CUCS-32-82

AVERAGE CASE OPTIMAL ALGORITHMS IN HILBERT SPACES

G.W. Wasilkowski
Department of Computer Science
Columbia University
New York, New York

(On leave from the University of Warsaw)

H. Woźniakowski

Institute of Informatics
University of Warsaw
Warsaw, Poland

and

Department of Computer Science
Columbia University
New York, New York

September 1982

ABSTRACT

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, an affine spline algorithm is optimal among <u>affine</u> algorithms. The affine spline algorithm is defined in terms of the correlation operator and the mean element of the measure. We provide a condition on the measure which guarantees that the affine spline algorithm is optimal among <u>all</u> algorithms. The problem of optimal information is also solved.

1. Introduction

In two recent monographs ([5] and [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 that paper we confined ourselves to linear problems in a finite dimensional space. See also [3] where a general error criterion for optimal approximation of a linear problem is studied.

In this paper we study linear problems in a separable Hilbert space equipped with a probability measure μ.

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

1. For all measures μ:

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

   b. In the class of affine algorithms, an affine spline algorithm, defined in terms of the correlation operator and the mean element 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 affine mapping and B is any Borel set. In the class of all algorithms, the affine spline algorithm is optimal.

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

The measures which satisfy (1.1) include Gaussian measures and the measures studied in [7]. In a forthcoming paper [8] we characterize measures satisfying (1.1).

We briefly summarize the contents of this paper. We formulate the problem in Section 2. In Section 3 we collect some facts on measures in Hilbert spaces. Section 4 deals with optimal algorithms. In subsection (i) we study linear algorithms, in subsection (ii) affine algorithms and in subsection (iii) general algorithms. Section 5 deals with optimal information. Our analysis and results are illustrated by two examples. The first is a finite dimensional Hilbert space equipped with a weighted Lebesgue measure; the second is a separable Hilbert space equipped with a Gaussian measure.

2. Formulation of the Problem.

Let F_1 and F_2 be real separable Hilbert spaces. Let $m = \dim(F_1)$, $m \leq +\infty$, be the dimension of F_1. Let

$$(2.1) \qquad S : F_1 \to F_2$$

be a continuous _linear_ operator. We call S a _solution operator_.

Our aim is to approximate $S(f)$, $\forall f \in F_1$, with an average error as small as possible. In order to define an average error we assume that the Hilbert space F_1 is equipped with a probability measure μ, $\mu(F_1) = 1$, which is defined on Borel sets of F_1. see e.g., [1] and [4].

To find an approximation to Sf we must know something about f. We assume that $N(f)$ is known where

$$(2.2) \qquad N : F_1 \to \mathbb{R}^n$$

is a continuous _linear_ operator whose range has dimension n. We call N an _information operator_ and $n = \text{card}(N)$ is called the _cardinality_ of N. We seek an approximation to Sf by $\varphi(N(f))$ where

$$(2.3) \qquad \varphi : N(F_1) \to F_2.$$

We call φ an _(idealized) algorithm using information N_.

The (global) average error of φ is defined as

$$(2.4) \qquad e^{avg}(\varphi,N) = \{ \int_{F_1} \|Sf - \varphi(N(f))\|^2 \mu(df) \}^{1/2}.$$

Note that the norm in (2.4) is the norm of the Hilbert space F_2.

Let $\Phi(N)$ be the class of all algorithms φ using N for which the average error is well defined, i.e., $\|Sf - \varphi(N(f))\|^2$ is a measurable function.

We wish to find an algorithm φ^* from $\Phi(N)$ with the smallest average error. Such an algorithm is called an optimal average error algorithm, its error is called the average radius of information and is denoted by $r^{avg}(N)$, i.e.,

$$(2.5) \qquad \inf_{\varphi \in \Phi(N)} e^{avg}(\varphi,N) = e^{avg}(\varphi^*,N) = r^{avg}(N).$$

3. Measure μ.

We collect some facts on measures in Hilbert spaces which will be used in the following sections. See e.g., [1] and [4].

(i) The <u>mean element</u> m_μ of μ is defined as

$$(3.1) \qquad (m_\mu, x) = \int_{F_1} (f,x)\mu(df), \quad \forall\, x \in F_1$$

where the integral in (3.1) is understood as the Lebesgue integral with respect to the measure μ.

(ii) The <u>correlation operator</u> S_c of μ is defined as

$$(3.2) \qquad (S_c x, y) = \int_{F_1} (f - m_\mu, x)(f - m_\mu, y)\mu(df), \quad \forall\, x,y \in F_1.$$

Throughout this paper we assume that

$$(3.3) \qquad \int_{F_1} \|f\|^2 \mu(df) < +\infty.$$

This guarantees the existence of m_μ (since $\int_{F_1} \|f\|\mu(df) \le \sqrt{\int_{F_1} \|f\|^2 \mu(df)} < +\infty$) and S_c.

The correlation operator is self-adjoint and nonnegative definite. Let $\zeta_1, \zeta_2, \ldots$ be an orthonormal basis of F_1 such that

$$(3.4) \qquad S_c \zeta_i = \lambda_i \zeta_i, \quad 0 \le \lambda_{i+1} \le \lambda_i, \quad \forall\, i.$$

Then $f - m_\mu = \Sigma_{i=1}^{m}(f-m_\mu, \zeta_i)\zeta_i$ and $\|f-m_\mu\|^2 = \Sigma_{i=1}^{m}(f-m_\mu, \zeta_i)^2$. Note that (3.2) implies that

(3.5) $$\lambda_i = \int_{F_1} (f - m_\mu, \zeta_i)^2 \mu(df),$$

i.e., λ_i is the average value of the squared ith component of $f - m_\mu$. Without loss of generality we can assume that $\int_{F_1}(f-m_\mu, \zeta_i)^2\mu(df)$ is positive for any i. This means that all λ_i are positive and S_c is a one-to-one mapping.

From (3.5) and (3.3) we conclude

(3.6) $$\text{trace}(S_c) = \Sigma_{i=1}^{m}\lambda_i = \int_{F_1} \|f-m_\mu\|^2\mu(df) < +\infty,$$

i.e., the trace of S_c is finite. Note that $m = +\infty$ implies $\lambda_i \to 0$. This yields that $S_c(F_1)$ is a proper subset of F_1.

(iii) The <u>covariance operator</u> S_μ of μ is defined as

(3.7) $$(S_\mu x, y) = \int_{F_1}(f,x)(f,y)\mu(df), \ \forall x,y \in F_1.$$

Due to (3.3), S_μ exists and can be expressed in terms of S_c and m_μ as

(3.8) $$S_\mu x = S_c x + (m_\mu, x)m_\mu, \ \forall x \in F_1.$$

Note that $(S_\mu x,x) = (S_c x,x) + (m_\mu,x)^2$ which yields that $S_\mu \geq S_c$ and $\text{trace}(S_\mu) = \text{trace}(S_\mu) + \|m_\mu\|^2$. Thus S_μ is self-adjoint, positive definite and has finite trace. If $m_\mu = 0$

then $S_\mu = S_c$.

We illustrate these concepts by two examples.

Example 1: Assume that F_1 is finite dimensional, $F_1 = \mathbb{R}^m$ with $m < +\infty$. As in [7] define

$$(3.9)\qquad \mu(B) = \int_B w(\|Tf\|)df$$

where B is a Borel set of $\mathbb{R}^m$ and the integral is understood as the usual Lebesgue integral in $\mathbb{R}^m$, $w : \mathbb{R}_+ \to \mathbb{R}_+$ is a function such that $\int_{\mathbb{R}^m} w(\|Tf\|)df = 1$ and $T : \mathbb{R}^m \to F_4 = T(\mathbb{R}^m)$ is a one-to-one linear operator onto a Hilbert space F_4. Assume that $d = \frac{1}{m} \int_{\mathbb{R}^m}\|Tf\|^2 \mu(df) = \frac{1}{m}\int_{\mathbb{R}^m}\|Tf\|^2 w(\|Tf\|)df$ is finite. Let $M = (T^*T)^{1/2}$ and $M\zeta_i = \lambda_i\zeta_i$ for orthonormal ζ_i and $\lambda_i > 0$. Then

$$(S_\mu\zeta_i\cdot\zeta_j) = \int_{\mathbb{R}^m}(f,\zeta_i)(f,\zeta_j)w(\|Mf\|)df$$

$$= \frac{\det M^{-1}}{\lambda_i\lambda_j}\int_{\mathbb{R}^m}(f,\zeta_i)(f,\zeta_j)w(\|f\|)df.$$

Since the last integrand is odd for $i \neq j$, we have

$$(S_\mu\zeta_i,\zeta_j) = 0 \qquad \text{for } i \neq j,$$

$$(S_\mu\zeta_i,\zeta_j) = \frac{\det M^{-1}}{\lambda_i^2}\int_{\mathbb{R}^m}(f,\zeta_i)^2 w(\|f\|)df$$

$$= \frac{\det M^{-1}}{m\,\lambda_i^2}\int_{\mathbb{R}^m}\|f\|^2 w(\|f\|)dt$$

$$= \frac{1}{m\lambda_i^2}\int_{\mathbb{R}^m}\|Tf\|^2 w(\|Tf\|)\mu(df) = \frac{d}{\lambda_i^2}.$$

Thus

$$S_\mu \zeta_i = \frac{d}{\lambda_i^2} \zeta_i = dM^{-2}\zeta_i$$

which yields

$$(3.10) \qquad S_\mu = d(T^*T)^{-1}.$$

We have $m_\mu = 0$ and $S_c = S_\mu$.

<u>Example 2</u>: A <u>Gaussian</u> measure μ in the Hilbert space F_1 is a measure such that

$$(3.11) \qquad \int_{F_1} e^{i(f,x)}\mu(df) = e^{i(a,x)-(Ax,x)/2}, \qquad \forall x \in F_1, \quad i = \sqrt{-1},$$

where $a \in F_1$ and $A : F_1 \to F_1$ is a self-adjoint positive definite operator with finite trace. (The left-hand side of (3.11) is called the <u>characteristic functional</u> of μ.) Then

$$(3.12) \qquad S_c = A \quad \text{and} \quad m_\mu = a.$$

See [4, p. 18]. From (3.11) it follows that

$$(3.13) \qquad \mu\{f \in F_1 : (f,x) \le d\} = \int_{-\infty}^{d} \frac{1}{\sqrt{2\pi\sigma_x}} e^{-(t-mx)^2/2\sigma_x} dt, \qquad \forall x \in F_1, \ \forall d \in \mathbb{R},$$

where $\sigma_x = (S_c x, x)$ and $m_x = (m_\mu, x)$. Thus the measurable function $(\cdot, x)$ is normally distributed. See [1, p. 28]. ■

4. Algorithms with Minimal Average Error.

In this section we pose the problem of characterizing algorithms with minimal average error. In three subsections we solve this problem for three classes of algorithms. The first subsection deals with linear algorithms. Linear algorithms are important since they are easy to implement in actual computation. We prove that a linear algorithm with minimal average error is a spline algorithm defined in terms of the covariance operator. In the second subsection we turn to affine algorithms. They are also easy to implement. We prove that an affine algorithm with minimal average error is an affine spline algorithm which is defined in terms of the correlation operator and the mean element. In the third subsection we deal with the class $\Phi(N)$ of idealized algorithms. We find a property of the measure μ such that the affine spline algorithm is an optimal average error algorithm. We assume that

$$(4.1) \qquad N(f) = [(f,g_1),(f,g_2),\ldots,(f,g_n)]$$

for linearly independent elements $g_1,g_2,\ldots,g_n$ of F_1.

(i) <u>Linear Algorithms</u>

Let S_μ be the covariance operator of μ,

$S_\mu : F_1 \to S_\mu(F_1)$. Define

$$(4.2) \qquad T_\mu = S_\mu^{-1/2} : G_\mu \to F_1$$

where $G_\mu = S_\mu^{1/2}(F_1)$. Note that T_μ is a self-adjoint, positive definite and one-to-one operator. If $m = \dim(F_1) = +\infty$ then T_μ is unbounded.

Let $\eta_1, \eta_2, \ldots, \eta_n$ be an orthnormal basis of the linear subspace $\mathrm{lin}(S_\mu^{1/2} g_1, S_\mu^{1/2} g_2, \ldots, S_\mu^{1/2} g_n)$. Then

$$(4.2) \qquad \eta_i = \Sigma_{j=1}^n c_{ij} S_\mu^{1/2} g_j, \quad (\eta_i, \eta_j) = \delta_{i,j},$$

$$i, j = 1, 2, \ldots, n$$

for some c_{ij}. Let

$$(4.3) \qquad C_\mu = (c_{ji})_{j,i=1}^n$$

be a $n \times n$ matrix. Note that C_μ is nonsingular.

The element η_i belongs to the domain G_μ and therefore $T_\mu \eta_i$ is well defined. From (4.2) and (4.3) we have

$$(4.4) \qquad [(f, T_\mu \eta_1), (f, T_\mu \eta_2), \ldots, (f, T_\mu \eta_n)] = N(f) C_\mu .$$

Thus, knowing $N(f)$ we can compute $(f. T_\mu \eta_i)$.

Define the element $\sigma = \sigma(N(f))$ as

$$(4.5) \qquad \sigma = \Sigma_{i=1}^n (f, T_\mu \eta_i) T_\mu^{-1} \eta_i .$$

Then $(\sigma, T_\mu \eta_i) = (f, T_\mu \eta_i)$ which yields $N(\sigma) = N(f)$. This means that σ "interpolates" f with respect to N.

Observe that σ does not depend on a particular choice of the orthonormal basis $\eta_1, \ldots, \eta_n$. Indeed, if $\beta_1, \ldots, \beta_n$ is also an orthonormal basis of $\mathrm{lin}(S_\mu^{1/2} g_1, \ldots, S_\mu^{1/2} g_n)$ then $\eta_i = \Sigma_{j=1}^n (\eta_i, \beta_j)\beta_j$ and

$$\sigma = \Sigma_{i=1}^n \Sigma_{j=1}^n (\eta_i, \beta_j)(f, T_\mu \beta_j) \Sigma_{k=1}^n (\eta_i, \beta_k) T_\mu^{-1} \beta_k$$

$$= \Sigma_{j=1}^n (f, T_\mu \beta_j) [\Sigma_{k=1}^n (\Sigma_{i=1}^n (\eta_i . \beta_j)(\eta_i, \beta_k)) T_\mu^{-1} \beta_k]$$

$$= \Sigma_{j=1}^n (f, T_\mu \beta_j) \Sigma_{k=1}^n (\beta_j, \beta_k) T_\mu^{-1} \beta_k = \Sigma_{j=1}^n (f, T_\mu \beta_j) T_\mu^{-1} \beta_j$$

as claimed.

Take an arbitrary $g \in G_\mu$ such that $N(g) = N(f)$. Then $\|T_\mu g\|^2 = \|T_\mu (g-\sigma)\|^2 + \|T_\mu \sigma\|^2 + 2(T_\mu (g-\sigma), T_\mu \sigma)$. Since $h = g - \sigma$ belongs to ker N then $(h, T_\mu \eta_i) = 0$ and $(T_\mu h, T_\mu \sigma) = (h, T_\mu^2 \sigma) = \Sigma_{i=1}^n (f, T_\mu \eta_i)(h, T_\mu \eta_i) = 0$. Thus $\|T_\mu g\| \geq \|T_\mu \sigma\|$ and $\|T_\mu g\| = \|T_\mu \sigma\|$ iff $g = \sigma$. Thus σ is the unique solution of

$$N(\sigma) = N(f),$$

(4.6)

$$\|T_\mu \sigma\| = \inf\{\|T_\mu g\| : g \in G_\mu, N(g) = N(f)\}.$$

The solution of (4.6) is called a <u>spline</u> (or a T_μ-spline) interpolating N(f). The algorithm

$$(4.7) \qquad \varphi^s(N(f)) = S\sigma(N(f)) = \Sigma_{i=1}^n (f, T_\mu \eta_i) ST_\mu^{-1} \eta_i$$

is called a <u>spline</u> algorithm. (A discussion and optimal properties of spline algorithms for the worst case may be found in [5].) From (4.4) we can express (4.7) in the equivalent form

$$\varphi^s(N(f)) = \Sigma_{i=1}^n (f, g_i) \Sigma_{j=1}^n c_{ji} ST_\mu^{-1} \eta_j.$$

An algorithm φ is <u>linear</u> if it has the form

$$\varphi(N(f)) = \Sigma_{i=1}^n (f, g_i) w_i$$

for some elements $w_1, w_2, \ldots, w_n$ of F_2. We are ready to prove that the spline algorithm φ^s has minimal average error among all algorithms that use N.

<u>Theorem 4.1</u>: The spline algorithm φ^s defined by (4.7) and (4.4) is a unique linear algorithm with minimal average error among linear algorithms using N and

$$(4.8) \qquad e^{avg}(\varphi^s, N) = \sqrt{\Sigma_{i=n+1}^m \|ST_\mu^{-1} \eta_i\|^2}.$$

where $\eta_1, \ldots, \eta_n, \eta_{n+1}, \ldots$ form an orthonormal basis of F_1. ∎

<u>Proof</u>: Let $\varphi(N(f)) = \Sigma_{i=1}^n (f, g_i) w_i$. Due to (4.4) it can be written as $\varphi(N(f)) = \Sigma_{i=1}^n (f, T_\mu \eta_i) z_i$ for some elements $z_1, \ldots, z_n$ of F_2. Then

$$e^{avg}(\varphi,N)^2 = \int_{F_1} \|Sf-\varphi(N(f)\|^2\mu(df) = \int_{F_1}\|Sf\|^2\mu(df) +$$
$$+ \int_{F_1}\|\varphi(N(f))\|^2\,\mu(df) - 2\int_{F_1}(Sf,\varphi(N(f)))\mu(df).$$ Note that $(Sf,\varphi(N(f))) = (f,\Sigma_{i=1}^n(f,T_\mu\eta_i)S^*z_i) = \Sigma_{i=1}^n(f,T_\mu\eta_i)(f,S^*z_i)$.

From (3.7) we have

$$\int_{F_1}(f,T_\mu\eta_i)(f,S^*z_i)\mu(df) = (S_\mu T_\mu\eta_i,S^*z_i)$$
$$= (ST_\mu^{-1}\eta_i,z_i).$$

Observe next that $\|\varphi(N(f))\|^2 = \Sigma_{i,j=1}^n(f,T_\mu\eta_i)(f,T_\mu\eta_j)(z_i,z_j)$ and

$$\int_{F_1}\|\varphi(N(f)\|^2\mu(df) = \Sigma_{i,j=1}^n(S_\mu T_\mu\eta_i,T_\mu\eta_j)(z_i,z_j)$$
$$= \Sigma_{i=1}^n\|z_i\|^2$$

since $(S_\mu T_\mu\eta_i,T_\mu\eta_j) = (\eta_i,\eta_j) = \delta_{ij}$. Thus

$$e^{avg}(\varphi,N)^2 = \int_{F_1}\|Sf\|^2\mu(df)$$
$$+ \Sigma_{i=1}^n\{\|z_i\|^2 - 2(ST_\mu^{-1}\eta_i,z_i)\}$$
$$= \int_{F_1}\|Sf\|^2\mu(df) + \Sigma_{i=1}^n\|z_i-ST_\mu^{-1}\eta_i\|^2$$
$$- \Sigma_{i=1}^n\|ST_\mu^{-1}\eta_i\|^2$$
$$= e^{avg}(\varphi^s,N)^2 + \Sigma_{i=1}^n\|z_i - ST_\mu^{-1}\eta_i\|^2.$$

This proves that the average error is minimized iff $z_i = ST_\mu^{-1}\eta_i$, i.e., φ^s is a unique linear algorithm with minimal average error. The error of φ^s is given by

$$(4.9)\qquad e^{avg}(\varphi^s,N)^2 = \int_{F_1} \|Sf\|^2\mu(df) - \Sigma_{i=1}^{n}\|ST_\mu^{-1}\eta_i\|^2.$$

We now compute $\int_{F_1}\|Sf\|^2\mu(df)$. Since $f = \Sigma_{i=1}^{m}(f,\xi_i)\xi_i$ where ξ_i are orthonormal eigenelements of S_μ, $S_\mu\xi_i = \lambda_i\xi_i$, we have

$$\int_{F_1}\|Sf\|^2\mu(df) = \int_{F_1}\Sigma_{i,j=1}^{m}(f,\xi_i)(f,\xi_j)(S\xi_i,S\xi_j)\mu(df)$$

$$= \Sigma_{i=1}^{m}\lambda_i\|S\xi_i\|^2 = \Sigma_{i=1}^{m}\|ST_\mu^{-1}\xi_i\|^2.$$

Let $K = [(ST_\mu^{-1})^*ST_\mu^{-1}]^{1/2}$. Then $\|ST_\mu^{-1}x\| = \|Kx\|$, $\forall\, x \in F_1$, where the first norm is in F_2 and the second one in F_1. Observe that $K\xi_i = \Sigma_{j=1}^{m}(K\xi_i,\eta_j)\eta_j$ and $\|K\xi_i\|^2 = \Sigma_{j=1}^{m}(K\xi_i\eta_j)^2$. Thus

$$\int_{F_1}\|Sf\|^2\mu(df) = \Sigma_{i=1}^{m}\|K\xi_i\|^2 = \Sigma_{j=1}^{m}\Sigma_{i=1}^{m}(\xi_i,K\eta_j)^2$$

$$= \Sigma_{j=1}^{m}\|K\eta_j\|^2.$$

This and (4.9) complete the proof. ∎

It is well known that spline algorithms enjoy many important optimiality properties for the worst case. Theorem 4.1 states that among linear algorithms spline algorithms are also optimal for the average case.

(ii) <u>Affine Algorithms</u>

Let S_c be the correlation operator of μ. $S_c : F_1 \to S_c(F_1)$.

Define

$$(4.10)\qquad T_c = S_c^{-1/2} : G_c \to F_1,$$

where $G_c = S_c^{1/2}(F_1)$. Let $\eta_1, \eta_2, \ldots, \eta_n$ be an orthonormal basis of the linear subspace $\text{lin}(S_c^{1/2} g_1, \ldots, S_c^{1/2} g_n)$. Then there exists a $n \times n$ nonsingular matrix $C_c = (c_{ji})$ such that

$$(4.11)\qquad [(f, T_c\eta_1), (f, T_c\eta_2), \ldots, (f, T_c\eta_n)] = N(f)C_c.$$

Define the element $\sigma = \sigma(N(f))$ as

$$(4.12)\qquad \sigma = \Sigma_{i=1}^{n}(f - m_\mu, T_c\eta_i)T_c^{-1}\eta_i + m_\mu$$

where m_μ is the mean element of μ. Then $(\sigma, T_c\eta_i) =$ $= (f - m_\mu, T_c\eta_i) + (m_\mu, T_c\eta_i) = (f, T_c\eta_i)$ which yields $N(\sigma) = N(f)$. Thus σ interpolates f with respect to N.

The algorithm

$$(4.13)\qquad \varphi^{as}(N(f)) = S\sigma(N(f)) = \Sigma_{i=1}^{n}(f - m_\mu, T_c\eta_i)ST_c^{-1}\eta_i + Sm_\mu$$

is called an affine spline algorithm. Note that φ^{as} differs from a linear algorithm by the constant element $w = Sm_\mu - \Sigma_{i=1}^{n}(m_\mu, T_c\eta_i)ST_c^{-1}\eta_i$. Due to (4.11) one can equivalently rewrite (4.13) as

$$\varphi^{as}(N(f)) = \Sigma_{i=1}^{n}(f, g_i)\{\Sigma_{j=1}^{n}c_{ji}ST_c^{-1}\eta_j\} + w.$$

This means that φ^{as} is affine in N(f).

It is easy to check that the element $\sigma - m_\mu$ is a unique solution of the problem

$$(4.14)\quad \begin{aligned} &N(\sigma-m_\mu) = N(f-m_\mu), \\ &\|T_c(\sigma-m_\mu)\| = \inf\{\|T_c g\| : g\in G_c, N(g) = N(f-m_\mu)\}. \end{aligned}$$

Thus $\sigma - m_\mu$ is a T_c-spline. Observe that

$$\varphi^{as}(N(f)) = S(\sigma-m_\mu) + Sm_\mu$$

where $S(\sigma-m_\mu)$ is a spline algorithm. Thus the algorithm φ^{as} is a spline algorithm translated by Sm_μ. This motives our terminology. Note that $m_\mu = 0$ implies $T_\mu = T_c$ and $\varphi^{as}(N(f)) = \varphi^{s}(N(f))$, i.e., (4.13) coincides with (4.7).

We are now ready to prove that the affine spline algorithm has minimal average error among all affine algorithms using N, i.e., among algorithms of the form

$$\varphi(N(f)) = \Sigma_{i=1}^{n}(f,g_i)w_i + w_0$$

for some elements $w_0, w_1, \ldots, w_n$ of F_2.

<u>Theorem 4.2</u>: The affine spline algorithm φ^{as} defined by (4.13) and (4.11) is a unique affine algorithm with minimal average error among affine algorithms using N and

$$(4.16)\qquad e^{avg}(\varphi^{as},N) = \sqrt{\Sigma_{i=n+1}^{m}\|ST_c^{-1}\eta_i\|^2}$$

where $\eta_1,\eta_2,\ldots \eta_n,\eta_{n+1},\ldots$ form an orthonormal basis of F_1.

<u>Proof</u>: Let $\varphi(N(f)) = \Sigma_{i=1}^{n}(f,g_i)w_i + w_0$. Due to (4.11) it can be written as $\varphi(N(f)) = \Sigma_{i=1}^{n}(f-m_\mu, T_c\eta_i)\, z_i + z_0$ for some elements $z_0,\ldots,z_n$ of F_2. Then $e^{avg}(\varphi,N)^2 =$ $\int_{F_1}\|Sf-\varphi(N(f))\|^2\mu(df) = \int_{F_1}\|S(f-m_\mu)-\varphi(N(f))+Sm_\mu\|^2\mu(df) =$ $\int_{F_1}\|S(f-m_\mu)\|^2\mu(df)+\int_{F_1}\|\varphi(N(f))-Sm_\mu\|^2\mu(df)-2\int_{F_1}(S(f-m_\mu),\varphi(N(f)-Sm_\mu)\mu(df)$. Note that $(S(f-m_\mu),\varphi(N(f))-Sm_\mu) =$ $(f-m_\mu,\Sigma_{i=1}^{n}(f-m_\mu,T_c\eta_i)S^*z_i + S^*(z_0-Sm_\mu)) =$ $\Sigma_{i=1}^{n}(f-m_\mu,T_c\eta_i)(f-m_\mu,S^*z_i)+(f-m_\mu,S^*(z_0-Sm))$. From (3.2) and (3.1) we have

$$\int_{F_1}(f-m_\mu,T_c\eta_i)(f-m_\mu,S^*z_i)\mu(df) = (S_cT_c\eta_i,S^*z_i)$$

$$= (ST_c^{-1}\eta_i,z_i),$$

$$\int_{F_1}(f-m_\mu,S^*(z_0-Sm_\mu))\mu(df) = (m_\mu,S^*(z_0-Sm_\mu))$$

$$-\ (m_\mu,S^*(z_0-Sm_\mu)) = 0.$$

Observe next that $\|\varphi(N(f)) - Sm_\mu\|^2 =$ $\Sigma_{i,j=1}^{n}(f-m_\mu,T_c\eta_i)(f-m_\mu,T_c\eta_j)(z_i,z_j) +$ $2\Sigma_{i=1}^{n}(f-m_\mu,T_c\eta_i)(z_i,z_0-Sm_\mu) + \|z_0-Sm_\mu\|^2$. Thus

$$\int_{F_1} \|\varphi(N(f))-Sm_\mu\|^2\mu(df) = \Sigma_{i,j=1}^n (S_cT_c\eta_j, T_c\eta_j)(z_i, z_j)$$

$$+ \|z_0-Sm_\mu\|^2 = \Sigma_{i=1}^n\|z_i\|^2 + \|z_0-Sm_\mu\|^2.$$

From this we have

$$e^{avg}(\varphi,N)^2 = \int_{F_1} \|S(f-m_\mu)\|^2\mu(df)$$

$$+ \Sigma_{i=1}^n\{\|z_i\|^2 - 2(ST_c^{-1}\eta_i,z_i)\} + \|z_0-Sm_\mu\|^2$$

$$= \int_{F_1} \|S(f-m_\mu)\|^2\mu(df) + \Sigma_{i=1}^n\|z_i-ST_c^{-1}\eta_i\|^2$$

$$+ \|z_0-Sm_\mu\|^2 - \Sigma_{i=1}^n\|ST_c^{-1}\eta_i\|^2.$$

This proves that the average error is minimized iff $z_i = ST_c^{-1}\eta_i$ and $z_0 = Sm_\mu$, i.e., φ^{as} is a unique affine algorithm with minimal average error. The error of φ^{as} is given by

$$e^{avg}(\varphi^{as},N)^2 = \int_{F_1} \|S(f-m_\mu)\|^2\mu(df) - \Sigma_{i=1}^n\|ST_c^{-1}\eta_i\|^2.$$

Repeating the last part of the proof of Theorem 4.1 one can show that

$$\int_{F_1} \|S(f-m_\mu)\|^2\mu(df) = \Sigma_{i=1}^m\|ST_c^{-1}\eta_i\|^2.$$

This completes the proof.

Theorem 4.2 establishes optimality of the affine spline algorithm in the class of affine algorithms. Note that the error of φ^{as} has a form similar to the error of the spline algorithm φ^{s}; although φ^{as} depends on T_c whereas φ^{s} depends on T_{μ}.

If $m_{\mu} = 0$ then the affine spline algorithm coincides with the spline algorithm. This yields the following corollary.

Corollary 4.1: If $m_{\mu} = 0$ then the spline algorithm is a unique linear algorithm with minimal average error among affine algorithms using N.

(iii) Optimal Average Error Algorithms

In this subsection we provide a condition on the measure μ which guarantees that the affine spline algorithm is an optimal average error algorithm. This condition is expressed in terms of the mapping $D : F_1 \to F_1$ defined as

$$(4.16) \qquad Df = 2(\Sigma_{i=1}^{n}(f-m_{\mu},T_c\eta_i)T_c^{-1}\eta_i + m_{\mu}) - f.$$

Here we use the notation of subsection (ii). The mapping D is affine and has two important properties

$$(4.17) \qquad \begin{aligned} N(f) &= N(Df), \\ D^2 f &= f. \end{aligned}$$

Indeed, $(Df,T_c\eta_i) = 2((f-m_{\mu},T_c\eta_i) + (m_{\mu},T_c\eta_i)) - (f,T_c\eta_i) = (f,T_c\eta_i)$. This and (4.11) yield $N(f) = N(Df)$. Then (4.16) can be rewritten as

$$f = 2(\Sigma_{i=1}^{n}(Df-m_\mu, T_c\eta_i)T_c^{-1}\eta_i + m_\mu) - Df = D(Df).$$

Thus D^2 is the identity operator, hence $D = D^{-1}$. Note that the mapping D depends on the information operator N.

We are ready to prove

Theorem 4.3: Assume that

$$(4.18) \qquad \mu(D(B)) = \mu(B)$$

for any Borel set B of F_1. Then the affine spline algorithm φ^{as} defined by (4.13) is a unique optimal average error algorithm, i.e.,

$$r^{avg}(N) = e^{avg}(\varphi^{as}, N) = \sqrt{\Sigma_{i=n+1}^{m}\|ST_c^{-1}\eta_i\|^2}.$$

Proof: Take an arbitrary algorithm φ from $\Phi(N)$. Observe that (4.18) implies

$$\int_{F_1} \|Sf-\varphi(N(f))\|^2\mu(df) = \int_{F_1} \|SDf-\varphi(N(Df))\|^2\mu(df).$$

Since $N(Df) = N(f)$, we can express the average error of φ as

$$(4.19) \qquad e^{avg}(\varphi, N)^2 = \frac{1}{2}\int_{F_1}\{\|Sf-\varphi(N(f))\|^2 + \|SDf-\varphi(N(f))\|^2\}\mu(df).$$

Observe that

$$(4.20)\qquad \|Sf-\varphi^{as}(N(f))\|^2 = \frac{1}{4}\|Sf-SDf\|^2$$

$$\leq \frac{1}{4}(\|Sf-\varphi(N(f))\| + \|SDf-\varphi(N(f))\|)^2$$

$$\leq \frac{1}{2}(\|Sf-\varphi(N(f))\|^2 + \|SDf-\varphi(N(f))\|^2).$$

From (4.19) we get

$$e^{avg}(\varphi,N)^2 \geq \int_{F_1} \|Sf-\varphi^{as}(N(f))\|^2 \mu(df)$$

$$= e^{avg}(\varphi^{as},N)^2.$$

Thus φ^{as} is an optimal average error algorithm. We now prove uniqueness. (Of course, uniqueness is understood to be up to a set of measure zero.) If φ is an optimal average error algorithm then (4.20) holds with equality almost everywhere. This yields

$$\|Sf-SDf\|^2 = 2(\|Sf-\varphi(N(f))\|^2+\|SDf-\varphi(N(f))\|^2.$$

Since

$$\|Sf-SDf\|^2 = \|Sf-\varphi(N(f))\|^2$$

$$+ \|SDf-\varphi(N(f))\|^2-2(Sf-\varphi(N(f)),SDf-\varphi(N(f)))$$

then

$$0 = \|Sf-\varphi(N(f))\|^2 + \|SDf-\varphi(N(f))\|^2$$

$$+ 2(Sf-\varphi(N(f)),SDf-\varphi(N(f)))$$

$$= \|Sf-\varphi(N(f)) + SDf-\varphi(N(f))\|^2.$$

Hence

$$\varphi(N(f)) = \frac{1}{2}(Sf+SDf) = \varphi^{as}(N(f)).$$

The average error of φ^{as}, which is equal to the average radius, is given by Theorem 4.2. Hence Theorem 4.3 is proven. ∎

Theorem 4.3 states that the invariance of the measure μ under the mapping D yields optimality of the algorithm φ^{as}. We now show that (4.18) holds for two examples which we presented in section 3. Measures satisfying (4.18) are characterized in [8].

<u>Example 1</u> (continued from Section 3): Since $m_\mu = 0$ and $T_c = T_\mu = d^{-1/2}(T^*T)^{1/2}$, we have

$$\mu(D(B)) = \int_{D(B)} w(d^{1/2}\|T_\mu f\|)df.$$

Let $g = Df$. Note that

$$\|T_\mu D^{-1}g\| = \|2\Sigma_{i=1}^{n}(g,T_\mu\eta_i)\eta_i - T_\mu g\| = \|T_\mu g\|.$$

Since D is linear and $|\det D| = 1$, we have

$$\mu(D(B)) = \int_B w(d^{1/2}\|T_\mu g\|)dg = \mu(B).$$

Thus (4.18) holds for every information operator N. ∎

<u>Example 2</u> (continued from Section 3): It is enough to show

that (4.18) holds for the sets B of the form $B = \{f \in F_1 : (f,x) \leq d\}$ where $x \in F_1$ and $d \in \mathbb{R}$. From (3.13) we have

$$\mu(B) = \frac{1}{\sqrt{2\pi\sigma_x}} \int_{-\infty}^{d} e^{-(t-m_x)^2/2\sigma_x}\, dt$$

where $\sigma_x = (S_c x, x)$ and $m_x = (m_\mu, x)$. We find $\mu(D(B))$. Note that

$$(Df,x) = (f,D_1x) + (b,x)$$

where $D_1x = 2\Sigma_{i=1}^{n}(T_c^{-1}\eta_i,x)T_c\eta_i - x$ and $b = 2(m_\mu - \Sigma_{i=1}^{n}(m_\mu,T_c\eta_i)T_c^{-1}\eta_i)$. From this we conclude that

$$D(B) = \{f \in F_1 : (f,D_1x) \leq d - (b,x)\}.$$

Observe that

$$(m_\mu,D_1x) = (Dm_\mu,x) - (b,x) = (m_\mu,x) - (b,x),$$

$$(S_cD_1x,D_1x) = \|2\Sigma_{i=1}^{n}(T_c^{-1}\eta_i,x)\eta_i - T_c^{-1}x\|^2$$

$$= \|T_c^{-1}x\|^2 = (S_cx.x).$$

Thus

$$\mu(D(B)) = \frac{1}{\sqrt{2\pi\sigma_x}} \int_{-\infty}^{d-(b,x)} e^{-(t-m_x+(b,x))^2/2\sigma_x}\, dt$$

$$= \frac{1}{\sqrt{2\pi\sigma x}} \int_{-\infty}^{d} e^{-(t-m_x)^2/2\sigma x}\, dt = \mu(B).$$

Hence (4.18) holds for every information operator N. ∎

We now show an example of a measure μ for which (4.18) does not hold and for which the algorithm (4.13) is far from being optimal.

Example 3: Let $F_1 = F_2 = \text{lin}(\zeta_1, \zeta_2, \ldots)$ where $(\zeta_i, \zeta_j) = \delta_{ij}$. Define

$$\mu(\{\zeta_i\}) = \mu(\{-\zeta_i\}) = p_i$$

where $\Sigma_{i=1}^{\infty} p_i = \frac{1}{2}$ for different positive p_i. Thus μ is an atomic measure concentrated on the elements $\zeta_1, -\zeta_1 \cdot \zeta_2, -\zeta_2, \ldots$. It is easy to check that $m_\mu = 0$ and

$$S_\mu f = S_c f = \Sigma_{i=1}^{\infty} 2p_i (f, \zeta_i)\zeta_i .$$

Let $N(f) = (f, g_1)$ with $g_1 = \Sigma_{i=1}^{\infty} p_i \zeta_i$ and let $Sf = f$. The algorithm (4.13) (and (4.7)) takes now the form

$$\varphi^{as}(N(f)) = \frac{(f, g_1)}{(S_\mu g_1, g_1)} S_\mu g_1$$

$$= (\Sigma_{i=1}^{\infty} 2p_i^3)^{-1} (f, g_1) \Sigma_{i=1}^{\infty} 2p_i^2 \zeta_i$$

and has positive average error.

Define the algorithm

$$\varphi^*(N(f)) = \begin{cases} \zeta_i & \text{if } N(f) = p_i, \\ -\zeta_i & \text{if } N(f) = -p_i, \\ g & \text{otherwise} \end{cases}$$

where g is an arbitrary element of F_1. The algorithm φ^* is <u>nonlinear</u> in $N(f)$ and discontinuous at zero. Observe that

$$\int_{F_1} \|f-\varphi^*(N(f))\|^2 \mu(df)$$

$$= \Sigma_{i=1}^{\infty} p_i(\|\zeta_i-\varphi^*(p_i)\|^2 + \|-\zeta_i-\varphi^*(-p_i)\|^2) = 0.$$

Thus φ^* is optimal and φ^{as} is <u>not</u>. The mapping D has now the form

$$Df = \frac{2(f,g_1)}{(S_\mu g_1,g_1)} \Sigma_{i=1}^{\infty} 2p_i^2\zeta_i - f.$$

To see that (4.18) is not satisfied, set $B = \{\zeta_i\}$. Then $\mu(B) = p_1$ and $\mu(DB) = 0$. ∎

We end this section by a property of optimal average error algorithms.

<u>Theorem 4.4</u>: An algorithm φ^* is an optimal average error algorithm iff

(4.21) $$\int_{F_1} (Sf-\varphi^*(N(f)),\varphi(N(f)))\mu(df) = 0, \quad \forall \varphi \in \Phi(N).$$ ∎

<u>Proof</u>: Assume that φ^* is an optimal average error algorithm

and let φ be an arbitrary algorithm using N. Define the algorithm

$$\varphi_1(N(f)) = \varphi^*(N(f)) + c\varphi(N(f))$$

for some real c. We have

$$(4.22)\qquad e^{avg}(\varphi_1,N)^2 = e^{avg}(\varphi^*,N)^2 - 2c\int_{F_1}(Sf-\varphi^*(N(f)),\varphi(N(f)))\mu(df) + c^2\int_{F_1}\|\varphi(N(f))\|^2\mu(df).$$

Since $e^{avg}(\varphi_1,N) \geq e^{avg}(\varphi^*,N)$ for an arbitrary c, then the coefficient multiplying c in (4.22) must vanish. This yields (4.21).

Assume now that (4.21) holds. Take an algorithm φ and let $\varphi_1(N(f)) = \varphi(N(f)) - \varphi^*(N(f))$. Then

$$e^{avg}(\varphi,N)^2 = \int_{F_1}\|Sf-\varphi(N(f))\|^2\mu(df)$$

$$= \int_{F_1}\|Sf-\varphi^*(N(f))\|^2\mu(df) - 2\int_{F_1}(Sf-\varphi^*(N(f)),\varphi_1(N(f)))\mu(df) + \int_{F_1}\|\varphi_1(N(f))\|^2\mu(df) = e^{avg}(\varphi^*,N)^2 + \int_{F_1}\|\varphi_1(N(f))\|^2\mu(df) \geq e^{avg}(\varphi^*,N)^2.$$

This proves that φ^* is an optimal average error algorithm which completes the proof. ∎

Observe that Theorem 4.4 can be rewritten in a somewhat stronger form. Namely, φ^* is optimal in a given subclass of algorithms $\Phi_0(N)$ iff (4.21) holds for all algorithms from $\Phi_0(N)$ whenever the subclass $\Phi_0(N)$ has the property: $\varphi^*, \varphi \in \Phi_0(N)$ implies $c_1\varphi^* + c_2\varphi \in \Phi_0(N)$ for real c_1 and c_2. This holds for the class of linear or affine algorithms.

5. Optimal Information

In the previous section we studied optimal average error algorithms using the information operator N of cardinality n of the form

$$N(f) = [(f,g_1),(f,g_2),\ldots,(f,g_n)].$$

In this section we find the optimal choice of elements $g_1,g_2,\ldots,g_n$. By optimal choice we mean elements for which the average error is minimized for a given class of algorithms.

We shall need the following result. Let $K : F_1 \to F_1$ be a self-adjoint nonnegative definite operator, $K = K^* \geq 0$. Let

$$(5.1) \qquad Kz_i = \lambda_i z_i, \qquad i = 1,2,\ldots$$

where $z_1,z_2,\ldots$ is an orthonormal basis and $\lambda_1 \geq \lambda_2 \geq \cdots$.

<u>Lemma 5.1</u>:

$$\max\{\Sigma_{i=1}^{n}(Kb_i,b_i) \ : \ (b_i,b_j) = \delta_{ij}\} = \Sigma_{i=1}^{n}(Kz_i,z_i)$$

$$= \Sigma_{i=1}^{n}\lambda_i.$$

<u>Proof:</u> Although Lemma 5.1 follows from Theorem 4.1.4 of [2] we provide a short proof for completeness. Let $b_i = \Sigma_{j=1}^{m} a_{ij}z_j$ where $a_{ij} = (b_i,z_j)$ and $n \leq m$. Since both b_i and z_j are orthonormal, $\Sigma_{j=1}^{m} a_{ij}^2 = \Sigma_{i=1}^{m} a_{ij}^2 = 1$. Then

$$\Sigma_{i=1}^{n}(Kb_i,b_i) = \Sigma_{i=1}^{n}\{\Sigma_{j=1}^{n} \lambda_j a_{ij}^2 + \Sigma_{j=n+1}^{m} \lambda_j a_{ij}^2\}$$

$$\le \Sigma_{i=1}^{n}\{\Sigma_{j=1}^{n} \lambda_j a_{ij}^2 + \lambda_{n+1}(1 - \Sigma_{j=1}^{n} a_{ij}^2)\}$$

$$= n\lambda_{n+1} + \Sigma_{j=1}^{n}(\lambda_j - \lambda_{n+1})\Sigma_{i=1}^{n} a_{ij}^2$$

$$\le n\lambda_{n+1} + \Sigma_{j=1}^{n}(\lambda_j - \lambda_{n+1}) = \Sigma_{j=1}^{n} \lambda_j .$$

Equality is obtained for $b_i = z_i$. ■

We now solve the problem of finding an information operator of cardinality n for which the average error of linear or affine algorithms is minimized.

Let $K_\mu = (ST_\mu^{-1})^* ST_\mu^{-1}$ and $K_c = (ST_c^{-1})^* ST_c^{-1}$. Let

(5.2) $$K_\mu z_{\mu,i} = \lambda_{\mu,i} z_{\mu,i}, \quad K_c z_{c,i} = \lambda_{c,i} z_{c,i}$$

where $\{z_{\mu,i}\}$ and $\{z_{c,i}\}$ are orthonormal basis and $\lambda_{\mu,i} \ge \lambda_{\mu,2}, \dots, \lambda_{c,1} \ge \lambda_{c,2} \ge \cdots$. Observe that

$$(K_\mu x,x) \le \|S\|^2 (T_\mu^{-1}x, T_\mu^{-1}x) = \|S\|^2 (S_\mu x,x),$$

$$(K_c x,x) \le \|S\|^2 (T_c^{-1}x, T_c^{-1}x) = \|S\|^2 (S_c x,x).$$

This yields that K_μ and K_c have finite traces and

$$\mathrm{trace}(K_\mu) = \Sigma_{i=1}^{m} \lambda_{\mu,i}$$

$$\mathrm{trace}(K_c) = \Sigma_{i=1}^{m} \lambda_{c,i}.$$

Note also that $K_\mu = T_\mu^{-1}S^*ST_\mu^{-1}$ and $\lambda_{\mu,i}T_\mu z_{\mu,i} = S^*ST_\mu^{-1}z_{\mu,i}$. Thus $T_\mu z_{\mu,i}$ is well defined whenever $\lambda_{\mu,i}$ is nonzero. Similarly, $T_c z_{c,i}$ is well defined whnever $\lambda_{c,i}$ is nonzero.

Define the information operators

$$N_{\mu,n}(f) = [(f,T_\mu z_{\mu,1}),(f,T_\mu z_{\mu,2}),\ldots,(f,T_\mu z_{\mu,k_1})],$$

$$N_{c,n}(f) = [(f,T_c z_{c,1}),(f,T_c z_{c,2}),\ldots,(f,T_c z_{c,k_2})]$$

where

$$k_1 = \min\{n,\max\{i : \lambda_{\mu,i} > 0\}\}$$

$$k_2 = \min\{n,\max\{i : \lambda_{c,i} > 0\}\}.$$

Note that $\lambda_{\mu,1} \geq \cdots \geq \lambda_{\mu,k_1} > 0$ and $\lambda_{c,1} \geq \cdots \geq \lambda_{c,k_2} > 0$ which yields that $N_{\mu,n}$ and $N_{c,\mu}$ are well defined. Of course, $\mathrm{card}(N_{\mu,n}) = k_1 \leq n$ and $\mathrm{card}(N_{c,n}) = k_2 \leq n$.

We say that N^{lo}, card $(N^{lo}) \leq n$, is <u>nth linearly optimal</u> iff

$$e^{avg}(\varphi^s,N^{lo}) = \inf\{e^{avg}(\varphi,N):\mathrm{card}(N) \leq n,\varphi\text{-linear}\},$$

N^{ao}, $\mathrm{card}(N^{ao}) \leq n$, is <u>nth affinely optimal</u> iff

$$e^{avg}(\varphi^{as},N^{ao}) = \inf\{e^{avg}(\varphi,N):\mathrm{card}(N)\leq n,\varphi\text{-affine}\}$$

and N^o, $\mathrm{card}(N^o) \leq n$, is <u>nth optimal</u> iff

$$r^{avg}(N^o) = \inf\{r^{avg}(N):\mathrm{card}(N) \leq n\}.$$

We are ready to prove the following theorem.

<u>Theorem 5.1:</u> The information operator $N_{\mu,n}$ is nth linearly optimal and

$$(5.4)\qquad e^{avg}(\varphi^{s}, N_{\mu,n}) = \sqrt{\Sigma_{i=n+1}^{m} \lambda_{\mu,i}}.$$

The information operator $N_{c,n}$ is nth affinely optimal and

$$(5.5)\qquad e^{avg}(\varphi^{as}, N_{c,n}) = \sqrt{\Sigma_{i=n+1}^{m} \lambda_{c,i}}.$$

If (4.18) holds for arbitrary information of cardinality at most n then the information operator $N_{c,n}$ is nth optimal and

$$(5.6)\qquad r^{avg}(N_{c,n}) = \sqrt{\Sigma_{i=n+1}^{m} \lambda_{c,i}}\ .$$ ■

<u>Proof</u>: To prove (5.4) we use (4.8) of Theorem 4.1 . Then $e^{avg}(\varphi, N)^2 \geq \Sigma_{i=n+1}^{m} \|ST_{\mu}^{-1}\eta_i\|^2 = \Sigma_{i=n+1}^{m}(K_{\mu}\eta_i, \eta_i)$ for orthonormal $\eta_1, \eta_2, \ldots, \eta_n, \eta_{n+1}, \ldots$ Since K_{μ} has finite trace, Lemma 5.1 yields

$$\Sigma_{i=n+1}^{m}(K_{\mu}\eta_i, \eta_i) = \text{trace}(K_{\mu}) - \Sigma_{i=1}^{n}(K_{\mu}\eta_i, \eta_i)$$

$$\geq \text{trace}(K_{\mu}) - \max\{\Sigma_{i=1}^{n}(K_{\mu}b_i, b_i) : (b_i, b_j) = \delta_{ij}\}$$

$$= \text{trace}(K_{\mu}) - \Sigma_{i=1}^{n} \lambda_{\mu,i} = \Sigma_{i=n+1}^{m} \lambda_{\mu,i}.$$

Thus

$$(5.7)\qquad e^{avg}(\varphi, N) \geq \sqrt{\Sigma_{i=n+1}^{m} \lambda_{n,i}}.$$

We now compute the average error of the spline algorithm φ^s using $N_{\mu,n}$. From (4.2) we have $\eta_i = S_\mu^{1/2} T_\mu z_{\mu,i} = z_{\mu,i}$, $i = 1,2,\ldots,k_1$. Thus (4.8) yields $e^{avg}(\varphi^s, N_{\mu,n})^2 = \Sigma_{i=k_1+1}^m \lambda_{\mu,i}$. If $k_1 < n$ then $\lambda_{\mu,i} = 0$, $i \geq k_1 + 1$ and $e^{avg}(\varphi^s, N_{\mu,n}) = 0$. Thus in both cases $k_1 < n$ and $k_1 = n$ we have

$$e^{avg}(\varphi^s, N_{\mu,n}) = \sqrt{\Sigma_{i=n+1}^m \lambda_{\mu,i}}.$$

This and (5.7) complete the proof of (5.4).

To prove (5.5) it is enough to repeat the same argument with T_μ and K_μ replaced by T_c and K_c.

Assume now that (4.18) holds for arbitrary information of cardinality at most n. Theorem 4.3 and (5.5) yield that

$$r^{avg}(N) = e^{avg}(\varphi^{as}, N) \geq e^{avg}(\varphi^{as}, N_{c,n})$$

$$= r^{avg}(N_{c,n}) = \sqrt{\Sigma_{i=n+1}^m \lambda_{c,i}}\ ,$$

which proves (5.6) and completes the proof. ∎

We stress that (4.18) holds for measures introduced in Examples 1 and 2, i.e., for measures of the form (3.9) in a finite dimensional Hilbert space and for the Gaussian measures.

<u>Remark 5.1</u>: We discuss uniqueness of optimal information. We say two information operators are equal iff they have the same kernel. From the proof of Theorem 5.1 it immediately follows that $N_{\mu,n}$ is unique nth linearly optimal whenever

$\lambda_{\mu,n} > \lambda_{\mu,n+1}$ and $N_{c,n}$ is unique nth affinely optimal whenever $\lambda_{c,n} > \lambda_{c.n+1}$. The information $N_{c,n}$ is also nth optimal whenever (4.18) holds for arbitrary information of cardinality at most n. Thus if the (n+1)st corresponding eigenvalue is strictly less than the nth we have unique optimal information operators. ∎

ACKNOWLEDGEMENTS

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

Bibliography

[1] Kuo, Hui-Hsuing, Gaussian Measures in Banach Spaces, Lecture Notes in Mathematics 463, Springer-Verlag, Berlin 1975.

[2] Marcus, M. and Minc, H., A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Inc., Boston, 1964.

[3] Micchelli, C.A., Orthogonal Projections are Optimal Algorithms, to appear.

[4] Skorohod, A.V., Integration in Hilbert Space, Springer Verlag, New York, 1974.

[5] Traub, J.F. and Woźniakowski, H., A General Theory of Optimal Algorithms, Academic Press, New York, 1980.

[6] Traub, J.F., Wasilkowski, G.W. and Woźniakowski, H., Information, Uncertainty, Complexity, Addison-Wesley, Reading, Mass. 1983.

[7] Traub, J.F., Wasilkowski, G.W., and Woźniakowski, H., Average Case Optimality for Linear Problems, Dept. of Computer Science Rep., Columbia Univ. (1981). To appear in Th. Comp. Sci.

[8] Woźniakowski, H., Can Adaption help on the Average?, in progress.