Can Adaption Help on the Average?*

G.W. Wasilkowski and H. Woźniakowski

Institute of Informatics, University of Warsaw, Warsaw, Poland and
Department of Computer Science, Columbia University, New York, New York, USA

Summary. We study adaptive information for approximation of linear problems in a separable Hilbert space equipped with a probability measure $\mu$. It is known that adaptation does not help in the worst case for linear problems. We prove that adaptation also does not help on the average. That is, there exists nonadaptive information which is as powerful as adaptive information. This result holds for "orthogonally invariant" measures. We provide necessary and sufficient conditions for a measure to be orthogonally invariant. Examples of orthogonally invariant measures include Gaussian measures and, in the finite dimensional case, weighted Lebesgue measures.


Introduction

We explain the setting of the problem using a simple integration example. Suppose one seeks an approximation to $\int_0^1 f(t)dt$ knowing $n$ values of $f$ at points $t_1, N(f) = \{f(t_1), f(t_2), \ldots, f(t_n)\}$, and knowing that $f$ belongs to a given class $F$ of functions. If the points $t_1, t_2, \ldots, t_n$ are given simultaneously then $N = N^n$ is called nonadaptive information. If the second point $t_2$ depends on the previously computed value $f(t_1)$, i.e., $t_2 = t_1\{f(t_1)\}$ and if the point $t_i$ depends on the previously computed values $f(t_2), \ldots, f(t_{i-1})$ i.e., $t_i = t_i\{f(t_1), \ldots, f(t_{i-1})\}$, then $N = N^i$ is called adaptive information.

The structure of adaptive information is much richer than the structure of nonadaptive information. Therefore one might hope that adaptive information can be much more powerful than nonadaptive information, i.e., an approxima-

---

* This research was supported in part by the National Science Foundation under Grant MCS-7823676
tion that uses adaptive information has much smaller error than an approximation that uses nonadaptive information.

What do we mean by error? It depends which model we have in mind. Consider first the worst case model. In this model the error of an algorithm \( \phi \) (for our simple example) is defined by

\[
\epsilon(\phi, N) = \sup_{f \in F} \int_0^1 f(t) dt - \phi(N(f))
\]  

(1.1)

By an algorithm we mean any mapping \( \phi \) which maps \( N(f) \) into \( \mathbb{R} \). Then

\[
r(N) = \inf_{\phi} \epsilon(\phi, N)
\]

(1.2)

is called the radius of information and \( \phi \) is optimal iff \( \epsilon(\phi, N) = r(N) \).

Does adaption help in the worst case? That is, does there exist a choice of points \( t_i = t_i(f(t_i), \ldots, f(t_{i-1})) \) such that

\[
r(N^a) < r(N^{non})
\]

A surprising answer is no, at least for some classes \( F \). More precisely, if the class \( F \) is convex and balanced (i.e., \( f \in F \) implies \( -f \in F \)) then there exist points \( t_i^* \) such that the nonadaptive information \( N^{non}(f) = [f(t_1^*), \ldots, f(t_k^*)] \) is as powerful as the adaptive information \( N^a \), i.e.,

\[
r(N^{non}) \leq r(N^a)
\]

This was established in [1] for arbitrary linear functionals. It was generalized to arbitrary linear operators and information consisting of linear functionals in [2] and [12, Theorem 7.1, Chap. 2]. A further generalization may be found in [11].

It is also known that there are nonlinear problems such that adaption does not help in the worst case; see [3, 8, 9, 14] and [17].

We stress that in the worst case there do exist nonlinear problems for which adaption is far more powerful. An example of such a problem is zero finding for scalar functions which change sign at the endpoints of the interval \([u, b]\). Then the optimal nonadaptive information has radius \((b - u)(2n - 1)\) whereas the optimal adaptive information is bisection information which has radius \((b - u)2^{-n-1}\); see [12, Theorem 2.1, Chap. 8] and [7].

As long as \( F \) is convex and balanced, adaption does not help in the worst case for linear problems. One may think that this is due to a model assumption, i.e., that the error of an algorithm is determined by its performance for the hardest \( f \). One might hope that with a more realistic definition of error, the converse result would be true, i.e., adaption helps, perhaps even significantly, for linear problems.

It seems natural to propose the average error of an algorithm as a more realistic measure of its performance. Technically, this means that we replace supremum in (1.1) by integral, i.e.,

\[
\epsilon^{ave}(\phi, N) = \int_{F} \left( \int_0^1 f(t) dt - \phi(N(f)) \right)^2 \mu(df)
\]

(1.3)
where $\mu$ is a probability measure on $F$. Note that even for our simple example, $F$ usually lies in an infinite dimensional space and therefore the analysis of (1.3) requires measure theory in infinite dimensional spaces. Thus the analysis of average case error is much harder than the analysis of worst case error.

Define

$$r^{*}(N) = \inf_{\phi} r^{*}(\phi, N)$$

as the average radius of information.

Does adaptation help on the average? That is, does there exist a choice of points $t_i = (f(t_1), \ldots, f(t_{i-1}))$ such that

$$r^{*}(N') < r(N^{\text{non}})?$$

The surprising answer is no for linear problems. This was established in [10] for a finite dimensional Hilbert setting with a weighted Lebesgue measure and with a general error criterion. In this paper we show that adaptation does not help on the average for linear problems in infinite dimensional Hilbert spaces with an "orthogonally invariant" measure $\mu$. Orthogonal invariance of $\mu$ means that the measure of a Borel set is invariant under certain linear orthogonal mappings. Examples of orthogonally invariant measures include Gaussian measures. For the finite dimensional case with $\mu$ absolutely continuous with respect to the Lebesgue measure, orthogonal invariance coincides with a weighted Lebesgue measure, see Corollary 3.1. Thus this coincides with measures studied in [10].

Our result holds for adaptive information operators which are measurable and which consist of arbitrary inner products. In particular, it holds for adaptive information operators used in practice which are usually continuous almost everywhere. We illustrate this point by an integration example. Usually the next point $t_{i+1}$ at which $f$ is to be evaluated, depends on whether $f(t_1), f(t_2), \ldots, f(t_i)$ for some of them satisfy a certain Boolean condition, i.e.,

$$t_{i+1} = \begin{cases} u_{1}, & \text{if } \text{Cond}(f(t_1), \ldots, f(t_i)) = \text{true}, \\ u_{2}, & \text{if } \text{Cond}(f(t_1), \ldots, f(t_i)) = \text{false}. \end{cases}$$

for some $u_1$ and $u_2$. Then $t_{i+1}$, as a function of the previously computed information, is a piecewise constant function. Thus it is not continuous but it is continuous almost everywhere for a reasonable choice of measure. See for example, the discussion on adaptive integration in [5, pp. 126-130].

We have given a number of references dealing with adaptive information for nonlinear problems in the worst case. There exist no such paper for the average case model. We hope that the study of nonlinear problems in the average case model will be one of the foci of future research.

We stress, by all means, that the worst and average case models are not the only interesting models to be studied. An asymptotic model, in which the total number of evaluations is not fixed a priori, should be analyzed. The question as to whether adaptation helps for linear problems in the asymptotic case is analyzed in (J.M. Trojan, in preparation). The answer is once more no. Some preliminary study indicates that adaptation does not help in the asymptotic average case. Results for this model will be reported in the future.
Why are we interested in the question whether adaption is more powerful than nonadaptation? There are a number of reasons which include:

(i) Intrinsic mathematical interest. Adaptation corresponds to certain non-linear operators whereas nonadaptation corresponds to linear operators. Mathematically the sentence "adaptation does not help" means that this nonlinearity is no more powerful than linearity.

(ii) Reduction of the search for optimal information. If adaptation does not help then we only have to look at the very special and relatively easy non-adaptive case to find optimal information.

(iii) Speedup for parallel computations. Nonadaptive information is naturally decomposable and can be computed very efficiently in parallel. Adaptive information is not decomposable an is ill-suited for parallel computations. For instance, for the integration example if a function evaluation costs unity and there are \( n \) processors then nonadaptive information costs unity and adaptive information costs \( n \).

A more detailed discussion of this subject may be found in [13].

We briefly summarize the contents of this paper. In Sect. 2 we formulate the problem, introduce the concept of orthogonal invariance and state the main theorem of this paper. The proof of the theorem requires some properties of orthogonally invariant measures. Therefore Sects. 3 and 4 deal with characterization and properties of orthogonally invariant measures. In particular, we prove that orthogonal invariance of \( \mu \) is equivalent to orthogonal invariance of its projections into finite dimensional subspaces. We also characterize orthogonal invariance for the finite dimensional case. We prove that the measure of a Borel set is invariant under a certain nonlinear mapping. This is basic to the proof in Sect. 5 that the spline algorithm is an optimal average error algorithm. The proof of the main theorem is given in Sect. 5.

2. Adaptive Information

Let \( F_1 \) and \( F_2 \) be real separable Hilbert spaces. Let \( S : F_1 \to F_2 \) be a linear continuous operator. Our aim is to approximate \( Sf \) for any \( f \) from \( F_1 \). We assume that instead of \( f \), we know \( N(f) \). Here \( N \) is an adaptive information operator defined by

\[
N(f) = \{ (f, g_1), (f, g_2, (y_1)), \ldots, (f, g_{n-1}, y_1, \ldots, y_{n-1}) \} \tag{2.1}
\]

where \( y_i = (f, g_i), i = (f, g_i, (y_1), \ldots, y_{i-1}, g_i(y_1), \ldots, y_{i-1}) \) is an element of \( F_1 \) and \( (\cdot, \cdot) \) is the inner product of \( F_1 \). The essence of (2.1) is that the choice of \( g_i(y_1), \ldots, y_{i-1} \) may depend on the \((i-1)\) previously computed inner products. For brevity we shall write

\[
g_i(f) = g_i, \quad g_i(f) = g_i(y_1, \ldots, y_{i-1}) \quad 2 \leq i \leq n \tag{2.2}
\]

To stress that \( N \) is adaptive we shall sometimes write \( N = N^a \). If each \( g_i(f) \) does not depend on \( f \), i.e., \( g_i(f) \equiv g \) for some \( g \) from \( F_1 \), then \( N \) is called non-adaptive and denoted by \( N = N^{\text{non}} \), i.e.,

\[
N^{\text{non}}(f) = \{ (f, g_1), (f, g_2, (y_1)), \ldots, (f, g_{n-1}) \} \tag{2.3}
\]
Note that nonadaptive information is a linear operator whereas adaptive information is in general nonlinear. Without loss of generality we assume that $g_1(f), g_2(f), ..., g_s(f)$ are linearly independent for each $f$ from $F_1$.

Knowing $N(f)$ we approximate $Sf$ by $\omega N(f)$ where $\omega$ is a mapping from $N(F_1)$ into $F_1$. We call such $\omega$ an (idealized) algorithm. We wish to approximate $Sf$ with an average error as small as possible. The average error of $\omega$ is defined as

$$e^{\omega N}(\omega, N) = \int_{F_1} |Sf - \omega N(f)|^2 \mu(df).$$  \tag{2.4}$$

Here $\mu$ is a probability measure defined on Borel sets of $F_1$ and the integral in (2.4) is understood as the Lebesgue integral. We assume that an algorithm $\omega$ is chosen such that (2.4) is well defined, i.e., $\|Sf - \omega N(f)\|^2$ is a measurable function. This assumption is not restricted as is shown in [15]. Let

$$r^{\omega N}(N) = \inf_{\omega \in \Phi(N)} e^{\omega N}(\omega, N)$$ \tag{2.5}$$

be the average radius of information where $\Phi(N)$ denotes the class of all algorithms using $N$ for which the average error is well defined.

The main problem addressed in this paper is to show that for a wide class of measures, adaptive information is not stronger than correspondingly chosen nonadaptive information. Thus the much more complicated structure of adaptive information operators does not supply more knowledge about linear problems than the relatively simple structure of nonadaptive information operators.

This result holds for "orthogonally invariant" measures $\mu$. This concept will be defined below. We assume that $\int_{F_1} f^2 \mu(df) < \infty$. Without loss of generality we can assume that the mean element of the measure $\mu$ is zero, i.e.

$$\int_{F_1} (f, x) \mu(df) = 0, \quad \forall x \in F_1, \quad \text{and} \quad \int_{F_1} (f, x)^2 \mu(df) > 0, \quad \forall x \in F_1, \quad x = 0.$$

Let $S_\mu$ be the covariance operator of $\mu$, i.e., $S_\mu : F_1 \to F_1$ and

$$(S_\mu, x, y) = \int_{F_1} (f, x)(f, y) \mu(df), \quad \forall x, y \in F_1.$$  \tag{2.6}$$

The operator $S_\mu$ is a linear self-adjoint, positive definite operator and has finite trace. If $\dim F_1 = + \infty$, then $S_\mu(F_1)$ is a proper dense subset of $F_1$ and $S_\mu^{-1} : S_\mu(F_1) \to F_1$ is a linear unbounded operator. See [4,6] and also [16]. Let

$$(x, y)_* = (S_\mu^{-1} x, y), \quad \forall x, y \in S_\mu(F_1).$$  \tag{2.7}$$

Then $x_* = (x, x)_* = (S_\mu^{-1} x, x)$. We say $\mu$ is orthogonally invariant iff

$$\mu(\Omega B) = \mu(B)$$  \tag{2.8}$$

for any Borel set $B$ and any linear mapping $Q, Q : F_1 \to F_1$, of the form

$$Qf = 2(f, h)S_\mu h - f$$  \tag{2.9}$$
for any \( h \) such that \( tS, h, h = 1 \) or \( h = 0 \). For \( h = 0 \), \( Qf = f \) and (2, 8) means that 
\[ \mu(B) = \mu(B) \quad \text{where} \quad -B = \{ t : -f \in B \}. \]
Note that \( f \in S_\alpha(F_1) \) implies that 
\[ Qf \in S_\alpha(F_1) \] and 
\[ Qf = (2\alpha f, h) S_\alpha^{-1} f; 2\alpha f, h) S_\alpha h = f. \]
Thus the mapping \( Q \) is orthogonal in the norm \( \| \cdot \| \). This explains why \( \mu \) is
called orthogonally invariant.

It is shown in [16] that Gaussian measures are orthogonally invariant as well as measures of the form
\[ \mu(B) = \int_B w(|S_\alpha^{-1} f|) \lambda(df) \]
for some measurable function \( w \) assuming that \( F_1 \) is finite dimensional and \( \lambda \) is the Lebesgue measure.

Note that \( Q \) resembles a Householder matrix. It is easy to check that
\[ Q^2 = I, \quad Q^{-1} = Q. \]
This important property will be extensively used in this paper. In Sect. 3 we characterize orthogonally invariant measures in detail.

We shall show in Sect. 5 that without loss of generality we can assume that
\[ (S_\alpha g, (f, g, f)) = \delta_{ij}. \]
Let
\[ a = \sup \left\{ \sum_{i=1}^n SS_\alpha g, (f, i) : f \in F_1 \right\}. \]
For simplicity assume that \( a \) is obtained for \( f^* \), i.e.,
\[ \sum_{i=1}^n SS_\alpha g, (f^*) = \sup \sum_{i=1}^n SS_\alpha g, (f). \]
Let \( g_i^* = g, (t^*) \). By \( N_i^{2n} \) we mean
\[ N_i^{2n+1} = \{ l f, g_i^*, l f, g_i^*, \ldots, l f, g_i^* \}. \]
Note that \( N_i^{2n} \) is nonadapted and is obtained by fixing \( g, (f) \) in the adaptive information \( N_i \).

We say that
\[ N_i(t) = \{ l f, g_i, (f), l f, g_i, (f) \} \]
is measurable if \( g, (t^*) \) is measurable, i.e., \( g, (t^*) \) is a Borel set for a Borel set \( B \) of \( R^{t^*} \), \( i = 2, 3, \ldots, n \).
We are ready to state the main result of this paper.

**Theorem 2.1.** Let \( \mu \) be an orthogonally invariant measure. Let \( N^i \) be measurable
adaptation information. Then
\[ r^{2n+1} (N^i) \geq r^{2n} (N_i^{2n}). \]

Thus adaptation does not help on the average for linear problems. As we
already mentioned in the introduction it does not help for the worst case model.
The proof of Theorem 2.1 depends heavily on the properties of orthogonally invariant measures. In Sect. 3 we characterize orthogonally invariant measures. The results of Sect. 3 are of intrinsic interest. In Sect. 4 we derive properties of orthogonally invariant measures. Section 5 contains the proof of Theorem 2.1. The proof is based on two results on orthogonally invariant measures. The first result is that for orthogonally invariant measures, the measure of a set is invariant under a certain nonlinear mapping. The second is that the measures \( \mu(N^{q-1}) \) are orthogonally invariant and independent of \( N^q \). Assuming these two results, the reader can skip Sects. 3 and 4 and turn to Sect. 5.

3. Orthogonal Invariance of Measure

We show in this section which measures are orthogonally invariant. Our analysis will be first done for a finite dimensional case, \( \dim(F_1) < +\infty \). We find, in particular, a condition for \( \mu \) to be orthogonally invariant whenever \( \mu \) is absolutely continuous with respect to the Lebesgue measure \( \lambda \). Next we consider the general case, \( \dim(F_1) \leq +\infty \). We show that orthogonal invariance of \( \mu \) is equivalent to orthogonal invariance of its finite dimensional projections.

(i) Assume in this subsection that \( m = \dim(F_1) < +\infty \). Then the operator \( S \) is bounded and

\[
T = S^{-1}: F_1 \to F_1
\]

is well defined. By \( \eta \otimes \eta \) we mean a linear operator from \( F_1 \) into \( F_1 \) such that \( \eta \otimes \eta(f) = \eta f \). Let \( Q \) be of the form (2.9). Then \( Qf = T^{-1}(2\eta \otimes \eta - I)Tf \) where \( \eta = S^2h \) and \( \eta = 1 \) or \( \eta = 0 \). Hence, the measure \( \mu \) is orthogonally invariant iff

\[
\mu(T^{-1}(2\eta \otimes \eta - I)TB) = \mu(B)
\]

for any Borel set \( B \) and any \( \eta \) such that \( \|\eta\| = 1 \) or \( \eta = 0 \).

We characterize orthogonally invariant measures \( \mu \) which are absolutely continuous with respect to the Lebesgue measure \( \lambda \). Recall that \( \mu \) is absolutely continuous w.r.t. to \( \lambda \) (denoted by \( \mu \ll \lambda \)) iff \( \lambda(A) = 0 \Rightarrow \mu(A) = 0 \) for every Borel set \( B \). If \( \mu \ll \lambda \), then the Radon-Nikodym theorem, see e.g. [6], guarantees the existence of a nonnegative measurable mapping \( g: F_1 \to \mathbb{R}_+ \) such that

\[
\mu(B) = \int_B g(f) \lambda(\text{d}f).
\]

For simplicity we assume that \( g \) is continuous almost everywhere, i.e., there exists a set \( A \), \( \lambda(F_1 - A) = 0 \), such that \( f \in A \) implies that \( g \) is continuous at \( f \).

Theorem 3.1. The measure \( \mu \) is orthogonally invariant iff

\[
g(f_1) = g(f_2) \text{ for any } f_1, f_2 \in A \text{ such that } f_1 \sim f_2 \sim f.
\]

Proof. Suppose \( \mu \) is orthogonally invariant. Take \( f_1 \) and \( f_2 \) from \( A \) such that \( f_1 \sim f_2 \). Define \( \eta = T(f_1 - f_2) \) for \( f_1 \neq f_2 \), and \( \eta = 0 \) for \( f_1 = f_2 \). Then

\[
\mu(T^{-1}(2\eta \otimes \eta - I)TB) = \mu(B)
\]

for any Borel set \( B \) and any \( \eta \) such that \( \|\eta\| = 1 \) or \( \eta = 0 \).

For simplicity we assume that \( g \) is continuous almost everywhere, i.e., there exists a set \( A \), \( \lambda(F_1 - A) = 0 \), such that \( f \in A \) implies that \( g \) is continuous at \( f \).

Theorem 3.1. The measure \( \mu \) is orthogonally invariant iff

\[
g(f_1) = g(f_2) \text{ for any } f_1, f_2 \in A \text{ such that } f_1 \sim f_2 \sim f.
\]

Proof. Suppose \( \mu \) is orthogonally invariant. Take \( f_1 \) and \( f_2 \) from \( A \) such that \( f_1 \sim f_2 \). Define \( \eta = T(f_1 - f_2) \) for \( f_1 \neq f_2 \), and \( \eta = 0 \) for \( f_1 = f_2 \).
\(-t_i\). Let \(Q = T^{-1}(2\eta \otimes \eta - 1)T\). We have \(2tt_1, Tt, t, t_1) = t, t_1, t') \) and \(Q f_1 = f_1\). Then (3.2) yields \(\mu(Q\mathcal{B}) = \mu(B)\) for any Borel set \(B\). Observe that \(\det Q = 1\). This and (3.3) yield
\[
\int_B (g(f) - g(Qf)) \lambda(df) = 0, \quad \forall B - \text{Borel set.} \tag{3.5}
\]
Note that \(g - gQ\) is continuous at \(f_1\) and \(g(f_1) - g(Qf_1) = g(f_1) - g(f_2)\). Suppose that \(g(f_1) - g(f_2) = 0\). Due to continuity of \(g - gQ\) at \(f_1\), there exists a positive \(r\) such that for \(f \in B = \{ f \in F_1 : |f - f_1| < r \}\) we have \(\text{sign}(g(f) - g(Qf)) = \text{constant}\). Since \(\lambda(B) > 0\) we have
\[
\int_B (g(f) - g(Qf)) \lambda(df) \neq 0
\]
which contradicts (3.5). Hence \(g(f_1) = g(f_2)\) as claimed.

Assume now that (3.4) holds. Then for an orthogonal \(Q\) in the norm \(\parallel \cdot \parallel\), we have \(\det Q = 1\) and
\[
\mu(Q\mathcal{B}) = \int_{Q\mathcal{B}} g(f) \lambda(df) = \int_{Q\mathcal{B} \cap A} g(f) \lambda(df) = \int_{B \cap A} g(Qf) \lambda(df) = \int_{B \cap A} g(f) \lambda(df).
\]
Note that \(f \in A \cap Q(A)\) implies \(Qf \in A\). Since \(Qf \parallel f\), then (3.4) yields \(g(Qf) = g(f)\). Thus we have
\[
\mu(Q\mathcal{B}) = \int_{B \cap Q(A) \cap A} g(f) \lambda(df) \leq \mu(B) \tag{3.6}
\]
for any Borel set \(B\). Setting \(B = Q(C)\) we have \(\mu(C) \leq \mu(Q(C))\) for any Borel set \(C\). Hence \(\mu(Q\mathcal{B}) = \mu(B)\). This means that \(\mu\) is orthogonally invariant. \(\square\)

The condition (3.4) means that \(g\) depends on the norm of \(f\). More precisely, let \(X = \{ f \parallel f \in A \}\). Define \(w : \mathbb{R} \to \mathbb{R}\) such that
\[
w(x) = \begin{cases} g(f) & x \in X, \\ 0 & x \in X'
\end{cases} \tag{3.7}
\]
where \(f \in A\) and \(f' = x\). Due to (3.4), \(w\) is well defined. For \(f \in A\) we have \(w(f) = g(f)\). Since
\[
\mu(B) = \int_{B \cap A} g(f) \lambda(df) = \int_{B \cap A} w(f) \lambda(df) = \int_B w(f) \lambda(df).
\]
Thus we have proven

**Corollary 3.1.** The measure \(\mu\) is orthogonally invariant iff
\[
\mu(B) = \int_B w(f) \lambda(df), \quad \forall B - \text{Borel set.} \tag{3.8}\]

The measures considered in [10] are of the form (3.8) and therefore they are orthogonally invariant.
We now turn to the general case \( \dim(F) \leq r \). If \( \dim(F) = -r \) then \( T = S_m^{-1} \) is unbounded and for \( t \notin S_m^{-1} F \), \( Tt \) is not well defined. Therefore the results of subsection (ii) do not hold.

We exhibit relations between orthogonal invariance of \( \mu \) and orthogonal invariance of its finite dimensional projections. Let \( \zeta_1, \zeta_2, \ldots \) be orthonormal eigenelements of the covariance operator \( S_m \), i.e.,

\[
S_m \zeta_i = \lambda_i \zeta_i.
\]

(3.9)

where \( \lambda_1 \geq \lambda_2 \geq \ldots \). Let \( X_m = \text{lin}(\zeta_1, \zeta_2, \ldots, \zeta_m) \) and let \( P_m \) be an orthogonal projection,

\[
P_m : F \to X_m.
\]

(3.10)

Let \( \mu_m \) be the projection of the measure \( \mu \) onto \( X_m \), i.e.,

\[
\mu_m(B) = \mu(P_m^{-1} B)
\]

(3.11)

for any Borel set \( B \) in \( X_m \), see [6]. We are ready to prove

**Theorem 3.2.** The measure \( \mu \) is orthogonally invariant iff the measures \( \mu_m \) are orthogonally invariant for \( m = 1, 2, \ldots \). □

**Proof:** Assume that \( \mu \) is orthogonally invariant. For any \( m \), take a mapping \( Q : X_m \to X_m \) of the form (2.9), i.e.,

\[
Qf = 2(f, h)S_m h - f
\]

where \( S_m \) is the covariance operator of the measure \( \mu_m \) and \( h \in X_m \), \( (S_m h, h) = 1 \) or \( h = 0 \). First of all we show that \( S_m x = S_m x, x \in X_m \). Indeed, for \( x, y \in X_m \) we have

\[
(S_m x, y) = \int f(x) (x, y) \mu_m(df) = \int (P_m f, x)(P_m f, y) \mu(df)
\]

\[
= (S_m P_m x, P_m y) = (S_m x, y).
\]

Since \( X_m \) is an invariant subspace of \( S_m \), \( S_m x \in X_m \) and \( S_m x = S_m x, x \in X_m \), as claimed. Thus \( Q \) can be extended to the space \( F \), with \( S_m \) replaced by \( S \). Let \( B \) be a Borel set in \( X_m \). Note that

\[
P_m^{-1} Q B = Q P_m^{-1} B.
\]

(3.12)

Indeed, if \( f \in P_m^{-1} Q B \) then \( f = Qb - f_i \) where \( b \in B \) and \( f_i \in X_m \). Since \( Qf_i = -f_i \), we have \( f = Q(b - f_i) \in Q(P_m^{-1} B) \). Assume now that \( f \in Q(P_m^{-1} B) \). Then \( f = Q(b - f_i) \) where \( b \in B \) and \( f_i \in X_m \). Thus \( f = Qb - f_i \in P_m^{-1} Q B \) as claimed.

From (3.11), (3.12) and orthogoninal invariance of \( \mu \) we have

\[
\mu_m(QB) = \mu(P_m^{-1} Q B) = \mu(Q P_m^{-1} B) = \mu(P_m^{-1} B) = \mu_m(B).
\]

(3.13)

Thus \( \mu_m \) is orthogonally invariant which completes this part of the proof.

Let \( \mu_m \) be orthogonally invariant. Let \( Q \) be of the form (2.9), i.e., \( Qf = 2(f, h)S_m h - f \) for some \( h \) such that \( (S_m h, h) = 1 \) or \( h = 0 \). Define

\[
Z = \{ B : B \text{ is a Borel set in } F, \mu(QB) = \mu(B) \}.
\]

(3.14)

Can Adaptation Help on the Average?
Observe that $Z$ is a $\sigma$-field. Indeed, if $B_i \in Z$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, then $Q_{B_i} \cap Q_{B_j} = \emptyset$ since $Q$ is one-to-one. Then

$$\mu \left( \bigcup_{i=1}^n B_i \right) = \mu \left( \bigcup_{i=1}^n Q B_i \right) = \sum_{i=1}^n \mu(Q B_i) = \sum_{i=1}^n \mu(B_i) = \mu \left( \bigcup_{i=1}^n B_i \right).$$

Thus $\bigcup_{i=1}^n B_i \in Z$. Of course $\emptyset \in Z$ and $B \in Z$ implies that $F_i - B \in Z$. Hence $Z$ is a $\sigma$-field as claimed.

We now show that each closed ball $B = \{f : f - a \leq r\}$ with $a \in X_m$ for some $m_n$ belongs to $Z$. Recall that $Q_B = \sum_j S_j h_j - f$ where $(S_j, h_j)$ is $1$ or $h_j = 0$. If $h_j = 0$ then take an index $j$ such that $P h_j \neq 0$. Define $h_j = c P h_j$ where $c = (S_j h_j P, P h_j)^{-1}$. If $h_j = 0$, set $h_j = 0$. Then $h_j \in X_j$ and $(S_j h_j, h_j) = 1$ of $h_j = 0$.

Define the mapping $Q_j : X_j \to X_j$ by

$$Q_j f = \sum_j S_j h_j - f.$$  

Note that $Q_j$ is of the form (2.9) for the space $X_j$. We have $h_j \to h$ and $Q_j f \to Q(f)$ as $j$ tends to $-\infty$. We now prove that

$$\bigcap_{j=1}^\infty \bigcup_{J \in J} P_j^{-1} Q_j P_j B \subseteq Q B.$$  \hspace{1cm} (3.15)

Indeed, let $x_j$ belong to the left hand side of (3.15). Then there exists a subsequence $j_k \to -\infty$ such that $x \in P_{j_k}^{-1} Q_{j_k} P_{j_k} B$. Thus $P_{j_k} x = Q_{j_k} (P_{j_k} b)$ where $h_{j_k} \in B$. From this we have

$$Q_{j_k} P_{j_k} x = P_{j_k} b \in P_{j_k} B.$$  

If $j_k \geq m_0$ then $P_{j_k} b - a = P_{j_k} (b - a) \leq b - a \leq r$ for any $b$ from $B$. Thus $P_{j_k} B \subseteq B$ and $Q_{j_k} P_{j_k} x \in B$. Since $B$ is closed then $\lim_{j_k} P_{j_k} x = Q x \in B$ and $x \in Q B$. This shows that the left hand side of (3.15) is contained in $Q B$. From (3.15) we have

$$\mu(Q B) \geq \lim_{-\infty} \mu \left( \bigcap_{j=1}^\infty \bigcup_{J \in J} P_j^{-1} Q_j P_j B \right) \geq \lim_{-\infty} \mu(P_j^{-1} Q_j P_j B)$$

$$= \lim_{-\infty} \mu(P_j P_j B).$$

Since $\mu$ is orthogonally invariant then

$$\mu(Q B) = \mu(P B) = \mu(P_j^{-1} P B) \geq \mu(B).$$

Thus

$$\mu(Q B) \geq \mu(B).$$  \hspace{1cm} (3.16)

To prove the opposite inequality we show that

$$Q B \subseteq \bigcup_{j=1}^\infty \bigcap_{J \in J} P_j^{-1} Q_j P_j B.$$  \hspace{1cm} (3.17)
where $B = \bigcap_{j=1}^{\infty} [-u, u]$. Indeed, $x \in QB$ means that $x = Qh$ and $h = u \leq r$. Note that $Q \pi_i Q h$ tends to $Q^2 h = h$ as $i \to \infty$. Thus there exists an index $i_0$ such that $Q \pi_i Q h - u \leq \varepsilon$ for $j \geq i_0$. Hence $Q \pi_i Q h \in B_j$. Since $Q \pi_i X_j = X_j$ then $Q \pi_i Q h = Q \pi_i Q h \pi_i Q h \pi_i P B_j$. Since $Q \pi_i = 1$ we have $Q \pi_i Q h \in Q \pi_i Q h \pi_i P B_j$ and $Q h \pi_i P B_j$ for $j \geq i_0$. Thus $x = Q h \in \bigcap_{j=i_0}^{\infty} P^{-1} Q \pi_j P B_j$ which completes the proof of (3.17). From this we have

$$\mu(QB) \leq \lim_{i \to \infty} \mu \left( \bigcap_{j=i}^{\infty} P^{-1} Q \pi_j P B_j \right) \leq \lim_{i \to \infty} \mu(P^{-1} Q \pi_j P B_j)$$

$$= \lim_{i \to \infty} \mu(Q \pi_j P B_j) = \lim_{i \to \infty} \mu(P B_j).$$

We now show that

$$B_e = \bigcap_{i=1}^{\infty} P^{-1} P B_i. \quad (3.19)$$

Since $B_e \subseteq P^{-1} P B_i, \forall i$, it is enough take $x \in \bigcap_{i=1}^{\infty} P^{-1} P B_i$ and show that $x \in B_e$. We have $P_i x \in P B_i$ and since $P_i a = a$ for $i \geq m_0$ we get $P_i x \in P B_i$. Note that $P_i x$ tends to $x$ and $B_e$ is closed which yields that $x \in B_e$ as claimed. Since $P_{i}^{-1} P_{i-1} B_i \subseteq P_{i-1} P \pi_i P B_i,$ then (3.19) yields

$$\mu(B_e) = \lim_{i \to \infty} \mu(P B_i).$$

This and (3.18) yield

$$\mu(QB) \leq \mu(B_e). \quad (3.20)$$

Note that (3.20) holds for any positive $e$. Let $e = k^{-1}$ with $k$ tending to infinity. Since $B = \bigcap_{i=1}^{\infty} B_{i+1}$ and $\mu(B) = \lim_{k \to \infty} \mu(B_k)$ we have from (3.20), $\mu(QB) \leq \mu(B)$. This and (3.16) yield

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

for any closed ball with center lying in $X_m$ for some $m_0$.

Thus $B \subseteq Z$. Since any closed ball $A = \bigcap_{i=1}^{\infty} [-u, u] = \bigcap_{i=1}^{\infty} [-P_i u, u]$ and $Z$ is a $\sigma$-field, $A$ belongs to $Z$. Hence $Z$ contains all closed balls and therefore it contains all Borel sets. Hence

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

for any Borel set $B$. Since $Q$ is an arbitrary mapping of the form (2.9), this proves that $\mu$ is orthogonally invariant. This completes the proof. $\Box$

4. Properties of Orthogonally Invariant Measures

The proof of Theorem 2.1 depends on properties of the orthogonally invariant measure $\mu$ which will be obtained in this section.
Let
\[ N^2(f) = \left[ I, g_1(f), (f, g_1(f)), \ldots, (f, g_1(f)) \right] \]  \hspace{1cm} (4.1)
be measurable adaptive information. This means that \( g_1(f), \ldots, g_1(f) \) are measurable and are of the form (2.2). Assume that \( (S, f) g_1(f) = \delta_f, \forall f \in F_1 \). (We show in Sect. 5 that this assumption is not restrictive.)

Define the mapping \( D : F_1 \to F_1 \) by
\[ D(f) = \sum_{i=1}^{n} (f, g_i(f)) S_{g_i(f)} g_i(f) - f. \]  \hspace{1cm} (4.2)

The mapping \( D \) plays an important role in our analysis. Observe that \( D \) is measurable. For nonadaptive information, i.e., \( g_1(f) = g_0 \), \( D \) is linear. For adaptive information \( D \) is nonlinear. The mapping \( D \) has four important properties
\[ N^2(D(f)) = N^2(f), \] \hspace{1cm} (4.3)
\[ D^{-1} = D, \] \hspace{1cm} (4.4)
\[ D(f) = 0 = f, \quad \forall f \in S_0(F_1), \] \hspace{1cm} (4.5)
\[ D(f) = - \prod_{i=1}^{n} (I - 2 S_{g_i(f)} g_i(f)) f, \quad \forall f \in F_1, \] \hspace{1cm} (4.6)

where \((x \otimes y)(f) = (f, y)(x)\). Indeed, observe that
\[ (D(f), g_i(f)) = 2 (f, g_i(f)) - (f, g_i(f)) = (f, g_i(f)), \quad i = 1, 2, \ldots, n. \]

Since \( g_i(f) \) is of the form (2.2) we have \( g_i(D(f)) = g_i \).

\[ g_i(D(f)) = g_i((Df, g_1) = g_i((f, g_1(f)) = g_2(f) \]
and similarly \( g_i(D(f)) = g_i(f) \). Thus \( N^2(D(f)) = N^2(f) \) which proves (4.3). To show (4.4) observe that
\[ D(D(f)) = \sum_{i=1}^{n} (f, g_i(f)) S_{g_i(f)} g_i(f) - D(f) \]
\[ = \sum_{i=1}^{n} (f, g_i(f)) S_{g_i(f)} g_i(f) - f \]
\[ = f. \]

Thus \( D^{-1}(f) = f \) which implies that \( D^{-1}(f) = D(f) \) as claimed.

To show (4.5) observe that \( f \in S_0(F_1) \) implies \( D(f) \in S_0(F_1) \) and \( D(f) = f \) is well defined. We have
\[ D(f) = (S^{-1} g_1(f), D(f)) = \left( \sum_{i=1}^{n} (f, g_i(f)) g_i(f) - S^{-1} f, D(f) \right) \]
\[ = \sum_{i=1}^{n} (f, g_i(f))^2 - \sum_{i=1}^{n} (f, g_i(f)) S^{-1} f, g_i(f) - (S^{-1} f, f) \]
\[ = (S^{-1} f, f) = f. \]
as claimed. Finally observe that

\[ (I - S_n g_1 f) \otimes g_1 f = (I - S_n g_1 f) \otimes g_1 f \]

\[ = f - 2(f \cdot g_1 f) S_n g_1 f - 2(f \cdot g_1 f) S_n g_1 f \]

and the repetitive use of this property yields (4.6).

Property (4.3) means that the mapping \( D \) does not change information, i.e., the elements \( f \) and \( D f \) are indistinguishable under \( X^k \). Property (4.4) means that \( D^2 \) is the identity operator. Property (4.5) means that \( D \) is orthogonal in the norm \( \|
\cdot \|_1 \) and Property (4.6) states the factorization of the operator \( D \).

We show that orthogonal invariance of the measure \( \mu \) implies that the mapping \( D \) does not change the measure of a Borel set.

**Theorem 4.1.** If \( \mu \) is orthogonally invariant then

\[ \mu(D(B)) = \mu(B) \quad (4.7) \]

for any Borel set \( B \).

**Proof.** The elements \( g_i(\cdot) \) which form the adaptive information \( X^k \) are of the form (2.2), i.e., \( g_i : \mathbb{R}^{n-1} \to F_i \). For \( y = [y_1, y_2, \ldots, y_{n-1}] \in \mathbb{R}^{n-1} \) denote \( g_i(y) = g_i(y_1, \ldots, y_{n-1}) \). Since \( g_i \) are measurable, they can be approximated by piecewise constant mappings,

\[ g_i(y) = \lim_{k \to \infty} g_{i,k}(y), \quad \forall y \in \mathbb{R}^{n-1}. \quad (4.8) \]

and \( g_{i,k}(y) = g_{i,k,j} \) for \( y \in A_{k,j} \), where \( A_{k,j} \) are disjoint Borel sets of \( \mathbb{R}^{n-1} \) whose union is \( \mathbb{R}^{n-1} \). Since \( g_i(y) = g_i \) and \( (S_n g_i(y), g_i(y)) = \delta_{y,1} \), we may assume the same properties for \( g_{i,k} \), i.e.,

\[ g_{i,k}(y) = g_i, \]

\[ (S_n g_{i,k}(y), g_{i,k}(y)) = \delta_{y,1} \quad (4.9) \]

for any \( y \in \mathbb{R}^{n-1} \) and any \( k = 1, 2, \ldots \).

Define the mapping

\[ D_k f = \frac{1}{2} \sum_{i=1}^{n} (f \cdot g_{i,k}) S_n g_{i,k} f - f \]

(4.10)

for \( X^k(f) \in A_{k,j} \). Due to (4.8) we have

\[ D(f) = \lim_{k \to \infty} D_k(f), \quad \forall f \in F_1. \quad (4.11) \]

Observe that \( D_k \) is piecewise linear. From (4.9) we have

\[ D_k f = - \sum_{i=1}^{n} (f - 2S_n g_{i,k} f) g_{i,k} f, \quad X^k(f) \in A_{k,j} \]

(4.12)

We now show that

\[ D(B) = E_{\hat{B}} \bigcup_{k=1}^{n} D_k^{-1}(S_k B) \]

(4.13)
for any open set $B$ of $F_i$. Indeed, let $x \in D_i(B)$. Then $x = D_i f, f \in B$. Since $D_i^2 = I$, $f = D_i x_i$. Due to (4.11), $D_i x_i$ approaches $D_i(x) = f \in B$. Since $B$ is open, $D_i x_i \in B$ for $k \geq k_0$. Thus $x \in D_i^{-1}(B)$ for all $k \geq k_0$. This proves (4.13).

Note that $D_i$ is measurable. Therefore $D_i^{-1}(B)$ and $E$ are Borel sets. From (4.13) we have

$$\mu(D_i(B)) \leq \mu(E) = \lim_{k \to \infty} \mu \left( \bigcap_{k=1}^n D_i^{-1}(B) \right) \leq \lim_{k \to \infty} \mu(D_i^{-1}(B)).$$  \hspace{1cm} (4.14)

Let $B_{k,j} = (\wedge k)^{-1} A_{k,j}$. The sets $B_{k,j}$ are disjoint Borel sets and their union is $F_1$. Then

$$\mu(D_i^{-1}(B)) = \sum_{j=1}^n \mu(D_i^{-1}(B \cap B_{k,j})).$$

Note that $D_i^{-1}(B \cap B_{k,j}) = D_i^{-1}(B \cap B_{k,j})$ where

$$D_{k,j}(f) = 2 \sum_{i=1}^n (f, g_{i,k_j}) S_{i,k_j} g_{i,k_j} - f = - \sum_{i=1}^n (I - 2 S_{i,k_j} g_{i,k_j}) f$$

for $f \in F_1$. The mapping $D_{k,j}$ is linear and (4.9) yields that $D_{k,j}^2 = I$. Thus $D_i^{-1} = D_{k,j}$. Orthogonal invariance of $\mu$ yields that $\mu(C) = \mu(C)$ and $\mu(QC) = \mu(C)$ for any Borel set $C$ and $Q = I - 2 S_{i,k_j} g_{i,k_j}$ where $(S_{i,k_j}, h) = 1$. Thus we have

$$\mu(D_i^{-1}(B \cap B_{k,j})) = \mu(D_{k,j}(B \cap B_{k,j}))$$

$$= \mu \left( \bigcap_{i=1}^n \left( I - 2 S_{i,k_j} g_{i,k_j} \right) (B \cap B_{k,j}) \right)$$

$$= \mu \left( \bigcap_{i=1}^n \left( I - 2 S_{i,k_j} g_{i,k_j} \right) (B \cap B_{k,j}) \right)$$

$$= \cdots = \mu(B \cap B_{k,j}).$$

Hence

$$\mu(D_i^{-1}(B)) = \sum_{j=1}^n \mu(B \cap B_{k,j}) = \mu(B).$$

Thus we have

$$\mu(D_i(B)) \leq \mu(B).$$  \hspace{1cm} (4.15)

for any open set $B$.

Take now a closed set $B$. Define $B_s = \{ f \in F_1 : \text{dist}(f, B) < s \}$, $s = 1, 2, \ldots$. Then $B_s$ is open, $B \subset B_{s+1}$, and $B = \bigcap_{s=1}^\infty B_s$. Due to this and (4.15) we have

$$\mu(D_i(B)) \leq \mu(D_i(B_s)) \leq \mu(B_s).$$

Thus $\mu(D_i(B)) \leq \lim \mu(B_s) = \mu(B)$. Hence (4.15) holds also for closed sets.

Take now an open set $B$. Then $F_1 - B$ is closed and

$$1 - \mu(D_i(B)) = \mu(F_1 - B) \leq \mu(F_1 - B) = 1 - \mu(B).$$

Thus $\mu(B) \leq \mu(D_i(B))$. This and (4.15) give

$$\mu(D_i(B)) = \mu(B).$$  \hspace{1cm} (4.16)
for any open set $B$. Since the set of $B$ for which (4.16) holds is a $\sigma$-field and contains all open sets, it contains all Borel sets. This completes the proof. $
$

Theorem 4.1 will be used in the proof of the main result to change variables. That is (4.7) implies that

$$
\frac{1}{\theta} H(f) \mu(df) = \frac{1}{\partial B} H(Df) \mu(df)
$$

for any measurable function $H$ and any Borel set $B$.

In order to prove Theorem 2.1 we need one more result. Let $N^\circ$ be given by (4.1). Define the probability measure $\mu_1(\cdot, N^\circ)$ as

$$
\mu_1(A, N^\circ) = \mu_1((N^\circ)^{\#}(A)) = \mu_1(\{ f \in F_1 : N^\circ(f) \in A \})
$$

where $A$ is a Borel set of $\mathbb{R}^n$. The measure $\mu_1$, called the probability induced by $N^\circ$, tells us the probability that $N(f) \in A$.

We prove that the measure $\mu_1$ is independent of $N^\circ$ and $\mu_1$ is orthogonally invariant with mean zero and the identity covariance operator.

**Theorem 4.2.** There exists a probability measure $\mu_1$ defined on Borel sets of $\mathbb{R}^n$ such that

$$
\mu_1(A, N^\circ) = \mu_1(A), \quad \forall A \in \mathcal{B}(\mathbb{R}^n),
$$

for any measurable adaptive information $N^\circ$ of the form (4.1).

**Proof.** We first consider nonadaptive information operators. Let

$$
N_1(f)[(f, \xi_1, \ldots, f, \xi_n)],
$$

$$
N_2(f)[(f, \eta_1, \ldots, f, \eta_n)]
$$

where $(S, \xi_1, \ldots, \xi_n) = (S, \eta_1, \eta_n) = \delta_{i,j}$. We prove

**Lemma 4.1.** There exists a linear one-to-one mapping $Q : F_1 \rightarrow F_1$ such that

$$
N_1 = N_2 Q
$$

$$
\mu(Q^{-1} B) = \mu(B), \quad \forall B \in \mathcal{B}(F_1).
$$

**Proof.** Let $X = \text{lin}\{S_1^i \xi_1, \ldots, S_n^i \xi_n, S_1^i \eta_1, \ldots, S_n^i \eta_n\}$. Let $p = \dim X$. Of course $p \in [n, 2n]$. There exist elements $\xi_{p+1}, \ldots, \xi_{2n}, \eta_{p+1}, \ldots, \eta_{2n}$ so that $\{S_1^i \eta_1, \ldots, S_n^i \eta_1\}$ and $\{S_1^i \xi_1, \ldots, S_n^i \xi_1\}$ are orthonormal bases of $X$. Define the mapping $H : F_1 \rightarrow F_1$.

$$
H f = \sum_{i=1}^p (f, S_i \eta_i, \xi_i) \xi_i - f.
$$

Since $S_1^i \eta_i = \sum_{i=1}^p (S_1^i \eta_i, S_1^i \xi_i) S_1^i \xi_i$, we get $\eta_i = \sum_{i=1}^p (\eta_i, S_1^i \xi_i) S_1^i \xi_i$ and

$$
H \eta_i = \sum_{i=1}^p (\eta_i, S_i \eta_i) \xi_i - \sum_{i=1}^p (\eta_i, S_i \eta_i) \xi_i - \eta_i = -\eta_i
$$
for $k = 1, 2, \ldots, p$. We define the mapping $Q$ as

$$Q^f = H^*f = \sum_{j=1}^{p} (f, \hat{e}_j) S_j \eta_j + \sum_{i=1}^{r} (f, \hat{e}_i) S_i \gamma_i = f - f_{\min}.$$

To prove (4.19), note that $N_{\gamma} = N_{\gamma} Q$ is equivalent to $(f, \hat{e}_i) = (Qf, \eta_i) = (f, Q^* \eta_i)$

$= (f, H \eta_i)$. This holds since $H \eta_i = \hat{e}_i$ (see (4.21)).

To prove (4.20), we decompose $H$ as

$$H = S_d^{-1} H_I S_d^*$$

where $H_I = \sum_{i=1}^{r} (f, S_i \eta_i, \hat{e}_i) S_i \gamma_i, \neg f$. Note that $H_I S_d^* (F_i) \subset S_d^* (F_i)$ and therefore $S_d^* (H_I S_d^*)$ is well defined. Let $X^\perp$ be an orthogonal complement of $X, F_i = X \otimes X^\perp$. Then $f \in X^\perp$ implies $(f, S_d^* \eta_i) = (f, S_d^* \gamma_i) = 0$ and

$$H_I f = -f, \quad \forall f \in X^\perp. \quad (4.22)$$

From (4.21) we have

$$H_I S_d^* \eta_i = S_d^* \gamma_i, \quad k = 1, 2, \ldots, p.$$}

Thus $H_I$, as well as $-H_I$, restricted to $X$ are orthogonal mappings onto $X$. We decompose $-H_I$ 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

$$-H_I f = D_1 D_2 \cdots D_p f, \quad \forall f \in X. \quad (4.23)$$

where $D_i = I - 2x_i \otimes x_i$.

For $f \in X^\perp$ we have $(f, x_i) = 0$ and we get $D_1 D_2 \cdots D_p f = f$. Thus, (4.23) holds also for $f \in X^\perp$ due to (4.22). Hence we proved that $H = -D_1 D_2 \cdots D_p$ and

$$H = -S_d^{-1} D_1 D_2 \cdots D_p S_d^*$$

$$= -S_d^{-1} (D_1 S_d^*) (D_2 S_d^*) \cdots (D_p S_d^*)$$

$$= Q^* Q^* \cdots Q^*$$

where $Q^* = I - 2h \otimes h$, and $h = S_d^* x_i$. Observe that $Q^* = I - 2S_d^* h \otimes h$. Thus we get

$$Q = -Q^* Q^* \cdots Q^*.$$

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

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

The orthogonal invariance of $\mu$ yields $\mu(Q, B) = \mu(B) = \mu(B)$ for any Borel set $B$ of $F_i$. We have therefore

$$\mu(Q^{-1} B) = \mu(Q, \cdots, Q_p B) = \mu(Q_1, \cdots, Q_p B) = \mu(Q, B) = \cdots = \mu(B),$$

which proves (4.20) and completes the proof of Lemma 4.1. \qed
Define the measure $\mu_1(A)$ as
\[
\mu_1(A) = \mu(A, N_k) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).
\] (1.24)

From Lemma 4.1 we immediately get
\[
\mu_1(A, N_2) = \mu(N_2^{-1}, A) = \mu(Q_1^{-1}, N_2^{-1}, A) \\
= \mu(N_2^{-1}, A) = \mu_1(A, N_1) = \mu_1(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).
\]

Thus (1.48) holds for any non-adaptive information of the form (4.1).

Take now any measurable adaptive information $N^a$. Using (1.8) and (1.9) define
\[
N_1(f) = \{(f, g_1, \ldots, f, g_{n+k})\} \quad \forall f \in F_1.
\]

Let $A$ be an open set of $\mathbb{R}^n$. Then
\[
(N(a))^{-1}(A) \subseteq E \overset{\text{def}}{=} \bigcap_{k=1}^{N_1} N_1^{-1}(A),
\] (1.25)

Indeed, if $f \in (N(a))^{-1}(A)$ then $y = N_1(f) \in A$. Let $y_k = N_1(f_k)$. Then $y_k = y \in A$.

Since $A$ is open, $y_k \in A$ for $k \geq k_0$. Thus $f_k \in N_1^{-1}(y_k) \subseteq N_1^{-1}(A)$ for $k \geq k_0$. This means that $f \in E$ as claimed. From (1.25) we have
\[
\mu_1(A, N^a) = \mu((N(a))^{-1}(A)) \leq \mu(E)
\]

\[
= \lim_{\nu \to 0} \mu\left(\bigcap_{k=1}^{\nu} N_1^{-1}(A)\right) \leq \lim_{\nu \to 0} \mu(N_1^{-1}(A)).
\]

Observe that
\[
\mu(N_1^{-1}(A)) = \sum_{j=1}^{N_1} \mu(N_1^{-1}(A \cap A_{k_j})).
\]

Since $N_1$ on each $A_{k_j}$ coincides with non-adaptive information, we have
\[
\mu(N_1^{-1}(A \cap A_{k_j})) = \mu_1(A \cap A_{k_j})
\]

and
\[
\mu(N_1^{-1}(A)) = \sum_{j=1}^{N_1} \mu_1(A \cap A_{k_j}) = \mu_1(A).
\]

Thus
\[
\mu_1(A, N^a) \leq \mu_1(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).
\] (1.26)

for any open set $A$. Take now a closed set $A$ and define $A_0 = \{x \in \mathbb{R}^{n+1} : \text{dist}(x, A) < 1, i, j = 1, 2, \ldots, n\}$. Then $A \subseteq A_0 \subseteq A_0 = \bigcap_{j=1}^{N_1} A_{k_j}$. Since $A_0$ is open we have due to (1.26),
\[
\mu_1(A, N^a) = \mu((N(a))^{-1}(A)) \leq \mu((N(a))^{-1}(A)) \leq \mu_1(A).
\]
Thus

\[ \mu_1(A, N) \leq \lim \mu_1_s(A) = \mu_1(A). \]

Hence (4.26) holds also for closed sets \( A \). Repeating the last part of the proof of Theorem 4.1 we complete the proof of (4.18). \( \square \)

Theorem 4.2 will be used to compute \( \int_{F_1} H(Nf) \mu(df) \) for any measurable \( H \) and \( N \) of the form (4.11). Due to (4.18) we have

\[ \int_{F_1} H(Nf) \mu(df) = \int_{\mathbb{R}^n} H(y) \mu_1(dy). \]

**Theorem 4.3.** The measure \( \mu_1 \) of Theorem 4.2 is orthogonally invariant with mean zero and the identity covariance operator. \( \square \)

**Proof.** We first show that

\[ (m_{\mu_1}, x) = \int_{\mathbb{R}^n} (y, x) \mu_1(dy) = 0, \quad \forall x = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n. \]  \( (4.27) \)

Take \( \bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_n \) from \( F_1 \) such that \( (S_{\mu}, \bar{\sigma}_j) = \delta_{\cdot j} \). Define

\[ N(f) = [(f, \bar{\sigma}_1), (f, \bar{\sigma}_2), \ldots, (f, \bar{\sigma}_n)]. \]  \( (4.28) \)

Let \( g = \sum_{i=1}^n x_i \bar{\sigma}_i \). Since \( m_{\mu_1} = 0 \), we have

\[ 0 = \int_{F_1} (f, g) \mu(df) = \int_{F_1} \sum_{i=1}^n x_i (f, \bar{\sigma}_i) \mu(df). \]

We change variables by setting \( y = [y_1, \ldots, y_n] = N(f) \). Theorem 4.2 states that \( \mu N^{-1} = \mu_1 \) regardless of \( N \). Thus

\[ 0 = \int_{F_1} \sum_{i=1}^n x_i (y, \mu_1(dy)) = \int_{F_1} (x, y) \mu_1(dy) \]

which proves (4.27). This yields \( m_{\mu_1} = 0 \) as claimed.

To show that \( S_{\mu_1} = I \), we show that

\[ \int_{\mathbb{R}^n} (y, x)(y, z) \mu_1(dy) = (x, z), \quad \forall x, z \in \mathbb{R}^n. \]  \( (4.29) \)

For \( g = \sum_{i=1}^n x_i \bar{\sigma}_i \) and \( h = \sum_{i=1}^n z_i \bar{\sigma}_i \), we have

\[ (S_{\mu}, g, h) = \int_{F_1} (f, g)(f, h) \mu(df) = \int_{\mathbb{R}^n} (y, x)(y, z) \mu_1(dy). \]

Since \( (S_{\mu}, g, h) = \sum_{i,j=1}^n x_i z_j (S_{\mu}, \bar{\sigma}_i, \bar{\sigma}_j) = (x, z) \), (4.29) follows.

We now prove that \( \mu_1 \) is orthogonally invariant, i.e.,

\[ \mu_1(Q(B)) = \mu_1(B), \quad \forall B \in \mathcal{B} \cap \mathbb{B}^n \]
where \( Q = 2(y, x) x - y, y \in \mathbb{R}^n, x = 1 \) or \( x = 0 \). Define the mapping
\[
Df = 2(f, g) \tilde{s}_x g - f, \quad f \in F_1.
\]
where, as before, \( g = \sum_{i=1}^{n} x_{i} \tilde{s}_{x_i} \). Then \( (\tilde{s}_x, g) = \sum_{i=1}^{n} x_{i} \tilde{s}_{x_i} = 1 \) or \( g = 0 \). Observe that
\[
N^{-1} Q B = D N^{-1} B, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).
\]
Indeed, \( f \in N^{-1} Q B \) iff \( N f \in Q B \) iff \( N f \in B \) since \( Q^2 = I \). Similarly \( f \in D N^{-1} B \) iff \( N D f \in B \) since \( D^2 = I \). Note that
\[
Q N f = 2(N f, x) x - N f = 2(f, g) x - N f,
\]
\[
N D f = 2(f, g) N(\tilde{s}_x g) - N f = 2(f, g) x - N f,
\]
which proves (4.30). From Theorem 4.2, (4.30) and orthogonal invariance of \( \mu \), we have
\[
\mu_1(Q(B)) = \mu(N^{-1} Q(B)) = \mu(D N^{-1} B) = \mu(N^{-1} B) = \mu_1(B)
\]
as claimed. This completes the proof of Theorem 4.3.

5. Proof of the Main Result

Using properties of orthogonally invariant measures we are ready to prove that adaption does not help on the average.

The proof consists of two steps. The first step is to show that the spline algorithm that uses \( N^* \) has minimal average error among all algorithms that use \( N^* \). The second step is to estimate from below the average radius of information for equivalently the average error of the spline algorithm.

Let
\[
N^* f = [(f, g_1(f)), (f, g_2(f)), \ldots, (f, g_n(f))]
\]
be a measurable adaptive information operator of the form (2.1) and (2.2). Thus \( g_1(f), \ldots, g_n(f) \) are measurable. First of all we show that without loss of generality we can assume that
\[
(\tilde{s}_x g(f), g(f)) = \delta_{i,j} \quad (5.1)
\]
Indeed, as in [16] let \( \eta_{i}(f), \ldots, \eta_n(f) \) be an orthonormal basis of the linear space \( \text{lin} \{ \tilde{s}_x g_1(f), \ldots, \tilde{s}_x g_n(f) \} \). Then there exists a nonsingular matrix \( M \) such that
\[
[(f, \tilde{g}_1(f)), \ldots, (f, \tilde{g}_n(f))] = N^* f M
\]
where \( \tilde{g}_i(f) = \tilde{s}_x^{-1} \eta_i(f) \) and \( (\tilde{s}_x g_i(f), \tilde{g}_i(f)) = \delta_{i,j} \). Thus, knowing \( N^* f \) we can compute \( (f, \tilde{g}_i(f)) \). The mappings \( \tilde{g}_i(f) \) are also measurable. The elements \( \tilde{g}_i(f) \) then play the role of \( g_i(f) \). This explains (5.1).

Define
\[
\sigma = \sigma(N^* f) = \sum_{i=1}^{n} (f, g_i(f)) \tilde{s}_x g_i(f), \quad (5.2)
\]
Note that $N^4(\sigma) = N^4(f)$, i.e., $\sigma$ interpolates $f$. Furthermore, take $h \in S_i(F_i)$ such that $N^4(h) = N^4(f)$. Then
\[ h - \sigma = \sigma - h = \sigma - h = 2S_i^{-1} \sigma, h - \sigma. \]

Since $(h - \sigma, g_i(f)) = 0$ we have $\|S_i^{-1} \sigma, h - \sigma\| = 0$ and $h \not\parallel \sigma$. Thus $\sigma$ has minimal norm $\| \cdot \|$ among elements which interpolate $f$ and lie in $S_i(F_i)$. Such an element is called a spline interpolating $f$. Let
\[ \phi^i(N^4(f)) = \sum_{i=1}^{n} (f, g_i(f)) S_i g_i(f) \]
be the spline algorithm.

We say an algorithm $\phi$ is an optimal average error algorithm iff
\[ \epsilon^{\text{av}}(\phi, N^4) = \epsilon^{\text{av}}(N^4). \]

**Theorem 5.1.** If $\mu$ is orthogonally invariant then the spline algorithm $\phi^i$ is an optimal average error algorithm and
\[ \epsilon^{\text{av}}(\phi^i, N^4) = \epsilon^{\text{av}}(N^4) = \frac{1}{f_i} \int_{f_i} S_f - \phi_i(N^4(f)) \|\mu(df). \quad (5.5) \]

**Proof.** The proof is essentially the same as the proof of Theorem 4.3 of [16]. For completeness we provide a sketch of it.

Orthogonal invariance of $\mu$ and (4.3) yield
\[ \frac{1}{f_i} \int_{f_i} S_f - \phi_i(N^4(f)) \|\mu(df) = \frac{1}{f_i} \int_{f_i} S_D(f_i) - \phi_i(N^4(f)) \|\mu(df) \]
where $\phi$ is an algorithm and $D$ is the mapping defined by (4.2); see Theorem 4.1. Thus
\[ \epsilon^{\text{av}}(\phi, N^4) = \frac{1}{f_i} \int_{f_i} S_f - \phi_i(N^4(f)) \|\mu(df) \]
\[ = \frac{1}{f_i} \int_{f_i} S_D(f_i) - \phi_i(N^4(f)) \|\mu(df). \]

Since $S_f - \phi_i(N^4(f)) \| \geq 1/\psi \int_{f_i} S_f - \phi_i(N^4(f)) \| = SD(f_i) - \phi_i(N^4(f)) \|^2$ we get

**This means that $\phi^i$ is an optimal average error algorithm.** To prove (5.5) note that
\[ a = S_f - \phi_i(N^4(f)) \|^2 = SD(f_i) - \phi_i(N^4(f)) \|^2 \]
\[ = S_f \|^2 - SD(f_i) \|^2 - 2(\phi_i(N^4(f))) \|^2 - 2(S_D(f_i), \phi_i(N^4(f))) \]
Since $D(f_i) = \phi_i(N^4(f)) - f$, then $SD(f_i) = 2\phi_i(N^4(f)) - f$ and
\[ a = S_f \|^2 - SD(f_i) \|^2 - 2(\phi_i(N^4(f))) \|^2. \]
This and (5.6) with $\phi = \phi^*$ yield (5.7).

Proof of Theorem 2.1. The radius $r^{*}(N \cdot N^{-1})$ is given by (5.5). In order to estimate it from below, note that Theorem 4.2 yields

$$
\int_{F_{1}} \phi^*(N \cdot N^{-1}) f \mu(dy) \leq \int_{\mathbb{R}^n} \phi^*(y) \cdot \mu_1(dy)
$$

$$
= \sum_{i=1}^{n} \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) SS_{g_i}g_i(y) \mu_1(dy)
$$

(5.7)

where, as in Sect. 4, $g_i(y) = g, y_1, y_2, \ldots, y_{i-1}$. Define the mapping

$$
Qy = y - 2(y, e_i)e_i = y - 2y_i e_i, \quad y \in \mathbb{R}^n,
$$

where $e_i$ is the ith unit vector. Then $Qe_i = -e_i$ and $Qe_j = e_j$. This yields $g_i(Qy) = g_i(y)$ for $j < i$. Since $\mu_1$ is orthogonally invariant we have

$$
a = \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) SS_{g_i}g_i(y) \mu_1(dy)
$$

$$
= \int_{\mathbb{R}^n} y, y_i Qe_i, Qe_i SS_{g_i}g_i(Qy) SS_{g_i}g_i(Qy) \mu_1(dy)
$$

$$
= -a.
$$

Hence $a = 0$ and (5.7) becomes

$$
\int_{F_{1}} \phi^*(N \cdot N^{-1}) f \mu(dy) = \sum_{i=1}^{n} \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) \mu_1(dy).
$$

(5.8)

For $i < n$ define the mapping

$$
Qy = y - 2y_i h, \quad h = (e_i, -e_i) \sqrt{2}, \quad y \in \mathbb{R}^n.
$$

Note that $h = 1$ and $Qe_j = e_j$ for $j < i$ and $Qe_i = e_i$. Then $g_i(Qy) = g_i(y)$ and

$$
\int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) \mu_1(dy) = \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) \mu_1(dy).
$$

From this, (5.8) and (2.11) we have

$$
\int_{F_{1}} \phi^*(N \cdot N^{-1}) f \mu(dy) \leq \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i(y) \mu_1(dy)
$$

$$
\leq (\sup_{y \in \mathbb{R}^n} SS_{g_i}g_i(y)^2) \int_{\mathbb{R}^n} y, y_i \mu_1(dy) = \sum_{i=1}^{n} SS_{g_i}g_i(y)^2 \mu_1(dy).
$$

(5.9)

For the nonadaptive information $N_{f}^{(m)}$, see (2.12), we have

$$
\int_{F_{1}} \phi^*(N \cdot N^{-1}) f \mu(dy) = \sum_{i=1}^{n} \int_{\mathbb{R}^n} y, y_i SS_{g_i}g_i, SS_{g_i}g_i \mu_1(dy)
$$

$$
= \sum_{i=1}^{n} SS_{g_i}g_i^2.
$$

(5.10)
From (5.5) of Theorem 5.1, (5.9) and (5.10) we have

$$r^{*4}(N^{*})^{2} = r^{*4}(N^{*})^{2} \geq \int_{F} \left[ \sum_{i=1}^{n} S_{i} \cdot g_{i}^{*} \right]^{2}$$

This completes the proof. \(\square\)

Acknowledgments: Our special thanks are to J. F. Traub for his valuable comments and friendly advice. We are also grateful to the referee for suggestions on how to simplify the concepts presented in this paper.

References

8. Sikorski, K., Wozniakowski, H.: For which error criteria can we solve nonlinear equations? Dept. of Computer Science Report, Columbia University, 1983

Received November 17, 1982 November 21, 1983 February 14, 1984