CUCS-70-83

LOCAL AVERAGE ERROR

G.W. Wasilkowski

Department of Computer Science

Columbia University

New York, New York

August 1983

This research was supported in part by the National Science Foundation under Grant MCS-7823676.

Abstract

An average case setting for linear problems has been studied in a series of recent papers. Optimal algorithms and optimal information were obtained for certain probability measures.

In this paper the local average error of algorithms and local average radius of information are defined. Using these concepts, optimal information and optimal algorithms can be found for nonlinear problems and arbitrary Borel measures.

1. Introduction

The average case setting for linear problems is studied in [4,7,8]. More precisely, the (global) average error of an algorithm and the (global) average radius of information are defined and optimal algorithms and information are found for certain probability measures.

In this paper we define the local average error and the local average radius. Analogous concepts occur, for example, in statistics, where they are used primarily for discrete and finite dimensional problems. Since we are primarily interested in infinite dimensional problems, we study the local average error and radius in an abstract setting. It is often assumed that the local average radius is a measurable function. We want to establish rather than assume the measurability of the local average radius since this is crutial to our study.

We motivate our interest in the local average error and the local average radius. These concepts

(i) lead to formulas which, in principle, give an optimal algorithm and optimal information for nonlinear problems and arbitrary Borel measure. Whether this leads to "practical" formulas depends on the problem.

(ii) enable us to study any algorithm, in general nonmeasurable. The measurability of optimal algorithms is proven, not assumed.

The results reported in this paper are primarily of theoretical interest. They will be applied to a variety of problems in future papers, the flavor of which is discussed in Section 6.

We summarize the contents of the paper. In Sections 2,3,4 we deal with problems defined on separable Banach spaces. This is only for simplicity; a generalization is discussed in Section 5. In Section 2 we recall properties of conditional measure which are crutial for our study. In Section 3 we define our basic concepts and prove that the local average radius and the local average error of any optimal algorithm are measurable. In Section 4 we illustrate these concepts for an orthogonally invariant measure. We exhibit optimal algorithms and optimal information and we establish some important properties of orthogonally invariant measures. In Section 5 we generalize all results to the case where the error is not necessarily measured by a norm.

2. Conditional measure.

In this section we recall the concept of conditional measure which will be needed to define local average error and local average radius. For simplicity we confine ourself to measures defined on separable Banach spaces. A generalization is given in Section 5.

Let F_1 be a separable Banach space and let μ be a probability measure defined on $\mathbf{B}(F_1)$, where $\mathbf{B}(F_1)$ is the σ-field of Borel sets from F_1. We shall assume that the measure μ is complete, i.e., $\mu(B) = 0$ implies that every subset of B is a Borel set. Let N,

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

be an operator (nonlinear in general) which maps F_1 onto $\mathbb{R}^n$. We shall also assume that N is measurable, i.e.,

$$(2.1) \qquad N^{-1}(A) \in \mathbf{B}(F_1), \qquad \forall \text{ measurable } A \ (A \in \mathbf{B}(\mathbb{R}^n)).$$

Define the measure $\mu_1 = \mu_1(\cdot, N)$ as

$$(2.2) \qquad \mu_1(A) = \mu(N^{-1}(A)) \qquad (= \mu(\{f \in F_1 : N(f) \in A\})), \ \forall A \in \mathbf{B}(\mathbb{R}^n).$$

Then μ_1, called the probability distribution of N, is a probability measure on $\mathbf{B}(\mathbb{R}^n)$, $\mu_1(\mathbb{R}^n) = 1$, and tells us the probability that $N(f) \in A$.

For given $y \in \mathbb{R}^n$ let

$$(2.3) \qquad V(N,y) = N^{-1}(y) \qquad (=\{f \in F_1 : N(f) = y\}).$$

From [2, Th. 8.1, p. 147] we know that there exists a unique (modulo set of $\mu_1(\cdot,N)$ measure zero) family of probability measures $\mu_2(\cdot|y) = \mu_2(\cdot|y,N)$ defined on $\mathbf{B}(F_1)$ such that

(i) $\qquad \mu_2(V(N,y)|y) = \mu_2(F_1|y) = 1, \qquad \forall y, \text{a.e.},$

(ii) $\qquad \mu_2(B|\cdot)$ is μ_1-measurable, $\quad \forall B \in \mathbf{B}(F_1)$,

(iii) $\qquad \mu(B) = \int_{\mathbb{R}^n} \mu_2(B|y)\mu_1(dy), \qquad \forall B \in \mathbf{B}(F_1).$

We shall call μ_2 the <u>conditional measure</u> with respect to N. Note that due to (i) we have $\mu_2(B \cap V(N,y)|y) = \mu_2(B|y)$ and if $y \notin N(B)$ then $B \cap V(n,y) = \emptyset$ and $\mu_2(B|y) = 0$. Hence if $N(B)$ is measurable then (iii) can be rewritten as

$$(2.4) \qquad \mu(B) = \int_{N(B)} \mu_2(B \cap V(N,y)|y)\mu_1(dy)$$

$$= \int_{N(B)} \Big[\int_{B \cap V(n,y)} \mu_2(df|y) \Big] \mu_1(dy).$$

Hence $\mu_2(B \cap V(N,y)|y)$ tells us the probability (measure) of the set B under the condition that $N(f) = y$. This justifies the name conditional measure.

<u>Remark 2.1</u>: In this paper we assume that N maps F_1 onto $\mathbb{R}^n$ since this assumption guarantees the existence and

uniqueness of the conditional measure $\mu_2(\cdot|y,N)$. This assumption can be weakened since in general one may consider any measurable map $N: F_1 \to H$ where H with its σ-field is a separable standard Borel space (see [2]) or, in particular, where H is a complete separable metric space with $\mathbb{B}(H)$ as its σ-field. Since for many problems $N(F_1) = \mathbb{R}^n$ we make the above assumption although all our results to be presented also hold if $N(F_1) = H$ is not necessarily equal to $\mathbb{R}^n$. ■

Let now $G, G: F_1 \to \mathbb{R}_+$, be a measurable nonnegative function. Then the integral $\int_{V(N,y)} G(f)\mu_2(df|y)$ is μ_1 measurable, as a function of y, and

$$(2.5) \qquad \int_{F_1} G(f)\mu(df) = \int_{\mathbb{R}^n}\{\int_{V(N,y)} G(f)\mu_2(df|y)\}\mu_1(dy).$$

(This can be easily proven by taking G to be a simple function, $G(f) = \Sigma_i c_i\chi_{B_i}(f)$.) The essence of (2.5) is that we first integrate G over all elements from B that have a fixed value of N, $N(f) = y$, and next over all values y.

We illustrate the concept of conditional measure by the following simple example.

<u>Example</u>: Suppose that $F_1 = \mathbb{R}^m$ and that μ is defined as follows

$$\mu(B) = \int_B w(f)d_m f$$

(here $d_m f$ stands for the m dimensional Lebesgue measure) for some positive function $w: \mathbb{R}^m \to \mathbb{R}_+$ such that $\int_{\mathbb{R}^m} w(f) d_m f = 1$. Then, μ is a probability measure. Let $N(f) = [f_1, f_2, \ldots, f_n]$ where $n < m$ and $f = [f_1, \ldots, f_m]$. To find the measure $\mu_1 = \mu_1(\cdot, N)$ take an arbitrary set $A \in \mathbf{B}(\mathbb{R}^n)$. Then

$$(2.6) \qquad \mu_1(A,N) = \mu(N^{-1}(A)) = \mu(\{f \in \mathbb{R}^m: [f_1, \ldots, f_n] \in A\})$$
$$= \int_A \left(\int_{\mathbb{R}^{m-n}} w([y, f^2]) d_{m-n} f^2\right) d_m y$$
$$= \int_A w_1(y) d_n y$$

where

$$(2.7) \qquad w_1(y) = \int_{\mathbb{R}^{m-n}} w([y, f^2]) d_{m-n} f^2$$

and $[y, f^2] = [y_1, \ldots, y_n, f_1^2, \ldots, f_{m-n}^2]$ for every $y = [y_1, \ldots, y_m]$ and every $f^2 = [f_1^2, \ldots, f_{m-n}^2]$. Hence $\mu_1(\cdot, N)$ is a weighted Lebesgue measure with the weight w_1 defined by (2.7). We now find $\mu_2(\cdot|y) = \mu_2(\cdot|y, N)$. Take $y \in \mathbb{R}^n$ and $B \in \mathbf{B}(\mathbb{R}^m)$. Let

$$B_{2,y} = \{f^2 \in \mathbb{R}^{m-n}: [y, f^2] \in B\}.$$

Observe that $B_{2,y} = \emptyset$ if $y \notin N(B)$. Then

$$\mu(B) = \int_{\mathbb{R}^n} \left(\int_{B_{2,y}} w([y, f^2]) d_{m-n} f^2\right) d_n y$$
$$= \int_{\mathbb{R}^n} w_1(y) \left(\int_{B_{2,y}} \frac{w([y, f^2])}{w_1(y)} d_{m-n} f^2\right) d_n y.$$

This gives the formula for $\mu_2(\cdot\,|y,N)$. Namely,

$$\mu_2(B|y,N) = \int_{B_{2,y}} \frac{w([y,f^2])}{w_1(y)}\, d_{m-n}f^2 \tag{2.8}$$

for every $B \in \mathbf{B}(\mathbf{R}^m)$ and every $y \in \mathbf{R}^n$ except $w_1(y) = 0$. ∎

3. Local average error and local average radius.

In this section we define the local average error and local average radius and we study their properties.

Suppose we want to approximate $S(f)$ where S, called a solution operator, is a measurable operator, $S: F_1 \to F_2$, and F_1, F_2 are separable Banach spaces with the Borel σ-fields $\mathbf{B}(F_1)$ and $\mathbf{B}(F_2)$ respectively. We assume that we possess information $N(f)$, where $N: F_1 \to \mathbf{R}^n$, called an information operator, satisfies all assumptions from the previous section. Then we approximate $S(f)$ by $\varphi(N(f))$ where φ, called an algorithm, is any mapping $\varphi: \mathbf{R}^n \to F_2$.

For the reader's convenience we now summarize the basic notions of the average case setting studied in [4,7,8]. For given N and φ, the (global) average error of φ is defined by

$$(3.1) \qquad e^{avg}(\varphi,N) = \{\int_{F_1} \|S(f) - \varphi(N(f))\|^2 \mu(df)\}^{1/2},$$

where μ is a probability measure on $\mathbf{B}(F_1)$. Observe that this definition requires the algorithm φ to be error measurable, i.e., $\|S(\cdot) - \varphi(N(\cdot))\|^2$ is a measurable function of f, and therefore the class of algorithms is restricted to the class, Φ_μ, of error measurable algorithms. Then an optimal algorithm φ^* is defined by $\varphi^* \in \Phi_\mu$ and

$$e^{avg}(\varphi^*, N) = r^{avg}(N)$$

where $r^{avg}(N)$, called the <u>(global) average radius</u> of N is given by

$$r^{avg}(N) = \inf_{\varphi \in \Phi_\mu} e^{avg}(N).$$

As we shall see in this section the concept of <u>local</u> average error enables us to extend the definition of global average error to the class of all algorithms, which means that we do not have to restrict ourselves to the class Φ_μ.

For an arbitrary algorithm φ, we define the <u>local average error</u> (l.a.e) of φ as

$$(3.2) \qquad e^{avg}(\varphi, N, y) = \left\{ \int_{V(N,y)} \|S(f) - \varphi(y)\|^2 \mu_2(df|y) \right\}^{1/2}$$

where $\mu_2(\cdot|y) = \mu_2(\cdot|y, N)$ is the conditional measure with respect to N defined as in Section 2. Hence $e^{avg}(\varphi, N, y)^2$ is the average value with respect to $\mu_2(\cdot|y)$ of the distance $\|S(f) - \varphi(y)\|^2$ between the solution $S(f)$ and the approximation $\varphi(y)$. Note, that l.a.e. is well defined for <u>every</u> algorithm φ (not necessarily error measurable). Indeed, since $\varphi(y)$ is a fixed element from F_2, the existence of the integral (3.2) follows from the measurability of S. However, $e^{avg}(\varphi, N, \cdot)$ need not be a μ_1 measurable function. Therefore, to define the global average error of φ we proceed as follows. Let

$$(3.3) \qquad \mathcal{H}(\varphi) = \{H: \mathbb{R}^n \to \mathbb{R}_+ : H(y) \geq e^{avg}(\varphi,N,y)^2, \ \forall y, \text{ a.e.}, \text{ and } H \text{ is } \mu_1 \text{ measurable}\}.$$

Then by the <u>global average error</u> (g.a.e.) of φ we mean

$$(3.4) \qquad e^{avg}(\varphi,N) = \inf_{H\in\mathcal{H}(\varphi)} \sqrt{\int_{\mathbb{R}^n} H(y)\mu_1(dy)}$$

if $\mathcal{H}(\varphi)$ is nonempty. Otherwise $e^{avg}(\varphi,N) = +\infty$. Note that now the global average error is well defined for <u>every</u> algorithm φ. Furthermore, if φ is error measurable then $e^{avg}(\varphi,N,\cdot)^2$ is μ_1 measurable and, due to (2.5), we have

$$(3.5) \qquad e^{avg}(\varphi,N)^2 = \int_{\mathbb{R}^n} e^{avg}(\varphi,N,y)^2\mu_1(dy).$$

This means for error measurable φ, the definitions (3.4) and (3.1) coincide.

We shall say that an algorithm φ^* that uses N is <u>optimal</u> iff

$$(3.6) \qquad e^{avg}(\varphi^*,N) = \inf_{\varphi} e^{avg}(\varphi,N).$$

Furthermore, we shall say that an algorithm φ^* that uses N is <u>strongly optimal</u> iff

$$(3.7) \qquad e^{avg}(\varphi^*,N,y) = \inf_{\varphi} e^{avg}(\varphi,N,y), \qquad \forall y, \text{a.e.}$$

Of course, a strongly optimal algorithm is also optimal. We shall prove that the opposite statement is also true. It will be also proven that every optimal algorithm is error measurable. Before that, we introduce the average radius of N.

We define the <u>local average radius</u> (l.a.r.) of N as

$$(3.8) \qquad r^{avg}(N,y) = \inf_{g \in F_2} \{ \int_{V(n,y)} \|S(f) - g\|^2 \mu_2(df|y) \}^{1/2}.$$

Of course,

$$(3.9) \qquad r^{avg}(N,y) = \inf_{\varphi} e^{avg}(\varphi,N,y),$$

which means that φ^* is strongly optimal iff $e^{avg}(\varphi^*,N,y) = r^{avg}(N,y)$ for almost every y. To define the global average radius of N we need the following lemma.

<u>Lemma 3.1</u>: The squared local average radius $r^{avg}(N,y)^2$ of N is μ_1 measurable as a function of y. ∎

<u>Proof</u>: We need only to prove that for any real number a the set $B(a) = \{y \in \mathbb{R}^n : r^{avg}(N,y)^2 > a\}$ is μ_1 measurable. For $y \in \mathbb{R}^n$ let $R(y,\cdot): F_2 \to \mathbb{R}_+$ where

$$R(y,g) = \int_{V(N,y)} \|S(f)-g\|^2 \mu_2(df|y).$$

Then $R(y,\cdot)$ is continuous and $r^{avg}(N,y)^2 = \inf\{R(y,g) : g \in F_2\}$.

Furthermore,

(3.10) $$B(a) = \{y \in \mathbb{R}^n : \forall g \in F_2,\ R(y,g) > a\} = \bigcap_{g \in F_2} B_g(a),$$

where $B_g(a) = \{y \in \mathbb{R}^n : R(y,g) > a\}$. Since F_2 is separable, then there exists a countable subset G which is dense in F_2, and, of course,

(3.11) $$B(a) \subseteq \bigcap_{g \in G} B_g(a).$$

We prove that $B(a) = \bigcap_{g\in G} B_g(a)$. For this purpose take $\bar{g} \in F_2$ and $y \in \bigcap_{g \in G} B_g(a)$. Then $R(y,g) \geq a$, $\forall g \in G$. Since $R(y,\cdot)$ is continuous and $\bar{g} = \lim_i g_i$ for some $g_i \in G$, we have $R(y,\bar{g}) = \lim_i R(y,g_i) \geq a$. Thus $y \in B_{\bar{g}}(a)$, $\forall \bar{g} \in F_2$, and $y \in B(a)$. Hence $\bigcap_{g \in G} B_g(a) \subseteq B(a)$ which with (3.11) yields

(3.12) $$B(a) = \bigcap_{g \in G} B_g(a)$$

as claimed. Every set $B_g(a)$ is μ_1 measurable since $R(\cdot,g)$ is μ_1 measurable for every g. Hence the set $B(a)$, as an intersection of countably many μ_1 measurable sets, is also μ_1 measurable. This completes the proof. ■

<u>Remark 3.1</u>: As we shall see, this lemma plays a crucial role in the study of optimal algorithms. In the proof, we intentionally did not use the fact that F_1 and F_2 are Banach spaces.

The only important assumptions are separability of F_1 and F_2 and continuity of $R(y,\cdot)$. This will enable us in Section 5 to generalize all results to the case where F_1 and F_2 are separable metric spaces. ∎

We define the <u>global average radius</u> (g.a.r) of N as

$$r^{avg}(N) = \{\int_{\mathbb{R}^n} r^{avg}(N,y)^2 \mu_1(dy)\}^{1/2}. \tag{3.13}$$

Due to Lemma 3.1, $r^{avg}(N)$ is well defined. Furthermore we have

<u>Theorem 3.1</u>: For every N

$$r^{avg}(N) = \inf_{\varphi} e^{avg}(\varphi,N). \tag{3.14}$$

If $r^{avg}(N)$ is finite then

(i) an algorithm φ is optimal iff φ is strongly optimal, and

(ii) every optimal algorithm φ is error measurable, i.e., $\|S(\cdot)-\varphi(N(\cdot))\|^2$ is μ_1 measurable. ∎

<u>Proof</u>: We begin with (3.14). Let

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

Since, $R \geq r^{avg}(N)^2$, to prove (3.14) we need only to show that $R \leq r^{avg}(N)^2$. For positive δ let φ_δ be an algorithm such that $e^{avg}(\varphi,N,y)^2 = r^{avg}(N,y)^2 + \delta$, $\forall y \in \mathbb{R}^n$ (a.e.). Of course such algorithm exists and, due to Lemma 3.1, $e^{avg}(\varphi,N)^2 = r^{avg}(N)^2 + \delta$. Hence $r^{avg}(N)^2 + \delta \geq R$. Since δ is arbitrary, this means that $R \leq r^{avg}(N)^2$ which completes the proof of (3.14).

Suppose now that $r^{avg}(N) < +\infty$. Since every strongly optimal algorithm is optimal, we need only to prove that optimality of φ implies strong optimality. To show this let

$$P = \{y \in \mathbb{R}^n: e^{avg}(\varphi,N,y) > r^{avg}(N,y)\}.$$

We prove that the set P is μ_1 measurable and that its measure is zero. Indeed, for $i = 1,2,\ldots$ let $Q_i = \{y \in \mathbb{R}^n: e^{avg}(\varphi,N,y)^2 \geq r^{avg}(N,y)^2 + \frac{1}{i}\}$. Then $Q_i \subset Q_{i+1}$ and $\bigcup_{i=1}^{\infty} Q_i = P$. Due to (3.3), there exists a sequence $\{H_k\}$ of μ_1 measurable functions such that $H_k(y) \geq e^{avg}(\varphi,N,y)^2$ and $\lim_k \int_{\mathbb{R}^n} H_k(y)\mu_1(dy) = e^{avg}(\varphi,N)^2$. Define

$$Q_{i,k} = \{y \in \mathbb{R}^n: H_k(y) \geq r^{avg}(N,y)^2 + \tfrac{1}{i}\}.$$

Then $Q_i \subset Q_{k,i}$ and $Q_i \subset \tilde{P}_i = \bigcap_{k=1}^{\infty} Q_{k,i}$. Observe now that

$$\begin{aligned} r^{avg}(N)^2 = e^{avg}(\varphi,N)^2 &= \lim_k \int_{\mathbb{R}^n} H_k(y)\mu_1(dy) \\ &= r^{avg}(N)^2 + \lim_k \int_{\mathbb{R}^n} (H_k(y) - r^{avg}(N)^2)\mu_1(dy) \\ &\geq r^{avg}(N)^2 + \lim_k \int_{\tilde{P}_i} \frac{1}{i}\mu_1(dy) = r^{avg}(N)^2 + \frac{1}{i}\mu_1(\tilde{P}_i) \end{aligned}$$

which means that $\mu_1(\tilde{P}_i) = 0$. Since $Q_i \subset \tilde{P}_i$ and μ_1 is complete (the completeness of μ_1 follows from the completeness of μ), then Q_i is μ_1 measurable and $\mu_1(Q_i) = 0$. This implies that also $P = \bigcup_{i=1}^{\infty} Q_i$ is measurable and $\mu_1(P) = 0$, as claimed. This means that $e^{avg}(\varphi,N,y) = r^{avg}(N,y)$, $\forall y \in \mathbb{R}^n$ (a.e.), which proves that φ is strongly optimal.

We now prove that for every optimal algorithm φ, $\|S(\cdot) - \varphi(N(\cdot))\|^2$ is μ-measurable. Indeed, since optimality of φ implies strong optimality, then $e^{avg}(\varphi,N,y)^2 = r^{avg}(N,y)^2$. Hence $e^{avg}(\varphi,N,\cdot)^2$ is μ_1 measurable and

$$e^{avg}(\varphi,N)^2 = \int_{\mathbb{R}^n} e^{avg}(\varphi,N,y)^2 \mu_1(dy)$$

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

which means that $\|S(\cdot)-\varphi(N(\cdot))\|^2$ is μ_1 measurable. This completes the proof of Theorem 3.1. ∎

Due to (3.14) we can see that the definition (3.13) of global average radius coincides with that from papers cited at the beginning of this section. Furthermore we do not have to assume the measurability of $r^{avg}(N,\cdot)^2$, since this is a conclusion. We can also conclude that in the average case model every optimal algorithm is strongly optimal unless the radius $r^{avg}(N)$ is infinite. The assumption that $r^{avg}(N) < +\infty$ is crutial since, as we shall see in the following example, there exists an optimal algorithm φ which is not strongly optimal if $r^{avg}(N) = +\infty$.

Example 3.1: Suppose that F_1 is a separable Hilbert space with an orthonormal basis $\eta_1, \eta_2, \ldots$. Let μ be so that $\mu(\{2^{2k}\eta_k\}) = 2^{-k}$, $k = 1,2,\ldots$. Then μ is concentrated on the set $\{2^2\eta_1, 2^4\eta_2, \ldots\}$. Let $S = I$ and $N(f) = (f,\eta_1)$. It should be obvious that for every algorithm φ,

$$e^{avg}(\varphi,N)^2 = 2^{-1}\|2^2\eta_1 - \varphi(1)\|^2 + \Sigma_{k=2}^{\infty}\ 2^{-k}\|2^{2k}\eta_k - \varphi(0)\|^2$$

$$= +\infty.$$

This means that

$$r^{avg}(N)^2 = +\infty,$$

and every algorithm is optimal. Consider now an algorithm φ^*, $\varphi^*(y) = 0$, $\forall y \in \mathbb{R}$. Then its local average error is

$$e^{avg}(\varphi^*,N,1)^2 = \|2^2\eta_1\|^2 = 2^4 > 0 = r^{avg}(N,1)^2.$$

Since $\mu_1(\{1\}) = 2^{-1} > 0$, then the algorithm φ^* is not strongly optimal, although it is optimal. ■

We now show how all these concepts can be simplified by assuming that F_2 is a Hilbert space and $\int_{F_1} \|Sf\|^2\mu(df) < \infty$. Let $m = m(S,y)$, called the conditional mean element of S, be defined by

$$(3.15) \qquad (m,g) = \int_{F_1} (S(f),g)\mu_2(df|y), \qquad \forall g \in F_2 .$$

The existence and uniqueness of m for almost every y follows from the fact that $\int_{F_1} \|Sf\|\mu(df) \leq \int_{F_1} \|Sf\|^2\mu(df)$.

For an error measurable algorithm,

$$e^{avg}(\phi,N,y)^2 = \int_{F_1} \|S(f)\|^2\mu_2(df|y)+\|\phi(y)\|^2 - 2\int_{F_2} (S(f),\phi(y))\mu_2(df|y)$$

$$= \int_{F_1} \|S(f)\|^2\mu_2(df|y) + \|\phi(y)\|^2 - 2(m(S,y),\phi(y))$$

$$= \int_{F_1} \|S(f)\|^2\mu_2(df|y) - \|m(S,y)\|^2 + \|\phi(y) - m(S,y)\|^2$$

$$\geq \int_{F_1} \|S(f)\|^2\mu_2(df|y) - \|m(S,y)\|^2$$

$$= \int_{F_1} \|S(f) - m(S,y)\|^2\mu_2(df|y).$$

Hence

$$r^{avg}(N,y)^2 = \inf_{\phi} e^{avg}(\phi,N,y)^2 = \int_{F_1} \|S(f) - m(S,y)\|^2\mu_2(df|y).$$

We summarize this in

Theorem 3.2: Let F_2 be a Hilbert space. Then the unique optimal algorithm ϕ^* is given by

$$(3.16)\qquad \varphi^*(y) = m(S,y),$$

where m(s,y) is the conditional mean element of S. Furthermore

$$(3.17)\qquad r^{avg}(N,y) = \{\int_{F_2} \|S(f)\|^2 \mu_2(df|y) - \|m(S,y)\|^2\}^{1/2},$$

$$\forall y \in \mathbb{R}^n (a.e.),$$

and

$$(3.18)\qquad r^{avg}(N) = \{\int_{F_1} \|S(f)\|^2 \mu(df) - \int_{\mathbb{R}^n} \|m(S,y)\|^2 \mu_1(dy)\}^{1/2}. \blacksquare$$

We end this section by defining optimal information operator. Until now, the information operator N was fixed and we were looking for an optimal algorithm that uses N. Suppose that we vary information. What is "optimal" information? More precisely, as in [5,6] let Ψ be a class of functionals L, $L: F_1 \to \mathbb{R}$. We assume that every L from Ψ is measurable. For an integer n, let $\Psi(n)$ be the class of all information operators N, $N: F_1 \to \mathbb{R}^n$, such that

$$(3.19)\qquad N(f) = [L_n(f),\ldots,L_n(f)]$$

for some $L_i \in \Psi$. Then, roughly speaking, $\Psi(n)$ consists of all information operators of <u>cardinality</u> n which can be used to solve our problem S.

We define the <u>n-th minimal coverage radius (for the</u>

class $\Psi(n)$) as

$$(3.20) \qquad r^{avg}(n,\Psi(n)) = \inf_{N\in\Psi(n)} r^{avg}(N).$$

Then by an n-th optimal information operator (in the class $\Psi(n)$) we mean any information operator $N^* \in \Psi(n)$ such that

$$(3.21) \qquad r^{avg}(N^*) = r^{avg}(n,\Psi(n)).$$

Of course, n-th optimal information N^* has the smallest radius among all information of the same cardinality and an optimal algorithm ϕ^* that uses N^* has the smallest error among all algorithms that use any information operator of cardinality n.

4. Orthogonally invariant measure.

In this section we study optimal algorithms and optimal linear information operators assuming that the measure μ is orthogonally invariant. We first present the definition of orthogonal invariant measures with their basic properties. See [8] for a more detailed discussion. In Subsection 4.1 we exhibit further properties of orthogonal invariant measures. In Subsection 4.2 we apply these properties to linear problems, and in Subsection 4.3 we apply them to the problem $S(f) = \|f\|^2$ which is an example of a nonlinear problem.

Through this section we shall assume that F_1 is a separable Hilbert space and that $\int_{F_1} \|f\|^2 \mu(df) < \infty$. Without loss of generality we can assume that the mean element m_μ of the measure μ is zero and that $\int_{F_1} (f,x)^2 \mu(df) > 0$ unless $x = 0$. Recall that the <u>mean element</u> of μ is defined by $(m_\mu, x) = \int_{F_1} (f,x)\mu(df)$. Let S_μ be the <u>covariant operator</u> of μ, i.e., $S_\mu : F_1 \to F_1$ and

$$(4.1) \qquad (S_\mu x, z) = \int_{F_1} (f,x)(f,z)\mu(df), \qquad \forall x,z \in F_1.$$

Of course, S_μ is a linear self-adjoint, positive definite operator with finite trace.

We present the definition of orthogonal invariance. (For a more detailed discussion see [8].) We say that μ is <u>orthogonally invariant</u> iff

(4.2) $\mu(QB) = \mu(B)$

for every Borel set $B \in \mathbf{B}(F_1)$ and any linear mapping Q, $Q: F_1 \to F_1$, of the form

(4.3) $Qf = 2 \sum_{i=1}^{k} (f,\eta_i) S_\mu \eta_i - f$

for any $k \geq 0$ and any η_i such that $(S_\mu \eta_i, \eta_j) = \delta_{ij}$. Every operator Q of the form (4.3) satisfies

(4.4) $QQ = I$

and

(4.5) $\|Qf\|_* = \|f\|_*, \quad \forall f \in S_\mu(F_1),$

where $\|f\|_* = \sqrt{(f,f)_*}$ and $(f,f)_* = (S_\mu^{-1} f, f)$ is an inner product in the Hilbert space $S_\mu(F_1)$ (see [8]). Hence Q is an orthogonal mapping in $S_\mu(F_1)$. This justifies the name orthogonal invariance.

Orthogonally invariant measures have very important and interesting properties studies in [8]. Here we exhibit further properties given in terms of conditional measures of μ.

Let N, $N: F_1 \to \mathbb{R}^n$, be a <u>linear</u> continuous information operator. Without loss of generality we can assume that

(4.6) $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$, where

$$(S_\mu \eta_i, \eta_j) = \delta_{ij}, \quad \forall i,j = 1,2,\ldots,n.$$

Then card(N) = n and $N(F_1) = \mathbb{R}^n$. Let $y = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^n$. Recall that by a <u>spline element interpolating y with respect to N</u> (or briefly <u>spline</u>) we mean an element $\sigma(y,N)$ such that

$$(4.7) \qquad \sigma(y,N) = \Sigma_{i=1}^{n} y_i S_\mu \eta_i.$$

Of course, $N(\sigma(y,N)) = y$ and

$$V(N,y) = N^{-1}(y) = \sigma(y,N) + \ker N.$$

We now present some properties of orthogonally invariant measures.

4.1 Properties of orthogonally invariant measures.

For N of the form (4.6) let $\mu_1(\cdot,N)$ and $\mu_2(\cdot|y,N)$ be defined as in Section 2.

<u>Theorem 4.1</u>: Let μ be orthogonally invariant.

(i) Let N_1 and N_2 be of the form (4.6). If card(N_1) = card(N_2) then

$$\mu_1(\cdot,N_1) = \mu_1(\cdot,N_2).$$

(ii) Let N be of the form (4.6) with card(N) = n. Then the mean element $m_{N,y}$ of the measure $\mu_2(\cdot|y,n)$ is

$$m_{N,y} = \sigma(y,N) = \Sigma_{i=1}^{n} y_i S_\mu \eta_i, \quad \forall y \in \mathbb{R}^n \text{(a.e.)}.$$

(iii) $\forall n$, $\exists h: \mathbb{R}^n \to \mathbb{R}_+$, h-measurable, $\forall N$ of the form (4.6) with card(N) = n:

$$S_{N,y} = h(y)\cdot(I-\sigma_N)S_\mu(I-\sigma_N^*)$$

is the correlation operator of the measure $\mu_2(\cdot|y,N)$, $\forall y \in \mathbb{R}^n$(a.e.). Furthermore

$$\int_{\mathbb{R}^n} h(y)\mu_1(dy,N) = 1.$$

Here $\sigma_N: F_1 \to F_1$ is a linear operator defined by $\sigma_N(f) = \sigma(N(f),N)$. ∎

<u>Proof:</u> See appendix. ∎

Recall that the <u>correlation operator of a measure</u> λ is defined to be the covariance operator of the translated measure $\tilde{\lambda}$, $\tilde{\lambda}(A) = \lambda(A-m_\lambda)$, or equivalently an operator $S_c: F_1 \to F_1$ such that

$$(4.8) \qquad (S_c x,z) = \int_{F_1} (f-m_\lambda,x)(f-m_\lambda,z)\lambda(df), \qquad \forall x,z \in F_1,$$

where m_λ is the mean element of λ.

Theorem 4.1 states that the measure $\mu_1(\cdot,N)$ is independent of N. It depends only on the cardinality of N. Hence $\mu_1(\cdot,N) = \mu_1(\cdot)$ for some measure on $\mathcal{B}(\mathbb{R}^n)$. From (ii) we know that the mean element of $\mu_2(\cdot|y,N)$ is spline $\sigma(y,N)$ and from (iii) we know that, regardless of the constant h(y), the conditional measure $\mu_2(\cdot|y,N)$ has the same correlation operator for almost every $y \in \mathbb{R}^n$.

It is shown in [7] that the Gaussian measures are orthogonally invariant. We now study their conditional measures.

Recall that by a Gaussian measure on a Hilbert space F_1 we mean a measure λ such that

$$(4.9) \qquad \int_{F_1} e^{i(f,x)}\lambda(df) = \exp\{i(a,x)-\frac{1}{2}(Ax,x)\}, \quad \forall x \in F_1, \quad (i=\sqrt{-1}),$$

where $A\colon F_1 \to F_1$ is a self-adjoint nonnegative definite operator with finite trace and a is an element of F_1. (The left hand side of (4.9) is called the characteristic functional of λ and is denoted by $\psi_\lambda(x)$.) Then the mean element m_λ of λ is given by

$$(4.10) \qquad m_\lambda = a$$

and the correlation operator S_λ, by

$$(4.11) \qquad S_\lambda = A$$

(see [1,2,3]).

Suppose now that μ is the Gaussian measure with mean element zero and covariance operator S_μ which is positive definite. (Observe that $m_\mu = 0$ implies $S_\mu = S_c$.) This is equivalent to the fact that

(4.12) $$\mu(\{f \in F_1 : (f,x) \le d\}) = \frac{1}{\sqrt{2\pi\sigma_x}} \int_{-\infty}^{d} \exp(-\frac{t^2}{2\sigma_x})\,dt,$$

$$\forall x \in F_1,\ \forall d \in \mathbb{R},$$

where $\sigma_x = (S_\mu x, x)$ (see [1]).

<u>Theorem 4.2</u>: Let μ be the Gaussian measure with mean element zero and positive definite covariance operator S_μ. Then for every information operator N of the form (4.6) with card(N) = n we have

(i) $\mu_1 = \mu_1(\cdot, N)$ is the Gaussian measure on $\mathcal{B}(\mathbb{R}^n)$ with mean element zero and covariance operator $S_1 = I$, i.e.,

$$\mu_1(A) = \frac{1}{\sqrt{(2\pi)^n}} \int_A \exp(-\frac{(x,x)}{2})\,d_n x. \qquad \forall A \in B(\mathbb{R}^n).$$

(ii) $\mu_2(\cdot|y,N)$ is the Gaussian measure on F_1 with mean element $m_{N,y} = \sigma(y,N)$ and correlation operator

$$S_{N,y} = (I-\sigma_N)S_\mu(I-\sigma_N^*). \qquad ■$$

<u>Proof</u>: Since the proof is very simple, we only sketch it. To prove (i), it is enough (due to (4.12)) to show that

$$\mu_1(A) = \frac{1}{\sqrt{2\pi(a,a)}}\int_{-\infty}^{d} \exp(-\frac{t^2}{2(a,a)})\,dt$$

for every set A of the form $A = \{y \in \mathbb{R}^n: (y,a) \le d^2\}$. This follows from the fact that $\mu_1(A) = \mu(N^{-1}(A)) = \mu\{f \in F_1: (f,g) \le d\}$ where $g = g(a) = \Sigma_{i=1}^{n} a_i \eta_i$.

To prove (ii) it is enough to show that for λ_2 defined as in (ii) the characteristic functional of μ is equal

$$\psi_\mu(x) = \int_{\mathbb{R}^n} \int_{F_1} e^{i(f,x)} \lambda_2(df|y,N)\mu_1(dy).$$

Since characteristic functional defines measure uniquely and since conditional measure is determined uniquely, we have $\mu_2 \equiv \lambda_2$ which proves the theorem. ∎

Having established properties of orthogonally invariant measures we study optimal algorithms and optimal linear information operators for certain problems. We begin with

4.2 Linear Problems

Suppose that $S: F_1 \to F_2$ is a continuous linear operator and that F_2 is a separable Hilbert space. From Theorem 3.2 we know that for every information operator N the optimal algorithm ϕ^* is of the form

$$\phi^*(y) = m(S,y)$$

where $m(S,y)$ is the conditional mean element of S, i.e.,

$$(4.13) \qquad (m(s,y),x) = \int_{F_1} (S(f),x)\mu_2(df|y,N), \qquad \forall x \in F_2.$$

Since S is now linear and F_1, F_2 are Hilbert spaces then (4.13) can be rewritten as

$$(m(S,y),x) = \int_{F_1} (f,S^*x)\mu_2(df|y,N)$$

$$= (m_{N,y}, S^*x) = (Sm_{N,y}, x), \qquad \forall x \in F_2,$$

where $m_{N,y}$ is the mean element of the conditional measure $\mu_2(\cdot|y,N)$. This implies that $m(S,y) = Sm_{N,y}$ and the optimal algorithm is given by

$$(4.14) \qquad \phi^*(y) = Sm_{N,y}.$$

Taking an orthonormal basis $h_1, h_2, \ldots$ of F_2 we get

$$(4.15) \qquad r^{avg}(N,y)^2 = \int_{F_1} \Sigma_{i=1}^{\infty} (S(f-m_{N,y}), h_i)^2 \mu_2(df|y,N)$$

$$= \Sigma_{i=1}^{\infty} \int_{F_1} (f-m_{N,y}, S^*h_i)^2 \mu_2(df|y,N)$$

$$= \Sigma_{i=1}^{\infty} (S_{N,y}S^*h_i, S^*h_i) = \text{trace}(SS_{N,y}S^*)$$

and

$$(4.16) \qquad r^{avg}(N)^2 = \int_{\mathbb{R}^n} \text{trace}(SS_{N,y}S^*)\mu_1(dy,N),$$

where $S_{N,y}$ is a correlation operator of $\mu_2(\cdot|y,N)$.

Suppose now that μ is orthogonally invariant and N is of the form (4.6). Then due to Theorem 4.1 (ii), ϕ^* is the <u>spline algorithm</u>, i.e.,

$$(4.17) \qquad \phi^*(y) = S\sigma(y,N) = \Sigma_{i=1}^{n} y_i S S_\mu \eta_i$$

and due to Theorem 4.1 (iii)

$$(4.18) \qquad r^{avg}(N,y)^2 = h(y)\cdot r^{avg}(N)^2$$

$$= h(y)\,\mathrm{trace}(S(I-\sigma_N)S_\mu(I-\sigma_N^*)S^*).$$

From this and from Theorem 4.2 we get

<u>Corollary 4.1</u>: If μ is a Gaussian measure then for almost every $y \in \mathbb{R}^n$,

$$r^{avg}(N,y) = r^{avg}(N). \qquad \blacksquare$$

The optimality of the spline algorithm was established in [7] without using the concept of local error and/or local radius. In [7] and [8] there is a simple formula on the global radius $r^{avg}(N)$ of N as well as the nth optimal linear information operator is given. Namely for given N, $N(f) = [(f,\eta_1),\dots,(f,\eta_n)]$ let $\eta_{n+1},\eta_{n+2},\dots$ be such that $F_1 = \lim\{\eta_1,\eta_2,\dots\}$ and $(S_\mu\eta_i,\eta_j) = \delta_{ij}$, $\forall i,j = 1,2,\dots$. Then

$$r^{avg}(N)^2 = \Sigma_{i=n+1}^{\infty}\|SS_\mu\eta_i\|^2.$$

This and (4.18) give immediately

$$(4.19)\qquad r^{avg}(N,y)^2 = h(y)\,r^{avg}(N)$$

$$= h(y)\,\Sigma_{i=m+1}^{\infty}\|SS_{\mu}\eta_i\|^2.$$

Furthermore, if $\zeta_1,\zeta_2,\ldots,\zeta_n$ are eigenvectors of $(SS_{\mu}^{1/2})^*(SS_{\mu}^{1/2})$ corresponding to the maximal eigenvalues, then

$$(4.20)\qquad N_n^*(f) = [(f,\eta_1^*),\ldots,(f,\eta_n^*)],\qquad \eta_i^* = S_{\mu}^{-1/2}\zeta_i,$$

is nth optimal among all linear information operators. Hence

$$(4.21)\qquad r^{avg}(n,\Psi(n)) = r^{avg}(N_n^*)$$

$$= \sqrt{\Sigma_{i=n+1}^{\infty}\|SS_{\mu}^{1/2}\zeta_i\|^2}$$

where Ψ is the class of continuous linear functionals.

We now proceed to another problem, which serves as a simple example of a nonlinear problem.

4.3 Norm evaluation problem.

Suppose we want to approximate $\|f\|^2$, i.e.,

$$(4.22)\qquad S(f) = \|f\|^2.$$

We assume that the measure μ is orthogonally invariant and $\int_{F_1}\|f\|^4\mu(df) < +\infty$. Let N be of the form (4.6), i.e., $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$ with $(S_{\mu}\eta_i,\eta_j) = \delta_{ij}$. Since $F_2 = \mathbb{R}$, F_2 is a Hilbert space and we can apply Theorem 3.2.

For this purpose we need to calculate $m(S,y)$,

$$m(S,y) = \int_{F_1} \|f\|^2 \mu_2(df|y,N) = \int_{F_1} \|f-m_{N,y}+m_{N,y}\|^2 \mu_2(df|y,N)$$

$$= \int_{F_1} \|f-m_{N,y}\|^2 \mu_2(df|y,N) + \|m_{N,y}\|^2 + 2\int_{F_1} (f-m_{N,y}, m_{N,y}) \mu_2(df|y,N)$$

$$= \int_{F_1} \|f-m_{N,y}\|^2 \mu_2(df|y,N) + \|m_{N,y}\|^2$$

$$= \text{trace } S_{N,y} + \|m_{N,y}\|^2.$$

Since μ is orthogonally invariant then, due to Theorem 4.1, we conclude that

$$m(S,y) = h(y)\Sigma_{i=m+1}^{\infty} \|S_\mu \eta_i\|^2 + \|\sigma(y,N)\|^2.$$

Hence the optimal algorithm φ^* is of the form

(4.23) $$\varphi^*(y) = h(y)\Sigma_{i=n+1}^{\infty} \|S_\mu \eta_i\|^2 + \|\sigma(y,N)\|^2.$$

The local radius of N is

(4.24) $$r^{avg}(N,y) = \{\int_{F_1} \|f\|^4 \mu_2(df|y,N) - (\varphi^*(y))^2\}^{1/2}$$

and global radius of N is

(4.24) $$r^{avg}(N) = \{\int_{F_1} \|f\|^4 \mu(df) - \int_{\mathbb{R}^n} (\varphi^*(y))^2 \mu_1(dy)\}^{1/2}.$$

We now calculate $r^{avg}(N)$ assuming that μ is Gaussian.

We begin with the following integral.

$$I_1 = \int_{\mathbb{R}^n} (\varphi^*(y))^2 \mu_1(dy).$$

Recall that now $h(y) = 1$, $\forall y$(a.e.), and μ_1 is the Gaussian measure with mean zero and covariance operator I. Denote $A = \Sigma_{i=n+1}^{\infty} \|S_{\mu} \eta_i\|^2$. Since $(\varphi^*(y))^2 = A^2 + 2A\|\sigma(y,N)\|^2 + \|\sigma(y,N)\|^4$ then

$$I_1 = A^2 + 2A\int_{\mathbb{R}^n} \|\sigma(y,N)\|^2 \mu_1(dy) + \int_{\mathbb{R}^n} \|\sigma(y,N)\|^4 \mu_1(dy).$$

It is easy to see that

$$\int_{\mathbb{R}^n} \|\sigma(y,N)\|^2 \mu_1(dy) = \Sigma_{i=1}^{n} \|S_{\mu} \eta_i\|^2$$

and that

$$\int_{\mathbb{R}^n} \|\sigma(y,N)\|^4 \mu_1(dy) = \int_{\mathbb{R}^n} \{\Sigma_{i=1}^{n} \|S_{\mu} \eta_i\|^2 y_i^2$$

$$+ 2\Sigma_{i=1}^{n-1} \Sigma_{j>i}^{n} y_i y_j (S_{\mu} \eta_i, S_{\mu} \eta_j)\}^2 \mu_1(dy)$$

$$= \Sigma_{i=1}^{n} \|S_{\mu} \eta_i\|^4 \int_{\mathbb{R}^n} y_i^4 \mu_1(dy)$$

$$+ 2\Sigma_{i=1}^{n-1} \Sigma_{j>i}^{n} \|S_{\mu} \eta_i\|^2 \|S_{\mu} \eta_j\|^2 \int_{\mathbb{R}^n} y_i^2 y_j^2 \mu_1(dy)$$

$$+ 4\Sigma_{i=1}^{n-1} \Sigma_{j>i}^{n} (S_{\mu} \eta_i, S_{\mu} \eta_j) \int_{\mathbb{R}^n} y_i^2 y_j^2 \mu_1(dy).$$

Since for Gaussian measure μ_1, $\int_{\mathbb{R}^n} y_i^4 \mu_1(dy) = 3$ and $\int_{\mathbb{R}^n} y_i^2 y_j^2 \mu_1(dy) = 1$ $(i \neq j)$, then

$$\int_{\mathbb{R}^n} \|\sigma(y,N)\|^4 \mu_1(dy) = 3\Sigma_{i=1}^n \|S_\mu \eta_i\|^4 + 2\Sigma_{i=1}^{n-1}\Sigma_{j>i}^n \|S_\mu \eta_i\|^2 \|S_\mu \eta_j\|^2$$
$$+ 4\Sigma_{i=1}^{n-1}\Sigma_{j>i}^n (S_\mu \eta_i, S_\mu \eta_j)^2$$
$$= (\Sigma_{i=1}^n \|S_\mu \eta_i\|^2)^2 + 2\Sigma_{i,j=1}^n (S_\mu \eta_i, S_\mu \eta_j)^2.$$

This means that

$$I_1 = (A + \Sigma_{i=1}^n \|S_\mu \eta_i\|^2)^2 + 2\Sigma_{i,j=1}^n (S_\mu \eta_i, S_\mu \eta_j)^2$$
$$= (\text{trace } S_\mu)^2 + 2\Sigma_{i,j=1}^n (S_\mu \eta_i, S_\mu \eta_j)^2.$$

We now calculate

$$I_2 = \int_{F_1} \|f\|^4 \mu(df).$$

Let $e_1, e_2, \ldots$ be eigenvectors of the operator S_μ, $\|e_i\| = 1$ and $S_\mu e_i = \lambda_i e_i$. Since $\|f\|^4 = \{\Sigma_{i=1}^\infty (f,e_i)^2\}^2$ $= \Sigma_{i=1}^\infty (f,e_i)^4 + 2\Sigma_{i=1}^\infty \Sigma_{j>i} (f,e_i)^2 (f,e_j)^2$, then

$$I_2 = \Sigma_{i=1}^\infty \int_{F_1} (f,e_i)^4 \mu(df) + 2\Sigma_{i=1}^\infty \Sigma_{j>i} \int_{F_1} (f,e_i)^2 (f,e_j)^2 \mu(df).$$

To calculate $\int_{F_1} (f,e_i)^4 \mu(df)$ take $N_i(f) = (f, e_i/\sqrt{\lambda_i})$. Let now $\mu_1(\cdot, N_1)$ and $\mu_2(\cdot \mid y, N_i)$ be the decomposition of μ corresponding to N_i. Since N_i is of the form (4.6) (with card(N_i) = 1), then due to theorem 4.2 we get

$$\int_{F_1} (f,e_i)^4 \mu(df) = \lambda_i^2 \int_{F_1} (f, e_i/\sqrt{\lambda_i})^4 \mu(df) = 3\lambda_i^2 \int_{\mathbb{R}} t^4 \mu_1(dt, N_1)$$
$$= 3\lambda_i^2 = 3\|S_\mu e_i\|^2.$$

Similar, taking $N_{i,j}(f) = [(f,e_i/\sqrt{\lambda_i}),\ (f,e_j/\sqrt{\lambda_j})]$, we can prove that

$$\int_{F_1}(f,e_i)^2(f,e_j)^2\mu(df) = \lambda_i\lambda_j = \|S_\mu e_i\|\|S_\mu e_j\|.$$

Hence

$$I_2 = 3\Sigma_{i=1}^\infty\|S_\mu e_i\|^2 + 2\Sigma_{i=1}^\infty\Sigma_{j>i}\|S_\mu e_i\|\|S_\mu e_j\|$$

$$= 2\Sigma_{i=1}^\infty\|S_\mu e_i\|^2 + (\text{trace } S_\mu)^2.$$

Since $r^{avg}(N)^2 = I_2 - I_1$, this yields that $r^{avg}(N)^2 = 2\{\Sigma_{i=1}^\infty\|S_\mu e_i\|^2 - \Sigma_{i,j=1}^n(S_\mu\eta_i,S_\mu\eta_j)^2\}$. Recall that $(S_\mu\eta_i,\eta_j) = \delta_{ij}$. This means that $\zeta_1,\zeta_2,\ldots,\zeta_i = S_\mu^{1/2}\eta_i$ form an orthonormal system for the space F_1 and therefore

$$\Sigma_{i=1}^\infty\|S_\mu e_i\|^2 = \Sigma_{i=1}^\infty\Sigma_{k=1}^\infty(S_\mu e_i,\zeta_k)^2 = \Sigma_{k=1}^\infty\Sigma_{i=1}^\infty(e_i,S_\mu\zeta_k)^2$$

$$= \Sigma_{k=1}^\infty\|S_\mu\zeta_k\|^2 = \Sigma_{k,l=1}^\infty(S_\mu\zeta_k,\zeta_l)^2$$

as well as $(S_\mu\eta_i,S_\mu\eta_j)^2 = (S_\mu\zeta_i,\zeta_j)^2$. Thus, finally, the global radius of N is

$$(4.26)\qquad r^{avg}(N) = \sqrt{2}\{\Sigma_{k=1}^n\Sigma_{j=n+1}^\infty(S_\mu\zeta_k,\zeta_j)^2 + \Sigma_{k=n+1}^\infty\|S_\mu\zeta_k\|^2\}^{1/2}$$

$$= \sqrt{2}\{\Sigma_{k=1}^n\Sigma_{j=n+1}^\infty(S_\mu^2\eta_k,\eta_j)^2 + \Sigma_{k=n+1}^\infty\|S_\mu^{3/2}\eta_k\|^2\}^{1/2}.$$

From (4.26) it follows that

$$r^{avg}(N)^2 \ge 2\Sigma_{k=n+1}^\infty\|S_\mu\zeta_k\|^2 = 2\{\text{trace}(S_\mu^2) - \Sigma_{k=1}^n\|S_\mu\zeta_k\|^2\}.$$

It is well known, see e.g. [7], that

$$\Sigma_{k=1}^{n} \|S_\mu \zeta_k\|^2 \le \Sigma_{k=1}^{n} \|S_\mu e_k\|^2$$

where $e_1,\dots,e_n$ correspond to the maximal eigenvalues of S_μ, $\lambda_1 \ge \lambda_2 \ge \dots \ge 0$. Define

$$(4.27) \qquad N_n^*(f) = [(f,\eta_1^*),\dots,(f,\eta_n^*)], \qquad \eta_i^* = e_i/\sqrt{\lambda_i}.$$

Then

$$r^{avg}(N)^2 \ge 2\{\mathrm{trace}(S_\mu^2) - \Sigma_{k=1}^{n}\|S_\mu e_k\|^2\}$$

$$= 2\ \Sigma_{k=n+1}^{\infty} \lambda_k^2 = r^{avg}(N_n^*)^2.$$

This shows that N_n^* is nth optimal among all linear information operators and

$$(4.28) \qquad r^{avg}(n,\psi(n)) = r^{avg}(N_n^*) = \sqrt{2\ \Sigma_{k=n+1}^{\infty} \lambda_k^2}. \qquad \blacksquare$$

5. General problems.

In the previous sections we studied an average case model for problems defined on separable Banach spaces with error criterion:

$$\|S(f) - \varphi(N(f))\| \text{ - small on the average.}$$

Of course, this is not the only interesting error criterion and therefore average case analysis should be applied to a wider class of problems. In this section we briefly discuss some generalizations.

As in [5], consider a problem defined as follows: given two sets F_1 and F_2 and a function

$$\text{dist}: F_1 \times F_2 \to \mathbb{R}_+,$$

construct an element $g = g(f) \in F_2$ such that $\text{dist}(f,g)$ is small, $\forall f \in F_1$. The function dist serves as an error criterion and for problems studied in the previous sections $\text{dist}(f,g) = \|S(f) - g\|$. In general dist need not be a metric; the name "dist" is chosen to be suggestive. In the worst case model, studied in [5], the error of an algorithm φ is defined by

$$\bar{e}(\varphi, N) = \sup_{f \in F_1} \text{dist}(f, \varphi(N(f))).$$

In the average case model, the average error is defined by

$$e(\varphi,N) = \{\int_{F_1} \operatorname{dist}^2(f,\varphi(N(f)))\mu(df)\}^{1/2}$$

where μ is a given probability measure on F_1, assuming that φ is error measurable (i.e. $\operatorname{dist}^2(f,\varphi(N(f)))$ is μ measurable). Of course, the same issues arise as in Section 3 but all of them can be dealt with in a similar way under some additional assumptions. For example, if F_1 is a separable metric space, $N: F_1 \to N(F_1) = H$ is measurable and H is a separable metric space then the measures μ_1 and $\mu_2(\cdot|y,N)$ exist (see [2, Th. 8.1 p. 147]). If additionally, F_2 is a separable metric space, $\operatorname{dist}^2(\cdot,g)$ is measurable for every $g \in F_2$ and $\operatorname{dist}^2(f,\cdot)$ is continuous for almost every $f \in F_1$, then the squared local radius $r^{avg}(N,\cdot)^2$ is μ_1 measurable, optimality is equivalent to strong optimality etc. Since for every countable set F there exists a metric under which F_1 is separable, all discrete problems satisfy the above assumptions.

We end this section by an example for which dist(f,g) cannot be defined by any norm or even any metric.

<u>Function minimum problem:</u> Suppose F_1 is a Hilbert space of continuous functions, $f: [0,1] \to \mathbb{R}$. In addition assume that F_1 is equipped with a reproducing kernel. This means that for every $x \in [0,1]$ there exists a function $\xi_x = \xi_x(\cdot) \in F_1$ such that

$$(5.1) \qquad f(x) = (f, \xi_x), \qquad \forall f \in F_1.$$

Consider now the following problem. Given $y = N(f)$, construct $\phi(y) \in [0,1]$ such that

$|f(\phi(y))|$ is small on the average.

This problem cannot be defined as in Section 3 since $|f(\phi(y))| \neq \|S(f) - \phi(y)\|$ for any operator S. However letting

$$\mathrm{dist}(f,g) = |f(g)|, \quad \forall f \in F_1, \quad \forall g \in F_2 = [0,1],$$

we have

$$e^{avg}(\phi,N) = \{\int_{F_1} \mathrm{dist}^2(f, \phi(N(f)))\mu(df)\}^{1/2}$$
$$= \{\int_{F_1} |f(\phi(N(f)))|^2 \mu(df)\}^{1/2}.$$

Let $\mu_2(\cdot|y,N)$ be the conditional measure. Then

$$e^{avg}(\phi,N,y) = \{\int_{F_1} |f(\phi(y))|^2 \mu_2(df|y,N)\}^{1/2}$$

and

$$r^{avg}(N,y) = \{\inf_{x\in[0,1]} \int_{F_1} |f(x)|^2 \mu_2(df|y,N)\}^{1/2}.$$

Since $f(x) = (f,\xi_x)$ then the squared local radius $r^{avg}(N,\cdot)^2$ is μ_1 measurable. Hence

$$r^{avg}(N) = \{\int_{R^n} r^{avg}(N,y)^2 \mu_1(dy)\}^{1/2}$$

is well defined and

$$r^{avg}(N) = \inf_{\varphi} e^{avg}(\varphi, N).$$

We now calculate $r^{avg}(N,y)^2$.

$$(5.2) \qquad r^{avg}(N,y)^2 = \inf_{x\in[0,1]} \int_{F_1} |f(x)|^2 \mu_2(df|y,N)$$

$$= \inf_{x\in[0,1]} \int_{F_1} (f,\xi_x)^2 \mu_2(df|y,N)$$

$$= \inf_{x\in[0,1]} \{(S_{N,y}\xi_x,\xi_x) + (m_{N,y},\xi_x)^2\}$$

where $S_{N,y}$ is the correlation operator and $m_{N,y}$ is the mean element of $\mu_2(\cdot|y,N)$.

Suppose now that μ is orthogonally invariant and that N is linear. Without loss of generality we can assume that $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$ where $(S_\mu \eta_i,\eta_j) = \delta_{ij}$, and due to Theorem 4.1 we have

$$(5.3) \qquad r^{avg}(N,y)^2 = \inf_{x\in[0,1]} \{h(y)[(S_\mu \xi_x,\xi_x) - \Sigma_{i=1}^{n}(S_\mu \xi_x,\eta_i)^2] + (\sigma(y,N),\xi_x)^2\}$$

$$= \inf_{x\in[0,1]} \{h(y)\Sigma_{i=n+1}^{\infty}((S_\mu \eta_i)(x))^2 + (\sigma(y,N)(x))^2\}$$

where S_μ is the covariance operator of μ and $\sigma(y,N) = \sigma(y,N)(\cdot) \in F_1$, as always, denotes the spline element interpolating y with respect to N. ∎

5. Concluding Remarks.

As we mentioned in the Introduction, all results reported in this paper are primarily of theoretical interest. They will be applied to a variety of problems some of them we discuss now.

(i) Adaption Versus Nonadaption: In this paper (specially in Sections 4 and 5) we assumed that N is nonadaptive, i.e., $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$ where η_i are chosen a priori. A very important generalization is adaptive information where η_i depends on previously computed information $[(f,\eta_1),\ldots,(f,\eta_{i-1})]$. In a recent paper [8] it is proven that adaptive information is not more powerful than nonadaptive assuming that S is linear and μ is orthogonally invariant. The concept of local average radius studied here enables us to generalize this result for S not necessarily linear and μ not necessarily orthogonally invariant.

(ii) Asymptotic-Probabilistic Case Model: In this paper we considered the following approach. Given N, $S(f)$ is approximated by $\phi(N(f))$, $\forall f \in F_1$. Hence N is fixed and independent of f. In practice however, we use very often a different approach which can be characterized as follows: given sequences

$\{N_n\}$ and $\{\varphi_n\}$, we approximate $S(f)$ by $\varphi_n(N_n(f))$ where the index $n = n(f)$ is choosen depending on some termination procedure T. In the asymptotic-probabilistic model we want to find $\{N_n^*\}, \{\varphi_n^*\}$ and T^* such that with a large probability $\varphi_n^*(N_n^*(f))$ approximates $S(f)$ with a small error and the cost of evaluating $\varphi_n^*(N_n^*(f))$ is minimal.

<u>(iii) Stochastic Information</u>: In this paper we assumed N to be <u>exact</u>, i.e., given f, we know $y = N(f)$ exactly. In practice we often have a different situation. Instead of $y = N(f)$ we know $z = y + \varepsilon$ where the error ε is a random variable depending on y. We will study such information using the results reported here.

Appendix.

We prove Theorem 4.1. We begin with

Lemma A.1: Let $N_1(f) = [(f,\zeta_1),\ldots,(f,\zeta_n)]$, $N_2(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$ where $(S_\mu\zeta_i,\zeta_j) = (S_\mu\eta_i,\eta_j) = \delta_{ij}$. Then there exists a linear one-to-one mapping Q, $Q: F_1 \to F_1$, such that

$$(A.1) \qquad N_1 = N_2 Q,$$

$$(A.2) \qquad \mu(Q^{-1}B) = \mu(B), \qquad \forall B \in \mathbf{B}(F_1),$$

$$(A.3) \qquad \mu_2(QB|y,N_2) = \mu_2(B|y,N_1), \quad \forall B \in \mathbf{B}(F_1), \forall y \in \mathbb{R}^n \text{(a.e.)}. \blacksquare$$

Proof: Let $X = \text{lin}\{S_\mu^{1/2}\zeta_1,\ldots,S_\mu^{1/2}\zeta_n,S_\mu^{1/2}\eta_1,\ldots,S_\mu^{1/2}\eta_n\}$. Let $p = \dim X$. Of course $p \in [n,2n]$. There exist elements $\zeta_{n+1},\ldots,\zeta_p,\eta_{n+1},\ldots,\eta_p \in F_1$ so that $\{S_\mu^{1/2}\eta_i\}_{i=1}^p$ and $\{S_\mu^{1/2}\zeta_i\}_{i=1}^p$ are orthonormal basises of X. Define the mapping $H: F_1 \to F_1$,

$$Hf = \Sigma_{i=1}^p (f,S_\mu(\eta_i + \zeta_i))\zeta_i - f.$$

Since $S_\mu^{1/2}\eta_k = \Sigma_{i=1}^p (S_\mu^{1/2}\eta_k,S_\mu^{1/2}\zeta_i)S_\mu^{1/2}\zeta_i$, we get $\eta_k = \Sigma_{i=1}^p(\eta_k,S_\mu\zeta_i)\zeta_i$ and

$$(A.4) \qquad H\eta_k = \Sigma_{i=1}^p(\eta_k,S_\mu\eta_i)\zeta_i + \Sigma_{i=1}^p(\eta_k,S_\mu\zeta_i)\zeta_i - \eta_k = \zeta_k$$

for $k = 1,2,\ldots,p$. We define the mapping Q as

$$Qf = H^*f = \Sigma_{i=1}^{p}(f,\zeta_i)S_\mu(\eta_i + \zeta_i) - f.$$

To prove (A.1) note that $N_1 = N_2Q$ is equivalent to $(f,\zeta_k) = (Qf,\eta_k) = (f,Q^*\eta_k) = (f,H\eta_k)$. This holds since $H\eta_k = \zeta_k$ (see (A.4)).

To prove (ii) we decompose H as

$$H = S_\mu^{-1/2}H_1S_\mu^{1/2}$$

where $H_1f = \Sigma_{i=1}^{p}(f,S_\mu^{1/2}(\eta_i + \zeta_i))S_\mu^{1/2}\zeta_i - f$. Note that $H_1S_\mu^{1/2}(F_1) \subset S_\mu^{1/2}(F_1)$ and therefore $S_\mu^{-1/2}(H_1S_\mu^{1/2})$ is well defined. Let $X^\perp$ be an orthogonal complement of X, $F_1 = X \oplus X^\perp$. Then $f \in X^\perp$ implies $(f,S_\mu^{1/2}\eta_i) = (f,S_\mu^{1/2}\zeta_i) = 0$ and

(A.5) $\qquad H_1f = -f, \qquad \forall f \in X^\perp.$

From (A.3) we have

$$H_1S_\mu^{1/2}\eta_k = S_\mu^{1/2}\zeta_k, \qquad k = 1,2,\ldots,p.$$

Thus H_1 as well as $-H_1$ restricted to X are orthogonal mappings onto X. We decompose $-H_1$ in X using a Householder transformation, i.e., there exist elements $x_i \in X$ such that $x_i = 0$ or $\|x_i\| = 1$ and

(A.6) $\qquad -H_1f = D_1D_2\cdot\ldots\cdot D_pf, \qquad \forall f \in X,$

where $D_i = I - 2x_i \otimes x_i$. Here $x \otimes y$ denotes the linear operator such that $(x \otimes y)(f) = (y,f)x$. Since $(f,x_i) = 0$ for $f \in X^{\perp}$, we get $D_1 D_2 \cdot \ldots \cdot D_p f = f$. Thus, (A.6) holds also for $f \in X^{\perp}$ due to (A.5). Hence we proved that $H_1 = -D_1 D_2 \cdot \ldots \cdot D_p$ and

$$H = -S_{\mu}^{-1/2} D_1 D_2 \cdot \ldots \cdot D_p S_{\mu}^{1/2}$$

$$= -(S_{\mu}^{-1/2} D_1 S_{\mu}^{1/2}) \cdot \ldots \cdot (S_{\mu}^{-1/2} D_p S_{\mu}^{1/2})$$

$$= -Q_1^* Q_2^* \ldots Q_p^*$$

where $Q_i^* = I - 2h_i \otimes S_{\mu} h_i$ and $h_i = S_{\mu}^{-1/2} x_i$. Observe that $Q_i = I - 2S_{\mu} h_i \otimes h_i$. Thus we get

$$Q = -Q_p Q_{p-1} \cdot \ldots \cdot Q_1.$$

Note that $Q_i^{-1} = Q_i$. Thus Q is one-to-one and

$$Q^{-1} = -Q_1 Q_2 \ldots Q_p.$$

The orthogonal invariance of μ yields $\mu(Q_i B) = \mu(B) = \mu(-B)$ for any Borel set B of F_1. We have therefore

$$\mu(Q^{-1}B) = \mu(-Q_1 \cdot \ldots \cdot Q_p B) = \mu(Q_1 \cdot \ldots \cdot Q_p B) = \mu(Q_i \cdot \ldots \cdot Q_p B)$$

$$= \ldots = \mu(B)$$

which proves (A.2).

To prove (A.3) take $\bar{\mu}_2(\cdot|y)$ defined by

$$\bar{\mu}_2(B|y) = \mu_2(Q^{-1}B|y,N_1), \qquad \forall B \in \mathcal{B}(F_1).$$

Then $\bar{\mu}_2(V(N_2,y)|y) = \mu_2(V(N_1,y)|y,N_1) = 1$, $\forall y \in \mathbb{R}^n$(a.e.), $\bar{\mu}_2(B|\cdot)$ is measurable, and

$$\mu(B) = \mu(Q^{-1}B) = \int_{\mathbb{R}^n} \mu_2(Q^{-1}B|y,N_1)\mu_1(dy)$$

$$= \int_{\mathbb{R}^n} \bar{\mu}_2(B|y)\mu_1(dy).$$

Hence $\bar{\mu}_2(\cdot|y)$ is also a conditional measure and the uniqueness of $\mu_2(\cdot|y,N_2)$ implies that $\bar{\mu}_2(\cdot|y) = \mu_2(\cdot|y,N_2)$ for almost every y. This yields

$$\mu_2(QB|y,N_2) = \mu_2(B|y,N_1)$$

which proves (A.3) and completes the proof of Lemma A.1. ∎

<u>Proof of Theorem 4.1 (i)</u>: Let $A \in \mathcal{B}(\mathbb{R}^n)$. Then (A.1) yields that $N_1^{-1}(A) = Q^{-1}N_2^{-1}(A)$. From the definition of μ and (A.2) we get

$$\mu_1(A,N_1) = \mu(N_1^{-1}(A)) = \mu(Q^{-1}N_2^{-1}(A)) = \mu(N_2^{-1}(A)) = \mu_1(A,N_2).$$

This completes the proof of Theorem 4.1 (i). ∎

To prove the remaining parts of Theorem 4.1 we need the following:

<u>Lemma A.2</u>: Let $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$, $(S_\mu\eta_i,\eta_j) = \delta_{ij}$. Let D be a linear continuous mapping, $D\colon F_1 \to F_2$, such that

$$D = D^{-1},\ ND = N,\ \mu(B) = \mu(DB),\quad \forall B \in \mathcal{B}(F_1).$$

Then the conditional measure $\mu_2(\cdot|y,N)$ is D-invariant almost everywhere, i.e., there exists a set $A = A(D) \subset \mathbb{R}^n$ such that $\mu_1(A) = 1$ and

$$\mu_2(B|y,N) = \mu_2(DB|y,N),\quad \forall B \in \mathcal{B}(F_1),\ \forall y \in A. \quad \blacksquare$$

Proof: Let

$$\bar{\mu}_2(B|y,N) = \mu_2(DB|y,N) \quad \forall B \in \mathcal{B}(F_1).$$

The measure $\bar{\mu}_2(\cdot|y,N)$ is well defined since DB is measurable set. From $D = D^{-1}$ and $ND = N$ we have $V(N,y) = DV(N,y)$. Thus

(A.7) $$\bar{\mu}_2(V(N,y)|y,N) = \mu_2(V(N,y)|y,N) = 1,\ \forall y \in \mathbb{R}^n \text{(a.e.)}.$$

Since $\mu_2(DB|\cdot,N)$ is μ_1-measurable, we get

(A.8) $\bar{\mu}_2(B|\cdot,N)$ is μ_1-measurable, $\forall B \in \mathcal{B}(F_1)$.

Take $B \in \mathcal{B}(F_1)$. Since μ is D-invariant,

(A.9) $$\mu(B) = \mu(DB) = \int_{\mathbb{R}^n} \mu_2(DB|y,N)\mu_1(dy) = \int_{\mathbb{R}^n} \bar{\mu}_2(B|y,N)\mu_1(dy).$$

Thus $\bar{\mu}_2(\cdot|y,N)$ is also a conditional measure of μ with respect to N. Since a conditional measure is determined uniquely (up to a set of μ_1-measure zero), we get

$$\bar{\mu}_2(\cdot|y,N) = \mu_2(\cdot|y,N),\quad \forall y \in \mathbb{R}^n \text{(a.e.)}.$$

Thus there exists a set A dependent on the mapping D, such that $\mu_1(A) = 1$ and

$$\mu_2(B|y,N) = \mu_2(DB|y,N), \quad \forall B \in \mathcal{B}(F_1), \forall y \in A.$$

This completes the proof of Lemma A.2.

Proof of Theorem 4.1 (ii): Let $m_{N,y}$ be the mean element of $\mu_2(\cdot|y,N)$. Then for every $g \in F_1$

$$(m_{N,y}, g) = \int_{V(N,y)} (f,g)\mu_2(df|y,N).$$

Take $Df = 2\sum_{i=1}^{n}(f,\eta_i)S_\mu\eta_i - f$. From Lemma A.2, $\mu_2(\cdot|y,N)$ is D-invariant for almost every y and therefore

$$(m_{N,y}, g) = \int_{V(N,y)} (D_N f,g)\mu_2(df|y,N)$$

$$= 2(\sigma(y,N),g) - \int_{V(N,y)} (f,g)\mu_2(df|y,N)$$

$$= 2(\sigma(y,N),g) - (m_{N,y},g).$$

Since g is arbitrary, $m_{N,y} = \sigma(y,N)$, $\forall y \in \mathbb{R}^n$(a.e.), which completes the proof of part (ii). ∎

We now prove the last part of Theorem 4.1.

Proof of Theorem 4.1 (iii): Consider first the information operator N, $N(f) = [(f,\eta_1),\ldots,(f,\eta_n)]$. Let $\eta_1,\ldots,\eta_n,\eta_{n+1},\ldots$

be a basis of F_1 such that $(S_\mu \eta_i, \eta_j) = \delta_{ij}$. We now calculate the values $\alpha_{i,j} = \alpha_{i,j}(y,N) = (S_{N,y}\eta_i, \eta_j)$, $i,j = 1,2,\ldots$. Of course, $\alpha_{i,j} = \alpha_{j,i}$ and, due to Theorem 4.1 (ii),

$$\alpha_{i,j} = \int_{V(N,y)} (f-\sigma(y,N), \eta_i)(f-\sigma(y,N), \eta_j)\mu_2(df|y,N),$$

$$\forall y \in \mathbb{R}^n (\text{a.e.}).$$

Suppose now that i is not greater than n. Then for every $f \in V(N,y)$, $(f-\sigma(y,N), \eta_i) = y_i - y_i = 0$ which means that

$$\alpha_{i,j} = \alpha_{j,i} = 0, \ \forall y \in \mathbb{R}^n (\text{a.e.}), \quad \text{if} \quad i \leq n.$$

Suppose therefore that $i,j > n$. Since now $(\sigma(y,N), \eta_i) = (\sigma(y,N), \eta_j) = 0$,

$$\alpha_{i,j} = \int_{V(N,y)} (f, \eta_i)(f, \eta_j)\mu_2(df|y,N), \ \forall y \in \mathbb{R}^n (\text{a.e.}),$$

$$\forall i,j > n.$$

Consider first $i \neq j$. Then for D, $Df = f - 2(f,\eta_i)S_\mu \eta_i$, Lemma A.2 is applicable and

$$\alpha_{i,j} = \int_{V(N,y)} (f, \eta_i)(f, \eta_j)\mu_2(df|y,N)$$

$$= \int_{V(N,y)} (Df, \eta_i)(Df, \eta_j)\mu_2(df|y,N)$$

$$= \int_{V(N,y)} \{(f, \eta_i)(f, \eta_j) - 2(f, \eta_i)(f, \eta_j)\}\mu_2(df|y,N)$$

$$= -\alpha_{i,j}$$

since $(S_\mu \eta_i, \eta_i) = 1$ and $(S_\mu \eta_i, \eta_j) = 0$. This means that

$$\alpha_{i,j} = 0, \quad \forall y \in \mathbb{R}^n (a.e.), \quad \forall i \neq j.$$

Hence $\alpha_{i,j}$ may be different from zero only if $i = j > n$. Let $\alpha = \alpha(y,N) \overset{df}{=} \alpha_{n+1,n+1}(y,N)$. We now prove that $\alpha_{i,j}(y,N) = \alpha(y,N)$, $\forall y \in \mathbb{R}^n (a.e.), \forall i > n$. Indeed, take $j > n+1$ and $Df = f - 2(f,(\eta_{n+1} + \eta_j)/\sqrt{2})S_\mu((\eta_{n+1} + \eta_j)/\sqrt{2})$. Observe that $(f,\eta_j)^2 = (Df,\eta_{n+1})^2$ and that D satisfies the assumptions of Lemma A.2. Hence

$$\alpha = \int_{V(N,y)} (f,\eta_{n+1})^2 \mu_2(df|y,N) = \int_{V(N,y)} (Df,\eta_{n+1})^2 \mu_2(df|y,N)$$

$$= \int_{V(N,y)} (f,\eta_j)^2 \mu_2(df|y,N) = \alpha_{j,j}$$

as claimed.

Up to now we have proven that for every $i,j = 1,2,\ldots$ there exists a set $A_{i,j}$ of μ_1-measure one such that $(S_{N,y}\eta_i,\eta_j) = \alpha_{i,j}(y,N)$, $\forall y \in A_{i,j}$, where $\alpha_{i,j} \equiv 0$ for $i \leq n$ or $i \neq j$ and $\alpha_{i,i}(y,N) = \alpha(y,N)$ for $i > n$. Since there are at most countably many such sets $A_{i,j}$ we can conclude that there exists a set A such that $\mu_1(A) = 1$ and

$$\text{(A.10)} \qquad (S_{N,y}\eta_i,\eta_j) = \begin{cases} 0 & \text{if } i \leq n \text{ or } i \neq j \\ \alpha(y,N) & \text{if } i = j > n \end{cases} \quad \forall y \in A.$$

Of course, (A.10) defines $S_{N,y}$ uniquely (up to a set of

μ_1-measure zero). Consider now the following operator $K_{N,y}: F_1 \to F_1$,

$$K_{N,y}f = \alpha(y,N)\cdot(I-\sigma_N)S_\mu(I-\sigma_N^*)f, \quad \forall f \in F_1.$$

It is easy to check that for every $y \in \mathbb{R}^n$, $(K_{N,y}\eta_i, \eta_j) = 0$ if $i \leq n$ or $i \neq j$ and $(K_{N,y}\eta_i, \eta_i) = \alpha(y,N)$ if $i > n$. This means that

$$\text{(A.11)} \qquad S_{N,y} = K_{N,y} = \alpha(y,N)\cdot(I-\sigma_N)S_\mu(I-\sigma_N^*), \qquad \forall y \in A.$$

Since $\alpha(y,N) = \int_{F_1} (f,\eta_{n+1})^2\mu_2(df|y,N)$, $\alpha(\cdot,N)$ is μ_1-measurable and

$$\int_{\mathbb{R}^n}\alpha(y,N)\mu_1(dy) = \int_{\mathbb{R}^n}\int_{F_1}(f,\eta_{n+1})^2\mu_2(df|y,N)\mu_1(dy)$$

$$= \int_{F_1} (f,\eta_{n+1})^2\mu(df) = (S_\mu\eta_{n+1},\eta_{n+1}) = 1.$$

To complete the proof we only need to show that the function $\alpha(\cdot,N) = \alpha(\cdot)$ does not depend on N, since letting $h(y) = \alpha(y)$ we shall prove (iii).

To prove this, take two information operators N_1 and N_2 of the form (4.6) with $\text{card}(N_1) = \text{card}(N_2) = n$. Let Q be as in Lemma A.1. Taking $\eta_{n+1} = Q^*\xi_{n+1}$ we have

$$\alpha(y,N_1) = \int_{F_1}(f,\eta_{n+1})^2\mu_2(df|y,N_1)$$

$$= \int_{F_1}(Qf,\xi_{n+1})^2\mu_2(df|y,N_1),$$

and due to (A.3) of Lemma A.1, we get

$$\alpha(y,N_1) = \int_{F_1} (f,\zeta_{n+1})^2 \mu_2(df|y,N_2) = \alpha(y,N_2).$$

This completes the proof of part (iii) as well as the proof of Theorem 4.1. ∎

Acknowledgements.

I am grateful to J.F. Traub and H. Woźniakowski for their advice and valuable comments concerning this paper.

References.

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

[2] Parthasarathy, K.R., Probability Measures on Metric Spaces, Academic Press, New York, 1967.

[3] Skorohod, A.V., Integration in Hilbert Space, Springer-Verlag, New York, 1979.

[4] Traub, J.F., Wasilkowski, G.W. and Woźniakowski, H., "Average Case Optimality for Linear Problems", Dept. of Computer Science Report, Columbia University, 1981.

[5] ____, Information, Uncertainty, Complexity, Addison-Wesley, Reading, Mass., 1983.

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

[7] Wasilkowski, G.W., and Woźniakowski, H., "Average Case Optimal Algorithm in Hilbert Spaces" Dept. of Computer Science Report, Columbia University, 1982.

[8] Woźniakowski, H., "Can Adaption Help on the Average?", Dept. of Computer Science, Columbia University, 1982.