JOURNAL OF APPROXIMATION THEORY 47, 17–25 (1986)

Average Case Optimal Algorithms in Hilbert Spaces

G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI

Institute of Informatics, University of Warsaw, Warsaw, Poland, and Department of Computer Science, Columbia University, New York, New York 10027, U.S.A.

Communicated by Oved Shisha

Received November 14, 1983; revised May 13, 1985

We study optimal algorithms and optimal information for an average case model. This is done for linear problems in a separable Hilbert space equipped with a probability measure. We show, in particular, that for any measure a (linear) spline algorithm is optimal among *linear* algorithms. The spline algorithm is defined in terms of the covariance operator of the measure. We provide a condition on the measure which guarantees that the spline algorithm is optimal among *all* algorithms. The problem of optimal information is also solved.

1. INTRODUCTION

In two recent monographs [5, 6], optimal reduction of uncertainty for a *worst case model* was studied. In [7] a corresponding study for an *average case model* was initiated. In particular, it is proven in [7] that the same algorithm and information are optimal in the worst case and average case models, and that adaptive information is not more powerful than nonadaptive information. For the case of nonadaptive information, this result was studied by Micchelli [3] using a more general error criterion.

Both [7, 3] deal with the average case model for linear problems defined on finite dimensional Hilbert spaces. In this paper we study the average case model for linear problems in separable Hilbert spaces.

The author of [3] writes in a short subsection "Average Case Optimality: The Infinite Dimensional Case" that his proofs from the finite dimensional case immediately extend to separable Hilbert spaces equipped with Gaussian measures. This is done by restricting a linear problem defined on a separable Hilbert space H to the linear subspace $S_\mu^{1/2}(H)$, where S_μ is a correlation operator of Gaussian measure μ. Then it is

0021-9045/86 $3.00

claimed that orthogonal projections with respect to the new norm $\|x\|_0 = \|S_\mu^{-1/2}x\|$, $\|\cdot\|$ being a norm in H, are optimal algorithms.

It seems to us that the author of [3] overlooked one important fact. The restriction of the original problem defined on the whole space H to the subspace $S_\mu^{1/2}(H)$ is not appropriate. It is known that for the infinite dimensional space H, the measure of $S_\mu^{1/2}(H)$ is zero. Hence restricting the problem as in [3] means considering optimality on a set of measure zero. Then the error of any algorithm is zero and all algorithms are optimal. This indicates that average case optimality on $S_\mu^{1/2}(H)$ is meaningless.

The purpose of this paper is to study linear problems in a separable Hilbert space equipped with a probability (not necessarily Gaussian) measure μ.

We seek optimal algorithms and optimal information. The following results are obtained:

1. For all measures μ: In the class of linear algorithms, a spline algorithm, defined in terms of the covariance operator of the measure μ, is optimal.

2. Let μ be any measure such that

$$\mu(D(B)) = \mu(B) \tag{1.1}$$

where D is a certain linear mapping and B is any Borel set. In the class of all algorithms, the spline algorithm is optimal.

3. For all measures μ, optimal information is obtained for the class of linear algorithms. If μ satisfies (1.1) then optimal information for the class of linear algorithms is also optimal for the class of all algorithms.

The measures which satisfy (1.1) include Gaussian measures with mean zero and the measures studied in [3, 7] for the finite dimensional case.

We now comment on the relation between optimal algorithms and orthogonal projections. For simplicity assume that μ is a Gaussian measure with mean zero. Then the spline algorithm is optimal. If the dimension of the Hilbert space H is *finite*, then the spline algorithm may be viewed as an orthogonal projection as was pointed out in [3]. If the dimension of the Hilbert space H is *infinite* then the spline algorithm may be viewed as an orthogonal projection *only* on the linear space $S_\mu^{1/2}(H)$ whose measure is zero. In this case optimal algorithms in the worst case may be significantly different than optimal algorithms in the average case. We conclude that worst and average case cannot be unified by the techniques used in [3].

This paper is a shortened version of [9], which contains a detailed discussion of the concepts and proofs presented here.

2. Formulation of the Problem

Let F_1 and F_2 be real separable Hilbert spaces. Let

$$S: F_1 \to F_2 \tag{2.1}$$

be a continuous linear operator. We assume that F_1 is equipped with a probability measure μ defined on Borel sets of F_1 such that

$$\int_{F_1} \|f\|^2 \mu(df) < +\infty. \tag{2.2}$$

Without loss of generality we assume that the mean element of μ is zero, $\int_{F_1} (f, x)\, \mu(df) = 0$, $\forall x \in F_1$, and the covariance operator S_μ is F_1 positive definite, i.e.,

$$(S_\mu x, y) = \int_{F_1} (f, x)(f, y)\, \mu(df), \qquad \forall x, y \in F_1 \tag{2.3}$$

and $S_\mu = S_\mu^* > 0$, $\text{trace}(S_\mu) = \int_{F_1} \|f\|^2 \mu(df) < +\infty$, see [1, 4].

We want to approximate Sf knowing Nf, where

$$N: F_1 \to \mathbb{R}^n \tag{2.4}$$

is a continuous linear operator. We call N an *information* operator. Without loss of generality we assume that

$$Nf = [(f, g_1), (f, g_2), \ldots, (f, g_n)] \tag{2.5}$$

where $(S_\mu g_i, g_j) = \delta_{ij}$.

We seek an approximation to Sf by $\varphi(Nf)$, where

$$\varphi: N(F_1) = \mathbb{R}^n \to F_2 \tag{2.6}$$

is called an (*idealized*) *algorithm using information* N. Assuming that φ is Borel measurable, the average error of φ is defined as

$$e^{\text{avg}}(\varphi, N) = \left\{ \int_{F_1} \|Sf - \varphi(Nf))\|^2 \mu(df) \right\}^{1/2}. \tag{2.7}$$

3. Algorithms with Minimal Average Error

We first consider linear algorithms using information N. That is, for $N(f) = [(f, g_1), \ldots, (f, g_n)]$ with $(S_\mu g_i, g_j) = \delta_{ij}$, we consider φ of the form

$$\varphi(Nf) = \sum_{i=1}^{n} (f, g_i)\, w_i \tag{3.1}$$

for some $w_i \in F_2$. An example of such an algorithm is provided by the spline algorithm

$$\varphi^s(Nf) = \sum_{i=1}^{n} (f, g_i)\, SS_\mu g_i. \tag{3.2}$$

Note that $\varphi^s(Nf) = S\sigma$, where $\sigma = \sigma(Nf) = \sum_{i=1}^{n} (f, g_i)\, S_\mu g_i$ is the unique solution of the following problem

$$\begin{aligned} &N(\sigma) = Nf \\ &(S_\mu^{-1}\sigma, \sigma) = \inf\{(S_\mu^{-1} g, g) : g \in S_\mu(F_1),\ N(g) = N(f)\}. \end{aligned} \tag{3.3}$$

The solution of (3.3) is called in the literature a spline; that is why we call the algorithm (3.2) a spline algorithm.

THEOREM 3.1. *For any probability measure, the spline algorithm* (3.2) *is a unique linear algorithm with minimal average error among linear algorithms using N and*

$$e^{\text{avg}}(\varphi^s, N)^2 = \int_{F_1} \|Sf\|^2 \mu(df) - \sum_{i=1}^{n} \|SS_\mu g_i\|^2.$$

Proof. Take a linear algorithm φ of the form (3.1). Note that

$$\begin{aligned} \int_{F_1} (Sf, \varphi(Nf))\, \mu(df) &= \sum_{i=1}^{n} \int_{F_1} (f, g_i)(f, S^* w_i)\, \mu(df) \\ &= \sum_{i=1}^{n} (S_\mu g_i, S^* w_i), \end{aligned}$$

and

$$\begin{aligned} \int_{F_1} \|\varphi(Nf)\|^2 \mu(df) &= \sum_{i,j=1}^{n} (w_i, w_j) \int_{F_1} (f, g_i)(f, g_j)\, \mu(df) \\ &= \sum_{i,j=1}^{n} (w_i, w_j)(S_\mu g_i, g_j) = \sum_{i=1}^{n} \|w_i\|^2. \end{aligned}$$

This yields

$$\begin{aligned} e^{\text{avg}}(\varphi, N)^2 &= \int_{F_1} \|Sf\|^2 \mu(df) - 2 \sum_{i=1}^{n} (SS_\mu g_i, w_i) + \sum_{i=1}^{n} \|w_i\|^2 \\ &= \int_{F_1} \|Sf\|^2 \mu(df) + \sum_{i=1}^{n} \|w_i - SS_\mu g_i\|^2 - \sum_{i=1}^{n} \|SS_\mu g_i\|^2. \end{aligned}$$

Thus, the average error is minimized iff $w_i = SS_\mu g_i$ which corresponds to the spline algorithm. ▌

Remark 3.1. If the mean element m_μ of μ is not necessarily zero then Theorem 3.1 can be generalized as follows. For $N(f) = [(f, g_1),..., (f, g_n)]$ we now assume that $(S_c g_i, g_j) = \delta_{ij}$, where S_c is the correlation operator of μ,

$$(S_c x, y) = \int_{F_1} (f - m_\mu, x)(f - m_\mu, y)\, \mu(df), \qquad \forall x, y \in F_1.$$

Then the affine spline algorithm

$$\varphi^{\text{as}}(Nf) = \sum_{i=1}^{n} (f - m_\mu, g_i)\, SS_c g_i + Sm_\mu$$

is a unique affine algorithm with minimal average error among affine algorithms using N and

$$e^{\text{avg}}(\varphi^{\text{as}}, N)^2 = \int_{F_1} \|Sf\|^2 \mu(df) - \sum_{i=1}^{n} \|SS_c g_i\|^2.$$

Note that $m_\mu = 0$ implies that $\varphi^{\text{as}} = \varphi^{\text{s}}$ which means that the spline algorithm is then a unique linear algorithm with minimal average error among affine algorithms using N. ▌

We now provide a condition on the measure μ which guarantees that the spline algorithm has minimal average error among *all* algorithms using N. Define the mapping D: $F_1 \to F_1$ as

$$Df = 2 \sum_{i=1}^{n} (f, g_i)\, S_\mu g_i - f. \tag{3.4}$$

The mapping D has two properties

$$Nf = NDf, \qquad D^2 f = f, \qquad \forall f \in F_1. \tag{3.5}$$

We are ready to prove

THEOREM 3.2. *Assume that*

$$\mu(D(B)) = \mu(B) \tag{3.6}$$

for any Borel set B of F_1. Then the spline algorithm (3.2) *is a unique algorithm which has minimal average error among all algorithms using N.*

Proof. Take an arbitrary algorithm φ. Observe that (3.6) and (3.5) imply

$$\int_{F_1} \|Sf - \varphi(Nf)\|^2 \mu(df) = \int_{F_1} \|SDf - \varphi(Nf)\|^2 \mu(df). \tag{3.7}$$

Due to

$$\begin{aligned} Sf - \varphi^s(Nf) &= \tfrac{1}{2}(Sf - SDf) \\ &= \tfrac{1}{2}(Sf - \varphi(Nf)) + \tfrac{1}{2}(\varphi(Nf) - SDf), \end{aligned} \tag{3.8}$$

we have

$$\begin{aligned} \|Sf - \varphi^s(Nf)\|^2 &\leqslant \tfrac{1}{4}(\|Sf - \varphi(Nf)\| + \|\varphi(Nf) - SDf\|)^2 \\ &\leqslant \tfrac{1}{2}(\|Sf - \varphi(Nf)\|^2 + \|SDf - \varphi(Nf)\|^2). \end{aligned} \tag{3.9}$$

From (3.7) and (3.9) we get

$$\begin{aligned} e^{\text{avg}}(\varphi, N)^2 &= \tfrac{1}{2} \int_{F_1} \{\|Sf - \varphi(Nf)\|^2 + \|SDf - \varphi(Nf)\|^2\} \, \mu(df) \\ &\geqslant e^{\text{avg}}(\varphi^s, N)^2. \end{aligned}$$

Thus, φ^s has minimal average error. Uniqueness, up to a set of measure zero, follows from the fact that optimality of φ implies that (3.9) holds with equality (almost everywhere) which, in turn, yields $\varphi = \varphi^s$. ∎

Remark 3.2. If (3.6) is satisfied, then the spline algorithm remains optimal even for a more general error criterion. Indeed, let P: $F_2 \to \mathbb{R}_+$ be convex and symmetric. Here it is enough to assume that F_2 is a linear space. Redefine the average error as

$$e^{\text{avg}}(\varphi, N) = \left\{ \int_{F_1} P(Sf - \varphi(Nf)) \, \mu(df) \right\}^{1/2}.$$

Then (3.8) yields

$$P(Sf - \varphi^s(N, f)) \leqslant \tfrac{1}{2} P(Sf - \varphi(Nf)) + \tfrac{1}{2} P(SDf - \varphi(Nf)).$$

Thus (3.6) implies

$$e^{\text{avg}}(\varphi^s, N) \leqslant e^{\text{avg}}(\varphi, N)$$

as claimed.

This improves Theorem 1 of [3], where it is required that dim $F_1 < +\infty$, the functional P has the form $P(g) = p(\|g\|^2)$ and p is a strictly increasing convex and continuously differentiable function such that

$$\int_{F_1} p'(\|Sf - \varphi^s(Nf)\|^2)\|\varphi^s(Nf) - \varphi(Nf)\|^2\mu(df)$$

is finite for any algorithm φ with finite average error. ∎

For which measures does (3.6) hold? This is the case if

(i) F_1 is finite dimensional and either μ is a weighted Lebesgue measure of the form as in [7] or μ is unitary invariant as assumed in [3],

(ii) F_1 is separable and μ is a Gaussian measure with mean zero, i.e., μ is such that

$$\int_{F_1} e^{i(f,x)}\mu(df) = e^{-(S_\mu x, x)/2}, \qquad \forall x \in F_1, \; i = \sqrt{-1},$$

where S_μ is a covariance operator of μ.

Measures satisfying (3.6) for *all* information operators are characterized in [8]. We emphasize that if (3.6) does not hold then the spline algorithm may be far from being optimal.

4. Optimal Information

We wish to find optimal information

$$N(f) = [(f, g_1), \ldots, (f, g_n)], \qquad (S_\mu g_i, g_j) = \delta_{ij}, \tag{4.1}$$

i.e., elements g_i for which the error of the optimal algorithm using N is minimal. This problem has been studied for the worst case in [5, 6] and for the average case for finite dimensional spaces in [7, 3].

We now consider the infinite dimensional case. The error form of the spline algorithm indicates that we should maximize $\sum_{i=1}^n \|SS_\mu g_i\|^2$. Let

$$K = (SS_\mu^{1/2})^* SS_\mu^{1/2} : F_1 \to F_1. \tag{4.2}$$

Then $K = K^* \geqslant 0$ and K has finite trace. Thus there exists an orthonormal basis $z_1, z_2, \ldots$ such that

$$Kz_i = \lambda_i z_i \tag{4.3}$$

and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$, $\sum_{i=1}^{\infty} \lambda_i < +\infty$. Note that $\eta_i = S_\mu^{1/2} g_i$ are orthonormal and

$$\sum_{i=1}^{n} \|SS_\mu g_i\|^2 = \sum_{i=1}^{n} (K\eta_i, \eta_i). \tag{4.4}$$

It is known, see Theorem 4.1.4 of [2], that (4.4) is maximized for $\eta_i = z_i$. Without loss of generality we can assume that $\lambda_n > 0$. Note that $z_i \in S_\mu^{1/2}(F_1)$ for $i \leqslant n$. Define

$$N_n^*(f) = [(f, S_\mu^{-1/2} z_1), (f, S_\mu^{-1/2} z_2), \ldots, (f, S_\mu^{-1/2} z_n)]. \tag{4.5}$$

The spline algorithm φ^s using N_n^* has now the form

$$\varphi^s(N_n^* f) = \sum_{i=1}^{n} (f, S_\mu^{-1/2} z_i)\, SS_\mu^{+1/2} z_i. \tag{4.6}$$

This and Theorems 3.1, 3.2 yield

THEOREM 4.1. (i) *For any measure* μ, *information* N_n^* *is linearly optimal, i.e.,*

$$e^{\text{avg}}(\varphi^s, N_n^*) = \inf\{e^{\text{avg}}(\varphi, N) \colon N \text{ of the form (4.1), } \varphi \text{ linear}\}$$

(ii) *For* μ *satisfying* (3.6) *for all* N, *information* N_n^* *is optimal, i.e.,*

$$e^{\text{avg}}(\varphi^s, N_n^*) = \inf\{e^{\text{avg}}(\varphi, N) \colon N \text{ of the form (4.1), } \varphi \text{ arbitrary}\}. \quad \blacksquare$$

It is easy to show that

$$e^{\text{avg}}(\varphi^s, N_n^*) = \sqrt{\int_{F_1} \|Sf\|^2 \mu(df) - \sum_{i=1}^{n} \lambda_i} = \sqrt{\sum_{i=n+1}^{\infty} \lambda_i}.$$

Thus the average error of the spline algorithm that uses optimal information goes to zero with speed characterized by the truncated series of eigenvalues of K.

ACKNOWLEDGMENTS

We wish to express our gratitude to J. F. Traub for his valuable suggestions and comments. We also thank K. Sikorski for a careful checking of the manuscript.

REFERENCES

1. H.-H. KUO, "Gaussian Measures in Banach Spaces," Lecture Notes in Mathematics 463, Springer-Verlag, Berlin, 1975.
2. M. MARCUS AND H. MINC, "A Survey of Matrix Theory and Matrix Inequalities," Allyn & Bacon, Boston, 1964.
3. C. A. MICCHELLI, Orthogonal projections are optimal algorithms, *J. Approx. Theory* **40** (1984), 101–110.
4. A. V. SKOROHOD, "Integration in Hilbert Space," Springer-Verlag, New York, 1974.
5. J. F. TRAUB AND WOŹNIAKOWSKI, "A General Theory of Optimal Algorithms," Academic Press, New York, 1980.
6. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, "Information, Uncertainty, Complexity," Addison–Wesley, Reading, Mass., 1983.
7. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, "Average Case Optimality for Linear Problems," Dept. of Computer Science Rep., Columbia Univ., 1981; *Theoret. Comput. Sci.* **29** (1984), 1–25.
8. G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI, Can adaption help on the average? *Numer. Math.* 44 (1984), 169–190.
9. G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI, "Average Case Optimal Algorithms in Hilbert Spaces," Dept. of Computer Science Rep., Columbia Univ., 1982.