JOURNAL OF APPROXIMATION THEORY **61**, 158–177 (1990)

Orthogonally Invariant Measures and Best Approximation of Linear Operators

NIELS JUUL MUNCH*

Matematisk Institut, Aarhus Universitet, Ny Munkegade, DK-8000 Aarhus C, Denmark

Communicated by T. J. Rivlin

Received July 27, 1987

This paper studies optimal information and optimal algorithms in Hilbert space for an n-dimensional average case model. The error in approximating a linear operator is the average of some error criterion E with respect to an *orthogonally invariant* measure. The orthogonally invariant measures are characterized and the problem of best approximation is solved for a wide range of error criteria E. In addition it is shown that *adaption does not help*.

1. INTRODUCTION

This paper is concerned with the general problem of estimating the action of a linear bounded operator A on a real, separable Hilbert space $\mathscr{H}$ when only finite information is available. Here *information* is provided by a map N from $\mathscr{H}$ into the space $\mathbb{R}^n$ of fixed finite dimension n. Knowing Nf, $f \in \mathscr{H}$, one seeks, the best recovery of Af by means of an *algorithm* φ, that is, a map φ: $\mathbb{R}^n \to \mathscr{H}$. In other words the difference $A - \varphi N$ should be as small as possible in a specified sense.

For a worst case error criterion this setup has been examined in [4, 5, 8–10] and others. Here we relate to an *average error* criterion as in [6–8, 12–16]. Assuming μ to be a Borel probability measure on $\mathscr{H}$ with mean zero and finite second moment $\int_{\mathscr{H}} \|f\|^2 \, d\mu(f)$, the error to be minimized is

$$e(\varphi, N) = \int_{\mathscr{H}} E(Af - \varphi Nf) \, d\mu(f) \tag{1}$$

for some function E: $\mathscr{H} \to \mathbb{R}_+ = [0, \infty[$. The classical choice for E is $E(f) = \|f\|^2$, which constitutes the average squared error. But also the

* Research supported by the Danish Natural Science Research Council.

0021-9045/90 $3.00

probabilistic or hit-and-miss criterion conforms to this framework if we choose $E(f) = 1_{[\varepsilon, \infty[}(\|f\|)$.

It is the striking result of [14] (see also [6, 7, 13, 15, 16]) that for the average squared error and a certain class of "orthogonally invariant" measures μ *adaptive* linear information is not superior to *non-adaptive* linear information. For such μ, possessing a high degree of spatial symmetry, the minimal error $e(\varphi, N)$ can be obtained even within the class of non-adaptive linear information operators N and corresponding linear spline algorithms φ.

Given the implications of this result it becomes of interest to determine its precise range of validity, i.e., to determine which measures μ are orthogonally invariant. Examples from [14] are Gaussian measures and, in finite dimensions, measures μ absolutely continuous with respect to Lebesgue measure m so that the Radon–Nikodym derivative $d\mu/dm$ is rotation invariant in a suitably perturbed inner product. It turns out that these examples are quite characteristic. Denote by C the covariance operator of μ and assume that the range of C is infinite dimensional. In that case μ is orthogonally invariant if and only if it has a representation

$$\mu = \int_0^\infty \mu_{tC} \, d\nu(t), \tag{2}$$

where μ_K is the Gaussian measure on $\mathscr{H}$ with mean zero and covariance operator K and ν is a Borel probability measure on $\mathbb{R}_+$. Further the representation (2) is uniquely determined under the condition $1 = \int_0^\infty t \, d\nu(t)$ in which case we denote the measure μ by μ_C^ν. One may note from (2) that the projection of μ_C^ν onto any finite dimensional subspace of $\mathscr{H}$ invariably is absolutely continuous w.r.t. Lebesgue measure.

The error criterion (1) has a simple probabilistic interpretation. When X is a second order random variable taking values in $\mathscr{H}$ and μ is the induced distribution on $\mathscr{H}$ the error $e(\varphi, N)$ is the expected value of $E((A - \varphi N)X)$. Of course Gaussian measures μ arise from Gaussian variables X. But suppose that X is in fact an instance, say X_1, of a stochastic process $\{X_t\}_{t \geqslant 0}$, where each X_t has the distribution μ_{tC}. If the observation time T is subject to noise it may differ from its nominal value $T = 1$, transforming the variable $X = X_1$ into $X = X_T$, where $X_T(\omega) = (X_{T(\omega)})(\omega)$ for each outcome ω. Under the hypothesis of statistical independence of $\{X_t\}_{t \geqslant 0}$ and T, the distribution of $X = X_T$ is μ_C^ν, where ν is the distribution of T. Thus even starting in a purely Gaussian setting it is possible to arrive naturally at measures of the type μ_C^ν.

In Section 2 of this paper we prove the structure theorem (2) and derive some ancillary properties of orthogonally invariant measures. As I have recently become aware it appears from a remark in [16, p. 363] that the connection between orthogonally invariant measures and the "elliptically contoured" measures [3] of the form (2) has been noted previously by Kwapien in private communication. However, since the present approach is quite different from the one in [3] and the results are sharper, I feel it is justified to present this material.

Next, in Section 3, we study the approximation problem for orthogonally invariant measures. For a quite general class of error criteria E it is proved that adaptive information N is not more powerful than non-adaptive information and that for a given non-adaptive N the natural spline algorithm is optimal. In particular there is a linear optimal algorithm. The set of E's considered includes the average squared error and the probabilistic error. Consequently our approach unifies and improves previous results for orthogonally invariant measures and the squared error [12, 14, 15] and Gaussian measures and more general criteria [13, 16, (6), (7)]. In addition a number of new results and uniqueness results are obtained.

2. Orthogonally Invariant Measures

Let $\mathscr{H}$ be a real Hilbert space of finite or countable dimension. Consider on $\mathscr{H}$ a Borel probability measure μ with mean zero, finite second moment $\int_{\mathscr{H}} \|f\|^2 \, d\mu(f)$, and covariance operator C_μ defined by

$$C_\mu = \int_{\mathscr{H}} (f \otimes f) \, d\mu(f).$$

Here the Hilbert–Schmidt operators on $\mathscr{H}$ are identified with the tensor product $\mathscr{H} \otimes \mathscr{H}$ through $(f_1 \otimes f_2) g = (g, f_1) f_2$. It is assumed that C_μ is injective and that μ is symmetric, i.e., $\int_{\mathscr{H}} F(f) \, d\mu(f) = \int_{\mathscr{H}} F(-f) \, d\mu(f)$.

Following Wasilkowski and Wozniakowski [14] we define the symmetric measure μ to be *orthogonally invariant* if $\mu = \mu \circ Q_f^{-1}$ for all $f \in \mathscr{H}$ normalized so that $(C_\mu f, f) = 1$. Here Q_f is the operator $Q_f = 2(f \otimes C_\mu f) - I$ which satisfies $Q_f^2 = I$ provided $(C_\mu f, f) = 1$.

Recall that the Fourier transform or characteristic functional $\hat{\mu}$ of μ is the function from $\mathscr{H}$ into $\mathbb{C}$ defined by

$$\hat{\mu}(f) = \int_{\mathscr{H}} \exp(i(g, f)) \, d\mu(g), \qquad f \in \mathscr{H},$$

and that $\hat{\mu}$ determines μ uniquely [11, pp. 11]. For any non-zero vector g in $\mathscr{H}$ denote by l_g the functional $l_g(f) = (C_\mu g, g)^{-1/2} (f, g)$, $f \in \mathscr{H}$.

Proposition 2.1. *For a symmetric measure μ the following are equivalent.*

(a) *The measure μ is orthogonally invariant.*

(b) *All measures $\mu \circ l_g^{-1}$, $g \in \mathscr{H} \setminus \{0\}$, are equal.*

(c) *There is a function g: $\mathbb{R}_+ \to \mathbb{R}$ so that*

$$\hat{\mu}(f) = g((C_\mu f, f)^{1/2}), \qquad f \in \mathscr{H}.$$

(d) *There is a twice continuously differentiable, positive definite function g_μ: $\mathbb{R} \to \mathbb{R}$ such that*

$$\hat{\mu}(f) = g_\mu((C_\mu f, f)^{1/2}), \qquad f \in \mathscr{H}$$

and

$$g_\mu(0) = 1, \qquad g'_\mu(0) = 0, \qquad g''_\mu(0) = -1.$$

Further if μ is orthogonally invariant and if for some function g: $\mathbb{R}_+ \to \mathbb{R}$ and self-adjoint operator C it holds that $\hat{\mu}(f) = g((Cf, f)^{1/2})$, $f \in \mathscr{H}$, then there is a constant $\gamma > 0$ such that $C_\mu = \gamma^2 C$ and $g(s) = g_\mu(\gamma s)$, $s \geqslant 0$.

Proof. (a) $\Rightarrow$ (c). Assume $(C_\mu f_1, f_1) = (C_\mu f_2, f_2)$. Let g be the vector $g = (C_\mu(f_1 + f_2), f_1 + f_2)^{-1/2} (f_1 + f_2)$. As $(C_\mu(f_1 + f_2), f_1 + f_2) = 2((C_\mu f_1, f_1) + (C_\mu f_1, f_2))$ one may verify that $Q_g^* f_1 = f_2$. Consequently $\hat{\mu}(f_1) = \widehat{\mu \circ Q_g^{-1}}(f_1) = \hat{\mu}(Q_g^* f_1) = \hat{\mu}(f_2)$.

(c) $\Rightarrow$ (a). It is straightforward to verify the relation $Q_f C_\mu Q_f^* = C_\mu$. Thus

$$\begin{aligned} \widehat{\mu \circ Q_f^{-1}}(g) &= \hat{\mu}(Q_f^* g) \\ &= g((C_\mu Q_f^* g, Q_f^* g)^{1/2}) \\ &= g((C_\mu g, g)^{1/2}) \\ &= \hat{\mu}(g), \qquad g \in \mathscr{H}. \end{aligned}$$

(c) $\Leftrightarrow$ (b). This equivalence is seen from

$$\widehat{\mu \circ l_g^{-1}}(s) = \hat{\mu}(s(C_\mu g, g)^{-1/2} g), \qquad s \in \mathbb{R}. \tag{3}$$

(c) $\Rightarrow$ (d). Denote the common value of $\mu \circ l_g^{-1}$ by $\bar{\mu}$. It is apparent from (3) that $\hat{\bar{\mu}}(s) = g(s)$, $s \geqslant 0$. Now (d) is simply the statement that $g_\mu = \hat{\bar{\mu}}$ is the transform of a probability measure with mean zero and second moment one.

To prove the final statement of the proposition assume that

$$\hat{\mu}(f) = g_\mu((C_\mu f, f)^{1/2}) = g((Cf, f)^{1/2}), \qquad f \in \mathscr{H}.$$

Consider any non-zero vector g in $\mathscr{H}$ and put $\alpha^2 = (C_\mu g, g)$, $\beta^2 = (Cg, g)$. Then $\hat{\mu}(sg) = g_\mu(|s| \alpha) = g(|s| \beta)$. If $\beta = 0$ then $\hat{\mu}(\mathbb{R} \cdot g) = \{1\}$ and μ is concentrated on the orthogonal complement of g, contradicting the standing assumption that C_μ is injective. Thus $g(s) = g_\mu(\gamma s)$, $s \geqslant 0$, holds with $\gamma = \alpha/\beta$. Since the identity $g_\mu(\gamma \cdot) = g$ can be true for at most one value of γ it follows that

$$(C_\mu g, g) = (\gamma^2 Cg, g), \qquad g \in \mathscr{H}.$$

The equality $C_\mu = \gamma^2 C$ is seen by polarization. ∎

As stated in the introduction we denote by μ_C the Gaussian measure on $\mathscr{H}$ with mean zero and covariance operator C. Similarly μ_C^ν denotes the measure given by

$$\mu_C^\nu(\mathscr{B}) = \int_0^\infty \mu_{tC}(\mathscr{B})\, d\nu(t)$$

for all Borel sets $\mathscr{B}$.

THEOREM 2.2. *Let μ be an orthogonally invariant measure on an infinite dimensional, separable real Hilbert space $\mathscr{H}$. Then there is a Borel probability measure ν on $\mathbb{R}_+$ with $1 = \int_0^\infty t\, d\nu(t)$ and positive nuclear operator $C = C_\mu$ such that $\mu = \mu_C^\nu$. The pair (C, ν) is unique.*

Proof. By the proposition we can express $\hat{\mu}$ as $\hat{\mu}(f) = g_\mu((C_\mu f, f)^{1/2})$, $f \in \mathscr{H}$. Since g_μ is continuous, $\hat{\mu}$ is positive definite, and C_μ has dense range it follows that the function $g_\mu(\|f\|)$ is positive definite on $\mathscr{H}$. Hence by a famous theorem of Schoenberg [2, p. 152] the function $g_\mu(\sqrt{t})$ for $t \geqslant 0$ is the Laplace transform $\mathscr{L}\nu$ of a Borel probability measure ν on $\mathbb{R}_+$. For convenience we express this as

$$g_\mu(t) = \int_0^\infty \exp(-t^2 s/2)\, d\nu(s), \qquad t \geqslant 0. \tag{4}$$

In turn (4) implies

$$\hat{\mu}(f) = \int_0^\infty \exp(-\tfrac{1}{2}(C_\mu f, f)t)\, d\nu(t)$$

$$= \int_0^\infty \widehat{\mu_{tC_\mu}}(f)\, d\nu(t), \qquad f \in \mathscr{H}.$$

Hence $\hat{\mu}$ equals the transform of the well defined mixed measure $\mu_{C_\mu}^\nu$ and the two must be equal. Since

$$C_\mu = \int_0^\infty (tC_\mu)\, d\nu(t)$$

clearly $1 = \int_0^\infty t\, d\nu(t)$.

If $\mu = \mu_{\bar{C}}^{\bar{\nu}}$ is some other representation one finds

$$\hat{\mu}(f) = g_\mu((C_\mu f, f)^{1/2}) = \bar{g}((\bar{C}f, f)^{1/2}),$$

where $\bar{g}(\sqrt{2t}) = (\mathscr{L}\bar{\nu})(t)$ and $g_\mu(\sqrt{2t}) = (\mathscr{L}\nu)(t)$, $t \geq 0$. The desired identification $(\bar{\nu}, \bar{C}) = (\nu, C_\mu)$ follows from combining $\bar{g}''(0) = -\int_0^\infty t\, d\bar{\nu}(t) = -1$ with the proposition above and the injectivity of the Laplace transform. ∎

It is apparent that the projection $\mu \circ p^{-1}$ of $\mu = \mu_C^\nu$ onto a finite dimensional subspace of $\mathscr{H}$ is absolutely continuous w.r.t. Lebesgue measure m with a Radon–Nikodym derivative $d(\mu \circ p^{-1})/dm$ which is $\mathscr{C}^\infty$ outside zero. If ν vanishes in a neighbourhood of zero $d(\mu \circ p^{-1})/dm$ even belongs to the Schwartz space $\mathscr{S}$. In contrast, for any finite dimension, normalized integration over the boundary of the unit ball is an orthogonally invariant measure singular w.r.t. Lebesgue measure. In dimension one this is $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$ with transform $\hat{\mu}(t) = \cos(t) = \cos((t^2)^{1/2})$ which is not even positive.

The two corollaries to Theorem 2.2 and Proposition 2.1 demonstrate that the Gaussian measures have properties which are quite distinct from those of a general orthogonally invariant measure.

COROLLARY 2.3. *If an orthogonally invariant measure μ is a product measure with respect to a non-trivial orthogonal decomposition $\mathscr{H} = \mathscr{H}_1 \oplus \mathscr{H}_2$, then μ is a Gaussian measure.*

Proof. Choose non-zero vectors f_i in $\mathscr{H}_i$ and put $\beta_{ij} = (C_\mu f_i, f_j)$. Denote by G_μ the function $G_\mu(t) = g_\mu(\sqrt{t})$, $t \geq 0$, which by l'Hospital's rule satisfies

$$\lim_{t \to 0^+} G'_\mu(t) = -\tfrac{1}{2}. \tag{5}$$

By hypothesis $\hat{\mu}(\lambda_1 f_1 + \lambda_2 f_2) = \hat{\mu}(\lambda_1 f_1)\, \hat{\mu}(\lambda_2 f_2)$. Consequently, as β_{12} is readily shown to be zero,

$$G_\mu(\lambda_1^2 \beta_{11} + \lambda_2^2 \beta_{22}) = G_\mu(\lambda_1^2 \beta_{11})\, G_\mu(\lambda_2^2 \beta_{22}), \qquad \lambda_i \in \mathbb{R}. \tag{6}$$

In combination with (5) the functional equation (6) implies $G_\mu(t) = \exp(-t/2)$, $t \geq 0$. Thus $\mu = \mu_{C_\mu}$. ∎

For any measure λ (on $\mathbb{R}_+$) and positive real number α denote by λ^α the measure

$$\int F(x)\, d\lambda^\alpha(x) = \int F(\alpha x)\, d\lambda(x).$$

COROLLARY 2.4. *Let μ be the convolution measure $\mu = \mu_1 * \mu_2$, where $\mu_i = \mu_{C_i}^{\nu_i}$.*

Then μ is orthogonally invariant only if either μ_i are both Gaussian or the covariances C_i are proportional.

Proof. The proof is based on (b) of Proposition 2.1. Since $C_\mu = C_1 + C_2$ is known we may set out to determine when $\mu \circ l_g^{-1}$, $g \in \mathscr{H} \backslash \{0\}$, are all equal. Notation will be as in the proof of Proposition 2.1.

Now $\mu \circ l_g^{-1} = (\mu_1 \circ l_g^{-1}) * (\mu_2 \circ l_g^{-1})$ and for

$$a = (C_1 g, g)^{1/2} ((C_1 + C_2) g, g)^{-1/2}$$

one finds that

$$\begin{aligned}\widehat{\mu \circ l_g^{-1}(s)} &= \widehat{(\mu_1 \circ l_g^{-1})}(s) \cdot \widehat{(\mu_2 \circ l_g^{-1})}(s)\\ &= \widehat{(\bar{\mu}_1)^a}(s) \cdot \widehat{(\bar{\mu}_2)^{(1-a^2)^{1/2}}}(s)\\ &= \hat{\bar{\mu}}_1(as)\, \hat{\bar{\mu}}_2((1-a^2)^{1/2} s)\end{aligned}$$

Thus the requirement is that the functions

$$g_a(s) = g_1(as)\, g_2((1-a^2)^{1/2} s) \tag{7}$$

should be independent of the parameter a as it ranges over the closure $I = K^-$ of the set

$$K = \{ \|C_1^{1/2} g\| \cdot \|(C_1 + C_2)^{1/2} g\|^{-1} \mid g \in \mathscr{H} \backslash \{0\} \}.$$

But I is precisely the set $\{ \|C_1^{1/2} (C_1 + C_2)^{-1/2} f\| \mid \|f\| = 1\}^-$ which in turn is identical to the square root $W^{1/2}$ of the numerical range

$$W = \{ ((C_1 + C_2)^{-1/2} C_1 (C_1 + C_2)^{-1/2} f, f) \mid \|f\| = 1\}^-.$$

In particular, I is an interval.

In case I, and hence W, is a singleton set we find by polarization that the C_i are proportional. Otherwise we may differentiate (7) with respect to a in the interior of I to obtain

$$\begin{aligned}&(1-a^2)^{1/2} g_1'(as)\, g_2((1-a^2)^{1/2} s)\\ &\quad = a g_1(as)\, g_2'((1-a^2)^{1/2} s), \qquad a \in I, s \in \mathbb{R}.\end{aligned}$$

Since g_i are everywhere positive this can be rewritten as

$$\begin{aligned}&(1-a^2) \frac{d}{ds} (\ln g_1(as))\\ &\quad = a^2 \frac{d}{ds} (\ln g_2((1-a^2)^{1/2} s)), \qquad a \in I, s \in \mathbb{R}.\end{aligned}$$

and consequently for any fixed a in I

$$g_1(as)^{(1-a^2)} = g_2((1-a^2)^{1/2}\, s)^{a^2},\ s \in \mathbb{R}. \tag{8}$$

On comparison with (7), with the common value of g_a denoted by $\bar{g}$, this yields

$$g_1(as) = \bar{g}(s)^{a^2}, \qquad a \in I,\ s \in \mathbb{R},$$

and it follows that the function $g_1(t)$ equals the function $\exp(t^2 \ln \bar{g}(1))$ through the interval I. Since $g_1(\sqrt{\ \cdot\ })$ is the Laplace transform of a probability measure it has an analytic continuation to the open right half plane. In turn g_1 has an analytic extension to the interior of a 45° cone symmetrically including the positive real axis. Thus from uniqueness of analytic continuation and the condition $g_1''(0) = -1$ the identity $g_1(t) = e^{-t^2/2}$, $t \in \mathbb{R}_+$, follows. Hence μ_1, and likewise μ_2, are Gaussian. ∎

Remarks. (a) To connect Corollary 2.4 with our introductory considerations regarding random variables X_T, let Z be a sum $Z = X_T + Y_S$ of two independent variables of this kind. Corollary 2.4 states that if the covariance parameters of $(X_t)_{t \geqslant 0}$ and $(Y_s)_{s \geqslant 0}$ are not proportional then Z has an orthogonally invariant distribution only if T and S are constants.

(b) One property, however, characteristic of Gaussian measures is preserved for orthogonally invariant measures $\mu = \mu_C^\nu$. When $\{e_j\}_{j=1}^\infty$ and $\{\lambda_i\}_{j=1}^\infty$ are the eigenvectors and corresponding eigenvalues of C the limit $\lim_{n \to +\infty} (1/n) \sum_{j=1}^n (f, e_j)^2/\lambda_j$ still exists for μ almost all f in $\mathscr{H}$. But it need no longer be equal to one, μ a.e. In fact ν is equal to $\mu \circ \rho^{-1}$ and μ_{tC}, $t \in \mathbb{R}_+$, are the conditional measures for $\{\rho(f) = t\}$, where

$$\rho(f) = \lim_{n \to +\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{(f, e_j)^2}{\lambda_j}$$

can be given any value on the set of non-convergence.

3. Approximation of Linear Operators

This section investigates the approximation of a linear bounded operator A: $\mathscr{H} \to \mathscr{H}$ with respect to some fixed orthogonally invariant measure $\mu = \mu_C^\nu$. First it is necessary to introduce further definitions and notations.

When the Hilbert space $\mathscr{H}$ is identified with its own dual space of functionals, an *adaptive* linear information operator N: $\mathscr{H} \to \mathbb{R}^n$ is any map of the form $Nf = (y_i)_{i=1}^n$, where $y_1 = (f, g_1)$, $y_2 = (f, g_2(y_1))$, ...,

$y_n = (f, g_n(y_1, ..., y_{n-1}))$, and g_i: $\mathbb{R}^{i-1} \to \mathscr{H}$ are measurable functions for $1 \leqslant i \leqslant n$. Thus the ith point of evaluation is allowed to depend (measurably) on the previous $(i-1)$ outcomes. The information operator or just information N is called *non-adaptive* if the g_i are constant functions, i.e., the points of evaluation have been chosen a priori. Given an error functional E the *error* of an algorithm φ is defined by

$$e(\varphi, N) = \int_{\mathscr{H}} E(Af - \varphi Nf)\, d\mu(f)$$

and the *radius* of an information operator N is

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

Without essential loss of generality it is assumed that

$$(Cg_i(y), g_j(y)) = \delta_{ij}$$

holds for almost all y in $\mathbb{R}^n$. For $y = (y_i)_{i=1}^n$ in $\mathbb{R}^n$ of course $g_i(y)$ means $g_i(y_1, ..., y_{i-1})$. Also for y in $\mathbb{R}^n$ we adopt the notation [13]

$$m(y) = \sum_{j=1}^{n} y_j C g_j(y)$$

$$\sigma(y) = \sum_{J=1}^{n} g_j(y) \otimes C g_j(y)$$

and

$$S(y) = (I - \sigma(y))\, C (I - \sigma(y))^*.$$

The measure μ_C^ν is transformed by N into the measure μ_I^ν on $\mathbb{R}^n$. This is readily verified when the g_i constantly equal suitably normalized eigenvectors for C; the general case then follows from [14, Theorem 4.2]. In [13, Theorem 3.1] it is shown that for $\mu = \mu_C$ the conditional measure for $\{Nf = y\}$ is the Gaussian measure $\mu_{m(y), S(y)}$ with mean $m(y)$ and covariance $S(y)$, i.e.,

$$\mu_C = \int_{\mathbb{R}^n} \mu_{m(y), S(y)}\, d\mu_I(y) \tag{9}$$

with each $\mu_{m(y), S(y)}$ supported on $\{f \mid Nf = y\}$. The next proposition determines the corresponding resolution of an orthogonally invariant μ with respect to N.

Denote by W_s the density function $W_s(y) = (2\pi s)^{-n/2} \exp(-\|y\|^2/2s)$, $y \in \mathbb{R}^n$, and by W^ν the Radon–Nikodym derivative

$$W^\nu(y) = \frac{d\mu_I^\nu}{dm}(y) = \int_0^\infty W_s(y)\, d\nu(s).$$

PROPOSITION 3.1. *It holds that*

$$\mu_C^\nu = \int_{\mathbb{R}^n} \mu^y \, d\mu_I^\nu(y), \tag{10}$$

where each probability measure

$$\mu^y = W^\nu(y)^{-1} \int_0^\infty \mu_{m(y), sS(s^{-1/2}y)} W_s(y)\, d\nu(s) \tag{11}$$

is supported on $\{f \mid Nf = y\}$.

Proof. Application of (9) to the covariance operators $\bar{C} = sC$ and the informations $\bar{N}$ given by $\bar{g}_i = s^{-1/2} g_i$ demonstrates that

$$\mu_{sC} = \int_{\mathbb{R}^n} \mu_{m(s^{1/2}y), sS(y)} \, d\mu_I(y)$$

and that each $\mu_{m(s^{1/2}y), sS(y)}$ is supported on $\{f \mid Nf = s^{1/2}y\}$. Thus

$$\mu_C^\nu = \int_0^\infty \int_{\mathbb{R}^n} \mu_{m(s^{1/2}y), sS(y)} \, d\mu_I(y)\, d\nu(s)$$

which after reshuffling, using

$$\int_0^\infty \int_{\mathbb{R}^n} F(s, y)\, d\mu_I(y)\, d\nu(s)$$
$$= \int_{\mathbb{R}^n} \int_0^\infty F(s, s^{-1/2}y)\, W_s(y)\, d\nu(s)\, dy,$$

becomes (10) and (11). ∎

In the sequel the following very general class of error functionals is considered. A measurable function $E\colon \mathscr{H} \to \mathbb{R}_+$ is called an *allowable* error functional if each set

$$\mathscr{B}_t = \{f \in \mathscr{H} \mid E(f) < t\}$$

is convex and balanced. This includes the average squared error and the error in probability. Moreover every convex function $E\colon \mathscr{H} \to \mathbb{R}_+$ with

$E(0)=0$ is allowable. We shall refer to E as a *standard* error function if $\mathscr{B}_t$ has the form $\mathscr{B}_t=F(t)\mathscr{B}$, where $\mathscr{B}$ is a bounded, convex, open set containing zero and F is a continuous bijection of $\mathbb{R}_+$. In this case E is given by $E(f)=G(p_{\mathscr{B}}(f))$ for $G=F^{-1}$ and the continuous Minkowski seminorm $p_{\mathscr{B}}(f)=\inf\{t>0 \mid f\in t\mathscr{B}\}$. The set of standard functionals includes in particular functions of the type $E(f)=H(\|f\|)$.

LEMMA 3.2. *Let E be an allowable error functional and let μ_C be a Gaussian measure. Then the function*

$$\chi(g)=\int_{\mathscr{H}} E(f-g)\,d\mu_C(f)$$

of g in $\mathscr{H}$ attains its minimum value at $g=0$. If $\chi(0)$ is finite and E is a standard functional this minimum is unique.

Proof. The main tool here is the identity

$$\chi(g)=\int_{\mathscr{H}} E(f)\,d\mu_{g,C}(f)=\int_0^\infty t\,d(\mu_{g,C}(\mathscr{B}_t)). \tag{12}$$

Optimality of $g=0$ follows from $\mu_{g,C}(\mathscr{B}_t)\leqslant\mu_C(\mathscr{B}_t)$, $t\in\mathbb{R}_+$, which holds by the hypothesis on $\mathscr{B}_t$, cf. [13, Lemma 3.1] and [1, Theorem 1].

If conversely, $\chi(0)=\chi(g)<+\infty$ then necessarily $\mu_{g,C}(\mathscr{B}_t)=\mu_C(\mathscr{B}_t)$, $t\in\mathbb{R}_+$. For the standard case $\{\mathscr{B}_t\}_{t\geqslant 0}$ is equal to $\{t\mathscr{B}\}_{t\geqslant 0}$. From (12) and the symmetry of μ_C

$$\begin{aligned}\int_{\mathscr{H}} p_{\mathscr{B}}(f)\,d\mu_C(f)&=\int_{\mathscr{H}} p_{\mathscr{B}}(f-g)\,d\mu_C(f)\\&=\int_{\mathscr{H}}\frac{1}{2}(p_{\mathscr{B}}(f-g)+p_{\mathscr{B}}(f+g))\,d\mu_C(f).\end{aligned}$$

Combined with the convexity of $p_{\mathscr{B}}$ this implies

$$2p_{\mathscr{B}}(f)=p_{\mathscr{B}}(f-g)+p_{\mathscr{B}}(f+g), \qquad \mu_C \text{ a.e.} \tag{13}$$

Take a sequence $f_n\to 0$ for which (13) holds. Then by the continuity of $p_{\mathscr{B}}$ ($\mathscr{B}$ open), $p_{\mathscr{B}}(g)=0$, and by the faithfulness of $p_{\mathscr{B}}$ ($\mathscr{B}$ bounded) $g=0$. ∎

When N is non-adaptive the constant values of $S(y)$ and $g_i(y)$ are simply denoted S and g_i.

THEOREM 3.3. *Assume that N is non-adaptive information, $\mu=\mu_C^y$ is an* *orthogonally invariant measure on* $\mathscr{H}$, *and E is an allowable error functional. Then*

(a) *The spline algorithm*

$$\varphi^s\colon (y_i)_{i=1}^n \to \sum_{i=1}^n y_i ACg_i \tag{14}$$

is an optimal algorithm. When $e(\varphi^s, N)$ *is finite and E is a standard functional* φ^s *is a unique optimal algorithm.*

(b) $r(N) = \int_{\mathscr{H}} E(f)\, d\mu^v_{ASA^*}(f)$.

(c) *When E is p-homogeneous, i.e.,* $E(\alpha f) = |\alpha|^p E(f)$,

$$r(N) = \left(\int_0^\infty s^{p/2}\, dv(s)\right) \int_{\mathscr{H}} E(f)\, d\mu_{ASA^*}(f).$$

Proof. (a) Due to Proposition 3.1

$$e(\varphi, N) = \int_{\mathbb{R}^n} \int_{\mathscr{H}} E(Af - \varphi(y))\, d\mu^y(f)\, d\mu^v_I(y), \tag{15}$$

where

$$\begin{aligned}\int_{\mathscr{H}} E(Af - \varphi(y))\, d\mu^y(f) &= W^v(y)^{-1} \int_0^\infty W_s(y) \\ &\quad \times \int_{\mathscr{H}} E(Af - \varphi(y))\, d\mu_{m(y), sS}(f)\, dv(s) \\ &= W^v(y)^{-1} \int_0^\infty W_s(y) \\ &\quad \times \int_{\mathscr{H}} E(f - (\varphi(y) - Am(y)) \\ &\quad \times d\mu_{sASA^*}(f)\, dv(s).\end{aligned} \tag{16}$$

From this and Lemma 3.2 it is clear that the algorithm $\varphi^s(y) = Am(y)$ (for almost all y) has the desired properties.

(b) Just combine (15), (16), and (a).

(c) This is a consequence of the general relation

$$\int_{\mathscr{H}} F(f)\, d\mu_{|\gamma|^2 K}(f) = \int_{\mathscr{H}} F(\gamma f)\, d\mu_K(f).$$

It is seen from 3.3(c) that for $E(f)=p_{\mathscr{B}}(f)^p$ and other p-homogeneous functions the approximation problem for μ_C^ν is equivalent to the one for μ_C.

Next we want to consider a restricted class of sets $\mathscr{B}$. But before we do so it is appropriate for us to touch on the problem of optimal information. Denote by R_n the operator

$$R_n = ASA^* = A(I-\sigma)\, C(I-\sigma)^*\, A^*$$

and define the nth *radius* of the approximation problem to be

$$r^n = \inf_N r(N).$$

In the next proposition it is tacitly assumed that all eigenvalues of ACA^* are non-degenerate. The general case is similar but more complicated to state.

PROPOSITION 3.4. *Assume that E is a standard error functional of the form $E(f)=H(\|f\|)$.*

Then $r^n = r(\bar{N})$, where the information $\bar{N}$ is given via the n principal eigenvalues and eigenvectors (λ_i, f_i) of ACA^* through $\bar{g}_i = \lambda_i^{-1/2}A^*f_i$. If N is any information then $r(N)=r(\bar{N})$ if and only if

$$Rg_1 + \cdots + \mathbb{R}g_n = \mathbb{R}\bar{g}_1 + \cdots + \mathbb{R}\bar{g}_n.$$

Proof. By 3.3(b) the value of $r(N)$ increases when the eigenvalues of ASA^* increase. Compute

$$\begin{aligned} R_n &= A\left(I - \sum_{i=1}^n g_1 \otimes Cg_i\right) C \left(I - \sum_{i=1}^n Cg_i \otimes g_i\right) A^* \\ &= AC\left(I - \sum_{i=1}^n Cg_i \otimes g_i\right)^2 A^* \\ &= AC\left(I - \sum_{i=1}^n Cg_i \otimes g_i\right) A^* \\ &= AC^{1/2}\left(I - \sum_{i=1}^n C^{1/2}g_i \otimes C^{1/2}g_i\right) C^{1/2}A^*. \end{aligned}$$

Then R_n is given by

$$R_n = AC^{1/2}(I-P)\, C^{1/2}A^*,$$

where P is the orthogonal projection onto the linear span of $\{C^{1/2}g_i\}_{i=1}^n$.

The non-zero eigenvalues of $R_n = (AC^{1/2}(I-P))(AC^{1/2}(I-P))^*$ equal those of

$$\tilde{R}_n = (AC^{1/2}(I-P))^* (AC^{1/2}(I-P))$$
$$= (I-P) C^{1/2}A^*AC^{1/2}(I-P).$$

Repeating the argument we note that the (non-zero) eigenvalues of $C^{1/2}A^*AC^{1/2}$ are $\{\lambda_i\}_{i=1}^{\infty}$. Consequently a minimal set of eigenvalues for R_n, namely $\{\lambda_i\}_{i=n+1}^{\infty}$, exists and is obtained if and only if

$$\mathbb{R}C^{1/2}g_1 + \cdots + \mathbb{R}C^{1/2}g_n = \mathbb{R}\eta_1 + \cdots + \mathbb{R}\eta_n, \tag{17}$$

where η_i are the n principal eigenvectors of $C^{1/2}A^*AC^{1/2}$. However, η_i are proportional to $C^{1/2}A^*f_i$ and (17) is equivalent to

$$\mathbb{R}g_1 + \cdots + \mathbb{R}g_n = \mathbb{R}A^*f_1 + \cdots + \mathbb{R}A^*f_n. \quad \blacksquare$$

The above proposition, which improves [13, pp. 738–741], is included at this point mainly to emphasize that the directions in $\mathscr{H}$ determined by the eigenvectors $\{f_i\}_{i=1}^{\infty}$ of ACA^* have a special significance. Thus prepared the reader will hopefully admit to the relevance of the sets $\mathscr{B}$ in the following corollaries to Theorem 3.3.

COROLLARY 3.5. *Let E be the functional $E(f) = G(p_{\mathscr{B}}(f))$, where $\mathscr{B}$ is defined by*

$$\mathscr{B} = \left\{ f \in \mathscr{H} \,\middle|\, \sum_{i=1}^{\infty} a_i (f, f_i)^2 < 1 \right\}$$

for some bounded set $\{a_i\}_{i=1}^{\infty}$ of positive numbers and G is a continuously differentiable bijection of $\mathbb{R}_+$. Let θ be the real part of the function

$$\varphi(\lambda) = \prod_{j=n+1}^{\infty} (1 - 2i\lambda\lambda_j a_j)^{-1/2}.$$

Then

$$r(\bar{N}) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \int_0^{\infty} G((ts)^{1/2})\, \hat{\theta}(t)\, dt\, dv(s), \tag{18}$$

where $\hat{\theta}$ denotes the Fourier transform.

Proof. First we claim that

$$\mu_{sA\bar{S}A^*}(t\mathscr{B}) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(t^2\lambda)}{\lambda}\, \theta(s\lambda)\, d\lambda.$$

We shan't go into the details of this. The proof is an application of the characteristic function trick that can be found for instance in [17, pp. 66]. Now for $F=G^{-1}$

$$\begin{aligned}
&\frac{d}{dt}(\mu_{sA\bar{S}A^*}(F(t)\mathscr{B}))\\
&\quad=\frac{1}{\pi}\int_{-\infty}^{\infty} 2F'(t)\,F(t)\cos(F(t)^2\,\lambda)\,\theta(s\lambda)\,d\lambda\\
&\quad=\sqrt{\frac{2}{\pi}}\,2s^{-1}F'(t)\,F(t)\,\hat{\theta}(s^{-1}F(t)^2)\\
&\quad=\sqrt{\frac{2}{\pi}}\,\frac{d}{dt}(s^{-1}F(t)^2)\,\hat{\theta}(s^{-1}F(t)^2)
\end{aligned}$$

and by Theorem 3.3(b)

$$\begin{aligned}
r(\bar{N})&=\sqrt{\frac{2}{\pi}}\int_0^{\infty}\int_0^{\infty} t\,\frac{d}{dt}(s^{-1}F(t)^2)\\
&\qquad\times\hat{\theta}(s^{-1}F(t)^2)\,dt\,dv(s)\\
&=\sqrt{\frac{2}{\pi}}\int_0^{\infty}\int_0^{\infty} G((ts)^{1/2})\,\hat{\theta}(t)\,dt\,dv(s).
\end{aligned}$$

This is (18). ∎

COROLLARY 3.6. *For each $E(f)=p_{\mathscr{B}}(f)^{2p}$ it holds that*

$$r(\bar{N})=r(\bar{N},p_{\mathscr{B}}^{2p})=(-i)^p\,\varphi^{(p)}(0)\int_0^{\infty} s^p\,dv(s). \tag{19}$$

In particular

$$r(\bar{N},p_{\mathscr{B}}^2)=\sum_{j=n+1}^{\infty}\lambda_j a_j$$

and

$$r(\bar{N},p_{\mathscr{B}}^4)=\left(2\sum_{j=n+1}^{\infty}(\lambda_j a_j)^2+\left(\sum_{j=n+1}^{\infty}\lambda_j a_j\right)^2\right)\int_0^{\infty} s^2\,dv(s).$$

Proof. From (18)

$$r(\bar{N},p_{\mathscr{B}}^{2p})=\sqrt{\frac{2}{\pi}}\int_0^{\infty} t^p\hat{\theta}(t)\,dt\int_0^{\infty} s^p\,dv(s).$$

Here the second factor may or may not be finite. Our objective is to determine the value of the first factor. Since $\theta(t)=\frac{1}{2}(\varphi(t)+\varphi(-t))$ it follows that $\hat{\theta}(t)=\frac{1}{2}(\hat{\varphi}(t)+\hat{\varphi}(-t))$. Hence

$$\sqrt{\frac{2}{\pi}}\int_0^\infty t^p\hat{\theta}(t)\,dt$$
$$=\frac{1}{\sqrt{2\pi}}\int_0^\infty t^p(\hat{\varphi}(t)+\hat{\varphi}(-t))\,dt$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty |t|^p\,\hat{\varphi}(t)\,dt. \tag{20}$$

When p is even this is

$$\frac{(-i)^p}{\sqrt{2\pi}}\int_{-\infty}^\infty (\varphi^p)\hat{}\,(t)\,dt=(-i)^p\,\varphi^{(p)}(0)$$

and we are done. For odd p (20) can be rewritten as

$$\frac{(-i)^p}{\sqrt{2\pi}}\int_{-\infty}^\infty \operatorname{sgn}(t)\,(\varphi^p)\hat{}\,(t)\,dt. \tag{21}$$

The function or tempered distribution sgn has Fourier transform $\widehat{\operatorname{sgn}}=(-i)\sqrt{2/\pi}\;Vp(1/t)$, where Vp denotes the Cauchy principal value. Thus (20) is equal to

$$\frac{(-i)^{p+1}}{\pi}\lim_{\varepsilon\to 0+}\int_{|t|\geqslant\varepsilon}\frac{\varphi^{(p)}(t)}{t}\,dt. \tag{22}$$

To estimate this integral we exploit the fact that $z^{-1/2}$ is an analytic function of z in the half plane $\{z\in\mathbb{C}\,|\,\operatorname{Re} z>0\}$. Indeed $z^{-1/2}=(1/\sqrt{\pi})\int_{-\infty}^\infty e^{-zx^2}\,dx$. In turn φ is analytic in the region $\{z\in\mathbb{C}\,|\,\operatorname{Im} z>-\gamma\}$, where $\gamma=(\max_j 2\lambda_j a_j)^{-1}$ and the integral of $\varphi(z)/z$ along the contour indicated in Fig. 1 is zero for any values of ε and R.

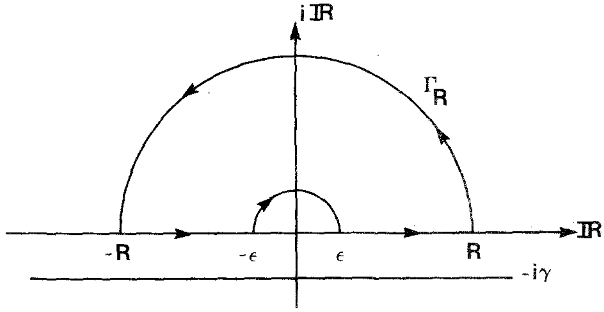

FIGURE 1

Since the integral along the semicircle Γ_R converges to zero it follows that the limit in (22) is in fact

$$\frac{(-i)^{p+1}}{\pi} \lim_{\varepsilon \to 0+} \int_0^\pi \frac{\varphi^{(p)}(\varepsilon e^{i\theta})}{\varepsilon e^{i\theta}} \, i\varepsilon e^{i\theta} \, d\theta$$
$$= (-i)^p \, \varphi^{(p)}(0).$$

Finally, this expression is calculated for the specific values $p=1$ and $p=2$ by use of

$$\varphi'(z) = \Lambda(z) \, \varphi(z), \tag{23}$$

where

$$\Lambda(z) = i \sum_{j=n+1}^{\infty} \frac{\lambda_j a_j}{(1 - 2iz\lambda_j a_j)}. \quad \blacksquare$$

Remark. By iterating (23) and using

$$\Lambda^{(k)}(0) = (i)^{k+1} \, 2^k \cdot k! \sum_{j=n+1}^{\infty} (\lambda_j a_j)^{k+1}$$

one may of course generate any desired instance of $(-i)^p \, \varphi^{(p)}(0)$. But we have not been able to find a closed expression for this.

COROLLARY 3.7. *For* $E(f) = \|f\|^{2p}$ *the* nth *radius* $r^n = r^n(\|\cdot\|^{2p})$ *of the approximation problem is*

$$r^n(\|\cdot\|^{2p}) = (-i)^p \, \varphi^{(p)}(0) \int_0^\infty s^p \, d\nu(s),$$

where $\varphi(z) = \prod_{j=n+1}^{\infty} (1 - 2iz\lambda_j)^{-1/2}$. *In particular*

$$r^n(\|\cdot\|^2) = \sum_{j=n+1}^{\infty} \lambda_j$$

and

$$r^n(\|\cdot\|^4) = \left(2 \sum_{j=n+1}^{\infty} \lambda_j^2 + \left(\bar{2} \sum_{j=n+1}^{\infty} \lambda_j\right)^2\right) \int_0^\infty s^2 \, d\nu(s).$$

Proof. Combine Corollary 3.6 and Proposition 3.4. ∎

For $E(f) = \|f\|^p$ and other standard functionals of the form $E(f) = H(\|f\|)$ we may derive a rather nice expression for the optimal approximation φN. From Theorem 3.3(a) and Proposition 3.4 it follows that φN is optimal if and only if $\mathbb{R}g_1 + \cdots + \mathbb{R}g_n = \mathbb{R}\bar{g}_1 + \cdots + \mathbb{R}\bar{g}_n$ and φ has the form (14). Again let P denote the projection onto the linear span of $\{\bar{g}_i\}_{i=1}^n$. Since $\{C^{1/2}g_i\}_{i=1}^n$ is an orthonormal basis in $C^{1/2}P$ one finds, for any f in the domain of $C^{-1/2}$,

$$\begin{aligned}\varphi^s N(f) &= \sum_{i=1}^n (f, g_i)\, ACg_i \\ &= AC^{1/2} \sum_{i=1}^n (C^{-1/2}f, C^{1/2}g_i)\, C^{1/2}g_i \\ &= AC^{1/2}QC^{-1/2}f,\end{aligned}$$

where Q is the projection onto $C^{1/2}P = \operatorname{span}\{\eta_i\}_{i=1}^n$ (cf. the proof of Proposition 3.4). Here the equation $\varphi^s N(f) = AC^{1/2}QC^{-1/2}f$ is independent of the choice of $\{g_i\}_{i=1}^n$. Consequently $AC^{1/2}QC^{-1/2}$ extends to a bounded operator in $\mathscr{H}$ and this operator is the unique optimal value of φN. Finally, using $g_i = \bar{g}_i = \lambda_i^{-1/2}A^*f_i$ and the very definition of f_i, one finds

$$\begin{aligned}(\varphi N)^{\text{optimal}}(f) &= \sum_{i=1}^n (f, \lambda_i^{-1/2}A^*f_i)\, AC(\lambda_i^{-1/2}A^*f_i) \\ &= \left(\sum_{i=1}^n A^*f_i \otimes f_i\right)(f) \\ &= \left(\sum_{i=1}^n f_i \otimes f_i\right) Af.\end{aligned}$$

Thus $(\varphi N)^{\text{opt}}$ is the composition of A and an orthogonal projection of rank n.

Finally, in closing this paper, we turn to the problem of adaptive versus non-adaptive information. When N is (adaptive) information let N_y, $y \in \mathbb{R}^n$, be the non-adaptive information given by $g_i = g_i(y)$.

The heart of the very elegant proof in [13] that "adaption doesn't help" is the equality $\mu^y(N) = \mu^y(N_y)$ between conditional measures. It is apparent from (11) that this does not hold generally for non-Gaussian measures μ_C^v. Nevertheless we have the following.

THEOREM 3.8. *For any allowable error functional E and any information N*

$$r(N) \geqslant \inf_{y \in \mathbb{R}^n} r(N_y). \tag{24}$$

Further if E is a standard error functional of the form $E(f)=H(\|f\|)$, $r(N)=r^n$ if and only if

$$\mathbb{R}g_1(y)+\cdots+\mathbb{R}g_n(y)=\mathbb{R}\bar{g}_1+\cdots+\mathbb{R}\bar{g}_n \tag{25}$$

holds for almost all y in $\mathbb{R}^n$.

Proof. Using the results of Proposition 3.1, Lemma 3.2, and Theorem 3.3 compute

$$\begin{aligned}
e(\varphi, N) &= \int_{\mathscr{H}} E(Af-\varphi Nf)\, d\mu(f) \\
&= \int_{\mathbb{R}^n} E(Af-\varphi(y))\, d\mu^y(f)\, d\mu_I^y(y) \\
&= \int_{\mathbb{R}^n} \int_0^\infty W_s(y) \int_{\mathscr{H}} E(Af-\varphi(y)) \\
&\quad \times d\mu_{m(y), sS(s^{-1/2}y)}(f)\, dv(s)\, dy \\
&\geqslant \int_{\mathbb{R}^n} \int_0^\infty W_s(y) \int_{\mathscr{H}} E(Af) \\
&\quad \times d\mu_{sS(s^{-1/2}y)}(f)\, dv(s)\, dy \\
&= \int_0^\infty \int_{\mathbb{R}^n} \int_{\mathscr{H}} E(Af)\, d\mu_{sS(y)}(f)\, d\mu_I(y)\, dv(s) \\
&\geqslant \inf_{y\in\mathbb{R}^n} \int_0^\infty \int_{\mathscr{H}} E(Af)\, d\mu_{sS(y)}(f)\, dv(s) \\
&= \inf_{y\in\mathbb{R}^n} r(N_y).
\end{aligned}$$

This proves (24).

For the final case to be considered it can be read of the above string of calculations that $r(N)=r^n$ if and only if

$$\varphi(y)=Am(y) \tag{26a}$$

$$r^n=r(N_y) \tag{26b}$$

holds for almost all y in $\mathbb{R}^n$. Combining (26b) with Proposition 3.4 one gets (25). The optimal algorithm is given by

$$\varphi((y_i)_{i=1}^n)=\sum_{i=1}^n y_i ACg_i(y_1, ..., y_{i-1})$$

for almost all $y=(y_i)_{i=1}^n$ in $\mathbb{R}^n$. ∎

REFERENCES

1. T. W. ANDERSON, The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proc. Amer. Math. Soc.* **6** (1955), 170–175.
2. C. BERG, J. P. R. CHRISTENSEN, AND P. RESSEL, "Harmonic Analysis on Semigroups," Springer-Verlag, New York/Berlin, 1984.
3. J. J. CRAWFORD, Elliptically contoured measures on infinite dimensional Banach spaces, *Stud. Math.* **60** (1977), 15–32.
4. W. DAHMEN, C. A. MICHELLI, AND P. W. SMITH, Asymptotically optimal sampling schemes for periodic functions, *Math. Proc. Cambridge Philos. Soc.* **99** (1986), 171–177.
5. M. A. KOWALSKI, Optimal complexity recovery of band and energy limited signals, *J. Complexity* **2** (1986), 239–254.
6. D. LEE, Approximation of linear operators on a Wiener space, *Rocky Mountain J. Math.* **16** (1986), 641–660.
7. D. LEE AND G. W. WASILKOWSKI, Approximation of linear functionals on a Banach space with a Gaussian measure, *J. Complexity* **2** (1986), 12–43.
8. C. A. MICCHELLI, Orthogonal projections are optimal algorithms, *J. Approx. Theory* **40** (1984), 101–110.
9. C. A. MICCHELLI AND T. J. RIVLIN, A survey of optimal recovery, *in* "Optimal Estimation in Approximation Theory" (C. A. Michelli and T. J. Rivlin, Eds.), Plenum, New York, 1977.
10. R. SCHARLACH, Optimal recovery by linear functionals, *J. Approx. Theory* **44** (1985), 167–172.
11. A. V. SKOROHOD, "Integration in Hilbert Space," Springer-Verlag, Berlin/New York, 1974.
12. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOZNIAKOWSKI, Average case optimality for linear problems, *Theoret. Comput. Sci.* **29** (1984), 1–25.
13. G. W. WASILKOWSKI, Optimal algorithms for linear problems with Gaussian measures, *Rocky Mountain J. Math.* **16** (1986), 727–749.
14. G. W. WASILKOWSKI AND H. WOZNIAKOWSKI, Can adaption help on the average?, *Numer. Math.* **44** (1984), 169–190.
15. G. W. WASILKOWKSI AND H. WOZNIAKOWSKI, Average case optimal algorithms in Hilbert spaces, *J. Approx. Theory* **47** (1986), 17–25.
16. H. WOZNIAKOWSKI, Information based complexity, *Ann. Rev. Comput. Sci.* **1** (1986), 319–380.
17. G. E. SILOV AND FAN DYK TIN, "Integral, mera i proizvodnaja no linejnych prostranstvach," Nauka, Moskva, 1967 [Russian].