Ann. Inst. Statist. Math.
32 (1980), Part A, 275-281

DESIGNS OF Φ-OPTIMAL CONTROL FOR SECOND-ORDER PROCESSES

DER-SHIN CHANG AND YUANG-CHIN CHIANG

(Received Nov. 6, 1979)

Abstract

Consider a realization of the process $y(t)=\sum_{k=1}^{n} \theta_k f_k(t)+\xi(t)$ on the interval $T=[0,1]$ for functions $f_1(t), f_2(t), \cdots, f_n(t)$ in $H(R)$, the reproducing kernel Hilbert space with reproducing kernel $R(s,t)$ on $T\times T$, where $R(s,t)=\mathrm{E}\,[\xi(s)\xi(t)]$ is assumed to be continuous and known. Problems of the selection of functions $\{f_k(t)\}_{k=1}^n$ to be Φ-optimal design are given, and an unified approach to the solutions of D-, A-, E- and D_s-optimal design problems are discussed.

1. Introduction

If a stochastic process

$$(1) \qquad y(t)=\sum_{k=1}^{n} \theta_k f_k(t)+\xi(t)\,, \qquad t\in T=[0,1]$$

is given with the noise process $\xi(t)$ having zero mean and known continuous covariance kernel $R(s,t)=\mathrm{E}\,[\xi(s)\xi(t)]$, $(s,t)\in T\times T$. Let $H(R)$ be the reproducing kernel Hilbert space (RKHS) with reproducing kernel (RK) $R(s,t)$ on $T\times T$, and let $\{f_k(t)\}_{k=1}^n$ be a linearly independent set of functions in $H(R)$. Then, by the Gauss-Markov theory of continuous time series, we obtain for $f=(f_1(t), f_2(t), \cdots, f_n(t))'$ (throughout this paper primes will denote transposes) the minimum variance unbiased estimate $\hat{\theta}=M^{-1}(f)\ (\langle y, f_1\rangle\sim, \cdots, \langle y, f_n\rangle\sim)'$, and its covariance matrix $\mathrm{Cov}\,[\hat{\theta}]=M^{-1}(f)$, where $\hat{\theta}=(\hat{\theta}_1, \cdots, \hat{\theta}_n)'$, $M(f)=[m_{ij}]_{i,j=1}^n$, $m_{ij}=\langle f_i, f_j\rangle_R$ if $\{f_k(t)\}_{k=1}^n\in H(R)$, and $\langle y, f_k\rangle\sim$, $k=1,2,\cdots,n$ are defined as if $y(t)$ were an element in $H(R)$. In [1], the author has given the definition of D-optimal, A-optimal and weighted optimal designs of the functions $\{f_k\cdot$

AMS 1970 subject classifications: Primary 62K05; secondary 93E20.
Key words and phrases: Second-order process, continuous sense of Gauss-Markov theory, reproducing kernel Hilbert space, Φ-optimal designs.

$(t)\}_{k=1}^n$ in some set $X \subset H(R)$, and give the analytic expression of optimal solution of $\{f_k(t)\}_{k=1}^n$ in $H(R)$. But they were treated separately and the methods were tedious also. In this paper, we use a result in [5] as a tool for obtaining an unified approach to these problems, and hopefully we can reach a more general results than [1] and make D-, A-, E-, D_s-optimal design to be our special cases. In Section 2, the criterion of Φ-optimal design of second-order process (1) will be given. In Section 3, we prove the tool theorem and use this theorem to describe the solution of D-, A- and E-optimal design problems. In Section 4, we will do the same approach to D_s-optimal design.

2. Design criterion

Suppose that (1) is given, then it is well kown (see [6], [7]) that the space of functions generated by $\{R_t(\cdot),\ t \in T \mid R_t(t')=R(t', t)\}$ is an RKHS denoted by $H(R)$, with RK $R(s, t)$ on $T\times T$. Since $R(s, t)$ is symmetric and positive definite (p.d.), then, by Mercer's theorem (see [8], pp. 242–246), we know that there exists a set of orthonormal functions $\{\phi_v(t)\}_{v=1}^\infty$ in $\mathcal{L}^2[T]$ and corresponding sequence of positive real numbers $\{\eta_v\}_{v=1}^\infty$ such that

$$R(s, t)=\sum_{v=1}^{\infty} \eta_v \phi_v(s)\phi_v(t)$$

is uniformly convergent in $T\times T$ if $R(s, t)$ is continuous. Also that the inner product in $H(R)$ is

$$\langle g, h\rangle_R=\sum_{v=1}^{\infty} g_v h_v/\eta_v \ ,$$

where $g_v=(g, \phi_v)_{\mathcal{L}^2}$ and $h_v=(h, \phi_v)_{\mathcal{L}^2}$, for any $g, h \in H(R)$. That is,

$$H(R)=\left\{h \,\middle|\, \sum_{v=1}^{\infty} h_v^2/\eta_v<\infty,\ h_v=(h, \phi_v)_{\mathcal{L}^2}\right\} .$$

Assume further that a set of linear independent functions $\{f_k(t)\}_{k=1}^n$ in $H(R)$ is given. Then, by [6] and [7], we have for $\theta=(\theta_1,\cdots, \theta_n)'$ and $f=(f_1(t),\cdots, f_n(t))'$ the minimum variance unbiased estimate

$$\hat{\theta}=M^{-1}(f)\cdot(\langle y, f_1\rangle\sim,\cdots, \langle y, f_n\rangle\sim)'$$

with $\mathrm{Cov}\,[\hat{\theta}]=M^{-1}(f)$, where

$$M(f)=[m_{ij}]_{i,j=1}^n\ , \qquad m_{ij}=\langle f_i, f_j\rangle_R,\ i, j=1,\cdots, n \tag{2}$$

and $\langle y, f_k\rangle\sim=\sum_{v=1}^{\infty}(f_{kv}y_v)/\eta_v$, $k=1, 2,\cdots, n$, with $y_v=(y, \phi_v)_{\mathcal{L}^2}$, the stochastic integral of $y(t)$ with respect to weight function $\phi_v(t) \in \mathcal{L}^2[T]$, $v=1$,

$2, \cdots$.

Since the Gauss-Markov estimate of θ needs the invertibility of $M(f)$, our discussion of designs will be restricted in the class of invertible matrices $M(f)$'s and in addition, since $M(f)$ is nonnegative definite (n.n.d.) (see [3]), we will furtherly restrict our discussion in p.d. matrices $M(f)$'s.

Now, following [1], [2], and [5], we can give our design criterion in following

DEFINITION. Let Φ be an real-valued function on some subset of p.d. matrices. A matrix C^* in this subset is called Φ-optimal if C^* minimizes $\Phi(C)$ for all C in this subset.

DEFINITION. Suppose that (1) and some set X in $H(R)$ is given. An experiment with $\{f_k^*(t)\}_{k=1}^n$ in X is said to be Φ-optimal (or Φ-optimal design) if $M(f^*)$ minimizes $\Phi(M(f))$ for all possible choices $\{f_k(t)\}_{k=1}^n$ from X.

Examples. Suppose Φ_t be given as in the following.

(i) $\Phi_0(M(f))=|M^{-1}(f)|$; an experiment with $\{f_k^*(t)\}_{k=1}^n$ which is Φ_0-optimal is also called D-optimal.

(ii) $\Phi_1(M(f))=\operatorname{tr} M^{-1}(f)$; an experiment with $\{f_k^*(t)\}_{k=1}^n$ which is Φ_1-optimal is also called A-optimal.

(iii) $\Phi_\infty(M(f))=\lambda_1(M^{-1}(f))$, where $\lambda_1(M^{-1}(f))$ is the largest eigenvalue of $M^{-1}(f)$; an experiment with $\{f_k^*(t)\}_{k=1}^n$ which is Φ_∞-optimal is also called E-optimal.

(iv) Let $M_s^*(f)$ be defined by

$$M_s^*(f)=[I_s \,\vdots\, 0]M^{-1}(f)\left(\frac{I_s}{0}\right),$$

where I_s is $s\times s$ identity matrix, 0 is zero matrix with appropriate size. Consider

$$\psi(M_s^*(f))=\Phi_0(M_s^*(f));$$

then an experiment with $\{f_k^*(t)\}_{k=1}^n$ which is ψ-optimal is also called D_s-optimal.

3. Optimality tool

In view of the examples of Section 2, we may extract the characteristics of Φ which guarantee the existence of Φ-optimal. Let $\mathscr{B}_n$ consist of the $n\times n$ p.d. matrices, and

$$\Phi:\ \mathscr{B}_n\to(-\infty,+\infty)$$

satisfies

(3)
(a) Φ is convex,
(b) $\Phi(bC)$ is nondecreasing in the scalar $b>0$,
(c) Φ is orthogonal invariant, i.e. for orthogonal matrix P, $\Phi(P'CP)=\Phi(C)$.

Let $\mathscr{D}$ be some index set, then, similarly as in [5], we have the following optimality tool, which is slightly different from [5] owing to the condition (c) being changed.

THEOREM 1. *If the class* $C=\{C_d;\ d\in\mathscr{D}\}\subset\mathscr{B}_n$ *contains a* C_{d^*} *which is a multiple of* I_n*, and maximizing* $\operatorname{tr} C_d$ *for* $d\in\mathscr{D}$*, then* C_{d^*} *is* Φ*-optimal for every* Φ *satisfying* (3).

PROOF. Suppose that there exists a $C_{d'}$ in C such that $\Phi(C_{d'})<\Phi(C_{d^*})$ and $P'C_{d'}P=[\lambda_i\delta_{ij}]_{i,j=1}^n$, where $P'P=I_n$ and $\lambda_1,\cdots,\lambda_n$ are the eigenvalues of $C_{d'}$. If σP is obtained from P by permuting the columns of P according to the permutation σ of the first n integers of natural numbers, then clearly $(\sigma P)'C_{d'}(\sigma P)=[\lambda_{\sigma_i}\delta_{ij}]_{i,j=1}^n$, where σ_i is the integer obtained from i moved by σ. Let $\overline{C_{d'}}=\sum_\sigma(\sigma P)'C_{d'}(\sigma P)/n!$. Since there are $(n-1)!\sigma$'s among $n!$ mapping i unchanged for every i, $1\leqq i\leqq n$, we easily have

$$\overline{C_{d'}}=\frac{1}{n}\sum_{i=1}^n\lambda_i I_n\ . \tag{4}$$

Although $\overline{C_{d'}}$ is not necessarily in C (but indeed in $\mathscr{B}_n$), we have, by (3) (c) and (a),

$$\begin{aligned}\Phi(C_{d^*})>\Phi(C_{d'})&=\sum_\sigma\Phi((\sigma P)'C_{d'}(\sigma P))/n!\\&\geqq\Phi(\sum_\sigma(\sigma P)'C_d(\sigma P)/n!)\\&=\Phi(\overline{C_{d'}})\ .\end{aligned}$$

Since C_{d^*} is multiple of I_n, then, by (4), $\overline{C_{d'}}$ is of the form bC_{d^*} for some $b>0$. More since $\operatorname{tr}\overline{C_{d'}}=\operatorname{tr}C_{d'}=\sum_{i=1}^n\lambda_i=b\cdot\operatorname{tr}C_{d^*}$, $\operatorname{tr}C_{d^*}$ maximizing $\operatorname{tr}C_d$ for $d\in\mathscr{D}$ implies $b\leqq 1$. But then, by (3) (b), $\Phi(C_{d'})\geqq\Phi(\overline{C_{d'}})\geqq\Phi(C_{d^*})$, which contradicts $\Phi(C_{d'})<\Phi(C_{d^*})$. This completes the proof.

Now rather than treat D- and A-optimal designs separately as in [1], we may use the theorem stated above to solve then as well as E- and D_s-optimal designs in an unified way. Because of the length of introducing the definition of D_s-optimal design, we will separate it for discussion in Section 4.

LEMMA 1. *Φ_0, Φ_1 and Φ_∞ all satisfies* (3).

This result is well known, and incidentally we can assert the convexity of $\log|A^{-1}|$ instead of $|A^{-1}|$ for proving Φ_0.

Application. Suppose we are given model (1), and $X=\{g|g\in H(R), \|g\|_R^2\leqq L\}$, where L is a positive number. Then $\operatorname{tr} M(f)=\sum_{i=1}^n m_{ii}=\sum_{i=1}^n \langle f_i, f_i\rangle_R=\sum_{i=1}^n \|f_i\|_R^2\leqq nL$ for $\{f_k(t)\}_{k=1}^n\subset X$. Now let $f_k^*(t)=\sqrt{L\eta_k}\phi_k(t)$, $k=1,\cdots,n$ as in [1], then $M(f^*)=LI_n$, and $\operatorname{tr} M(f^*)=nL$ reaches the maximum value of $\operatorname{tr} M(f)$ in X. Therefore, by Theorem 1 and Lemma 1, we know that $\{f_k^*(t)\}_{k=1}^n$ is Φ_0-, Φ_1- and Φ_∞-optimal simultaneously in X with

$$\min_{\{f_k\}\subset X}|M^{-1}(f)|=|M^{-1}(f^*)|=L^{-n},$$

$$\min_{\{f_k\}\subset X}\operatorname{tr} M^{-1}(f)=\operatorname{tr} M^{-1}(f^*)=n/L$$

and

$$\min_{\{f_k\}\subset X}\lambda_1(M^{-1}(f))=\lambda_1(M^{-1}(f^*))=1/L .$$

Remark. The above solutions of Φ_0- and Φ_1-optimal designs of model (1) are same as in [1]. But the weighted optimal design treated there is not a criterion satisfying (3) because it is not invariant under orthogonal transformation.

4. D_s-optimal

Although every parameter of $\theta_1,\cdots,\theta_n$ exert influence upon the investigated stochastic process $y(t)$ of (1), an experimenter frequently is interested in only some of the parameters, say, $\theta_1,\cdots,\theta_s$, $s<n$. In such a case, it turns out that D-optimum (Φ_0-optimum) is somewhat inappropriate and does not reflect the needs of the experimenter. This leads us to the consideration of the "generalized variance" of the estimates of the first s parameters $\theta_1,\cdots,\theta_s$. Since the "generalized variance" of the estimates of the first s parameters is also influenced by the estimates of the rest parameters $\theta_{s+1},\cdots,\theta_n$, thus it serves as a different criterion of optimality as discussed in [4] and [2]. We call such a criterion of "D_s-optimal".

Consider the class of p.d. matrices $M(f)$, $f=(f_1(t),\cdots,f_n(t))'$ such that $\{f_k(t)\}_{k=1}^n\subset X$. Let

$$(5)\qquad M(f)=\begin{bmatrix} M_1(f) & M_2'(f) \\ M_2(f) & M_3(f) \end{bmatrix}$$

where $M_1(f)$ is an $s\times s$ matrix, $M_3(f)$ is an $(n-s)\times(n-s)$ matrix. Then, following the well known Frobenius formula (see [3], p. 16), we can define the following.

DEFINITION. For any $\{f_k(t)\}_{k=1}^n$ contained in X, let its information matrix be $M(f)$ as in (2) and be partitioned as in (5). Let

$$(6)\qquad M_s^*(f)=[I_s \vdots 0]M^{-1}(f)\left(\frac{I_s}{0}\right)=M_1(f)-M_s'(f)M_3^{-1}(f)M_2(f)\ .$$

Then a design $\{f_k^*(t)\}_{k=1}^n$ in X is called D_s-optimal if $\{f_k^*(t)\}_{k=1}^n$ minimizes $\psi(M_s^*(f))=\Phi_0(M_s^*(f))=|M_s^{*-1}(f)|$ for all possible choices $\{f_k(t)\}_{k=1}^n$ in X.

Similarly as in Lemma 1, we have

LEMMA 2. *ψ satisfies* (3).

Now by Lemma 2 and apply Theorem 1 to ψ, we have following.

THEOREM 2. *Suppose* (1) *is given,* $X=\{g(t),\ t\in T\,|\,g\in H(R),\ \|g\|_R^2\leqq L\}$ *and* $M^{-1}(f)$ *the dispersion matrix of the Gauss-Markov estimate of* θ *as in Section* 1. *Then, we have*

$$\min_{\{f_k\}_{k=1}^n\subset X}|M_s^{*-1}(f)|=L^{-s}\ ,$$

which is attainable at $f_k^*(t)=\sqrt{L\eta_k}\,\phi_k(t)$, $k=1,\cdots,s$ *and any linearly independent set of functions* $\{f_{s+v}^*(t)\}_{v=1}^{n-s}$ *orthogonal to* $\{f_k^*(t)\}_{k=1}^s$.

PROOF. By direct computations of $\langle f_i^*, f_j^*\rangle_R$, $i, j=1, 2,\cdots, n$, we have

$$M(f^*)=\begin{bmatrix} LI_s & 0 \\ 0 & M_3(f^*) \end{bmatrix},$$

which implies

$$(7)\qquad M_s^*(f^*)=LI_s\ .$$

Furtherly, by (6) and same arguments stated in [4] (see Lemma 6.3, p. 801), we know that for any $\{f_k(t)\}_{k=1}^n\subset X$,

$$(8)\qquad M_s^*(f)\leqq M_1(f)\ .$$

Thus, by (7) and (8), we have

$$(9)\qquad sL=\operatorname{tr} M_1(f^*)=\operatorname{tr} M_s^*(f^*)\leqq\max_{\{f_k\}\subset X}\operatorname{tr} M_1(f)=sL\ ,$$

the last equality follows from $\operatorname{tr} M_1(f)=\sum_{i=1}^{s} m_{ii} \leqq sL$ and the maximum attained as $f_k^*(t)=\sqrt{L\eta_k}\phi_k(t)$, $k=1,\cdots,s$, which says that $\operatorname{tr} M_s^*(f^*)$ maximizing $\operatorname{tr} M_s^*(f)$ in X. Therefore, by (7), (9), Lemma 2 and Theorem 1, we justify the validity of the result, which is another approach to [2].

Acknowledgement

We wish to thank Professor C. S. Cheng for suggesting this approach.

NATIONAL TSING HUA UNIVERSITY

REFERENCES

[1] Chang, D. S. (1979). Designs of optimal control for a regression problem, *Ann. Statist.*, **7**, 1078-1085.

[2] Chang, D. S. (1981). Design of D_s- and E-optimal control for a regression problem, to appear in *Bulletin of the Institute of Mathematics Academia Sinica*.

[3] Federov, V. V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

[4] Karlin, S. and Studden, W. J. (1966). Optimal experimental design, *Ann. Math. Statist.*, **37**, 783-815.

[5] Kiefer, J. (1975). Construction and optimality of generalized Youden designs, *A Survey of Statistical Design and Linear Models* (ed. Srivastava, J. N.), North-Holland.

[6] Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I, *Technical Report* No. 23, Department of Statistics, Stanford University.

[7] Parzen, E. (1961). An approach to time series analysis, *Ann. Math. Statist.*, **32**, 951-989.

[8] Riesz, F. and Nagy, B. Sz. (1956). *Functional Analysis*, Ungar, New York.