} < \frac{t_k}{v_k},$$

and from (B.8) we see that

$$\frac{\partial^2 h}{\partial t_k^2} < 0. \tag{B.11}$$

From (B.10) and (B.11) it seems plausible to conjecture that no discrete design can be optimal.

Acknowledgments. We thank the referees and Associate Editor for their careful reading of our paper. Their insightful comments and pointers to related work greatly improved the presentation of our work.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

CURRIN, C., MITCHELL, T., MORRIS, M. and YLVISAKER, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86** 953–963.

DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Methuen, London.

HUTSON, V. and PYM, J. S. (1980). *Applications of Functional Analysis and Operator Theory*. Academic Press, London.

IGLEHART, D. L. (1978). The regenerative method for simulation analysis. In *Current Trends in Programming Methodology—Software Modeling* (K. M. Chandy and R. T. Yeh, eds.). Prentice-Hall, Englewood Cliffs, NJ.

KIEFER, J. and WOLFOWITZ, J. (1959). "Optimum designs in regression problems. *Ann. Math. Statist.* **30** 289–312.

KOEHLER, J. R. and OWEN, A. B. (1996). Computer experiments. In *Handbook of Statistics* (S. Ghosh and C. R. Rao, eds.) **13** 261–308. North-Holland, Amsterdam.

PLASKOTA, L. (1992). Function approximation and integration on the Wiener space with noisy data. *J. Complexity* **8** 301–323.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York.
RIPLEY, B. D. (1981). *Spatial Statistics*. Wiley, New York.
RITTER, K., WASILKOWSKI, G. W. and WOZNIAKOWSKI, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks–Ylvisaker conditions. *Ann. Appl. Probab.* **5** 518–540.
SACKS, J., SCHILLER, S. B. and WELCH, W. J. (1989). Designs for computer experiments. *Technometrics* **1** 41–47.
SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–423.
SEARLE, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.
SHIRYAYEV, A. N. (1984). *Probability*. Springer, New York.
SIMON, B. (1997). Convergence of 'natural' rootfinding procedures. Unpublished manuscript.
STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63.
TRICOMI, F. G. (1985). *Integral Equations*. Dover, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF COLORADO
P.O. BOX 173364
DENVER, COLORADO 80217
E-MAIL: jkoehler@carbon.cudenver.edu
puhalski@math.cudenver.edu
bsimon@math.cudenver.edu