Numer. Math. 44, 169-190 (1984)

Numerische Mathematik

CUCS-80-83

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 μ. It is known that adaption does not help in the worst case for linear problems. We prove that adaption 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.

Subject Classifications: AMS(MOS): 68C25, CR: F2.1.

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_i, $N(f)=[f(t_1), f(t_2), \ldots, f(t_n)]$, 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 $N = N^{non}$ is called *nonadaptive* information. If the second point t_2 depends on the previously computed value $f(t_1)$, i.e., $t_2 = t_2(f(t_1))$ and if the point t_i depends on the previously computed values $f(t_1), \ldots, f(t_{i-1})$ i.e., $t_i = t_i(f(t_1), \ldots, f(t_{i-1}))$, then $N = N^a$ 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-

Address reprint requests to: H. Woźniakowski, Department of Computer Science, Columbia University, New York, NY 10027, USA

* 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 ϕ (for our simple example) is defined by

$$e(\phi, N) = \sup_{f \in F} \left| \int_0^1 f(t)\,dt - \phi(N(f)) \right|. \tag{1.1}$$

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

$$r(N) = \inf_{\phi} e(\phi, N) \tag{1.2}$$

is called the *radius of information* and ϕ is optimal iff $e(\phi, N) = r(N)$.

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

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

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 $N^{\text{non}}(f) = [f(t_1^*), \ldots, f(t_n^*)]$ is as powerful as the adaptive information N^a, i.e.,

$$r(N^{\text{non}}) \leq r(N^a).$$

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 $[a, b]$. Then the optimal nonadaptive information has radius $(b-a)/(2(n+1))$ whereas the optimal adaptive information is bisection information which has radius $(b-a)2^{-(n+1)}$; 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.,

$$e^{\text{avg}}(\phi, N) = \left\{ \int_F \left| \int_0^1 f(t)\,dt - \phi(N(f)) \right|^2 \mu(df) \right\}^{1/2} \tag{1.3}$$

where μ 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

$$r^{avg}(N) = \inf_{\phi} e^{avg}(\phi, N) \tag{1.4}$$

as the *average radius of information.*

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

$$r^{avg}(N^a) < r(N^{non})?$$

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 adaption does not help on the average for linear problems in infinite dimensional Hilbert spaces with an "orthogonally invariant" measure μ. Orthogonal invariance of μ 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 μ 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 $f(t_1)$, $f(t_2), \ldots, f(t_i)$ (or some of them) satisfy a certain Boolean condition, i.e.,

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

for some $a_{1,i}$ and $a_{2,i}$. 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 adaption 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 adaption 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 then nonadaption? There are a number of reasons which include:

(i) Intrinsic mathematical interest. Adaption corresponds to certain nonlinear operators whereas nonadaption corresponds to linear operators. Mathematically the sentence "adaption does not help" means that this nonlinearity is no more powerful than linearity.

(ii) Reduction of the search for optimal information. If adaption does not help then we only have to look at the very special and relatively easy nonadaptive 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 μ 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

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

where $y_1 = (f, g_1)$, $y_i = (f, g_i(y_1, \ldots, y_{i-1}))$, $g_i(y_1, \ldots, y_{i-1})$ 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

$$g_1(f) = g_1, \quad g_i(f) = g_i(y_1, \ldots, y_{i-1}), \quad 2 \leq i \leq n. \tag{2.2}$$

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., $g_i(f) \equiv g_i$ for some g_i from F_1, then N is called *nonadaptive* and denoted by $N = N^{non}$, i.e.,

$$N^{non}(f) = [(f, g_1), (f, g_2), \ldots, (f, g_n)]. \tag{2.3}$$

Note that nonadaptive information is a *linear* operator whereas adaptive information is in general *nonlinear*. Without loss of generality we assume that $g_1(f), g_2(f), \ldots, g_n(f)$ are linearly independent for each f from F_1.

Knowing $N(f)$ we approximate Sf by $\varphi(N(f))$ where φ is a mapping from $N(F_1)$ into F_2. We call such φ an (idealized) *algorithm*. We wish to approximate Sf with an average error as small as possible. The average error of φ is defined as

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

Here μ 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 φ is chosen such that (2.4) is well defined, i.e., $\|Sf - \varphi(N(f))\|^2$ is a measurable function. This assumption is not restricted as is shown in [15]. Let

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

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 μ. 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 μ is zero, i.e.

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

Let S_μ be the *covariance operator* of μ, 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_μ 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_μ^{-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 $\|x\|_* = \sqrt{(x, x)_*} = \sqrt{(S_\mu^{-1} x, x)}$.

We say μ is *orthogonally invariant* iff

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

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 $(S_\mu h, h)=1$ or $h=0$. For $h=0$, $Qf=-f$ and (2.8) means that $\mu(-B)=\mu(B)$ where $-B=\{f: -f\in B\}$. Note that $f\in S_\mu(F_1)$ implies that $Qf\in S_\mu(F_1)$ and

$$\|Qf\|_*^2=(2(f,h)h-S_\mu^{-1}f, 2(f,h)S_\mu h-f)$$
$$=(S_\mu^{-1}f,f)=\|f\|_*^2.$$

Thus the mapping Q is orthogonal in the norm $\|\cdot\|_*$. This explains why μ is called orthogonally invariant.

It is shown in [16] that Gaussian measures are orthogonally invariant as well as measures of the form

$$\mu(B)=\int_B w(\|S_\mu^{-\frac{1}{2}}f\|)\lambda(df)$$

for some measurable function w assuming that F_1 is finite dimensional and λ is the Lebesgue measure.

Note that Q resembles a Householder matrix. It is easy to check that

$$Q^2=I, \quad Q^{-1}=Q. \tag{2.10}$$

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 $(S_\mu g_i(f), g_j(f))=\delta_{i,j}$. Let

$$a=\sup\left\{\sum_{i=1}^n \|SS_\mu g_i(f)\|^2: f\in F_1\right\}.$$

For simplicity assume that a is obtained for f^*, i.e.,

$$\sum_{i=1}^n \|SS_\mu g_i(f^*)\|^2=\sup_{f\in F_1}\sum_{i=1}^n \|SS_\mu g_i(f)\|^2. \tag{2.11}$$

Let $g_i^*=g_i(f^*)$. By N_f^{non} we mean

$$N_f^{\text{non}}(f)=[(f,g_1^*),(f,g_2^*),\ldots,(f,g_n^*)]. \tag{2.12}$$

Note that N_f^{non} is *nonadaptive* and is obtained by fixing $g_i(f)$ in the adaptive information N^a.

We say that $N^a(f)=[(f,g_1(f)),\ldots,(f,g_n(f))]$ is measurable iff $g_i(\cdot)$ is measurable, i.e., $g_i^{-1}(B)$ is a Borel set for a Borel set B of $\mathbb{R}^{i-1}$, $i=2,3,\ldots,n$. We are ready to state the main result of this paper.

Theorem 2.1. *Let μ be an orthogonally invariant measure. Let N^a be measurable adaptive information. Then*

$$r^{\text{avg}}(N^a)\geqq r^{\text{avg}}(N_f^{\text{non}}). \quad \square \tag{2.13}$$

Thus adaption 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 $\mu(N^a)^{-1}$ are orthogonally invariant and independent of N^a. 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 μ to be orthogonally invariant whenever μ is absolutely continuous with respect to the Lebesgue measure λ. Next we consider the general case, $\dim(F_1) \leqq +\infty$. We show that orthogonal invariance of μ 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_μ is bounded and

$$T = S_\mu^{-\frac{1}{2}}: F_1 \to F_1 \tag{3.1}$$

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

$$\mu(T^{-1}(2\eta \otimes \eta - I)TB) = \mu(B) \tag{3.2}$$

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

We characterize orthogonally invariant measures μ which are absolutely continuous with respect to the Lebesgue measure λ. Recall that μ is absolutely continuous w.r.t. to λ (denoted by $\mu \ll \lambda$) iff $\lambda(B) = 0 \Rightarrow \mu(B) = 0$ 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

$$\mu(B) = \int_B g(f)\,\lambda(df). \tag{3.3}$$

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 μ is orthogonally invariant* iff

$$g(f_1) = g(f_2) \quad \textit{for any } f_1, f_2 \in A \quad \textit{such that} \quad \|f_1\|_* = \|f_2\|_*. \quad \square \tag{3.4}$$

Proof. Suppose μ is orthogonally invariant. Take f_1 and f_2 from A such that $\|f_1\|_* = \|f_2\|_*$. Define $\eta = T(f_1 - f_2)/\|f_1 - f_2\|_*$ for $f_1 \neq -f_2$, and $\eta = 0$ for $f_1 =$

$-f_2$. Let $Q=T^{-1}(2\eta\otimes\eta-I)T$. We have $2(Tf_1, T(f_1-f_2))=\|f_1-f_2\|_*^2$ and $Qf_1=f_2$. Then (3.2) yields $\mu(QB)=\mu(B)$ for any Borel set B. Observe that $|\det Q|=1$. This and (3.3) yield

$$\int_B \{g(f)-g(Qf)\}\,\lambda(df)=0, \qquad \forall B-\text{Borel set}. \tag{3.5}$$

Note that $g-gQ$ is continuous at f_1 and $g(f_1)-g(Qf_1)=g(f_1)-g(f_2)$. Suppose that $g(f_1)-g(f_2)\neq 0$. 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 $\operatorname{sign}(g(f)-g(Qf))\equiv\text{constant}$. 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 $\|\cdot\|_*$ we have $|\det Q|=1$ and

$$\begin{aligned}\mu(Q(B))&=\int_{Q(B)} g(f)\,\lambda(df)=\int_{Q(B)\cap A} g(f)\,\lambda(df)\\ &=\int_{B\cap Q(A)} g(Qf)\,\lambda(df)=\int_{B\cap Q(A)\cap A} g(Qf)\,\lambda(df).\end{aligned}$$

Note that $f\in A\cap Q(A)$ implies $Qf\in A$. Since $\|Qf\|_*=\|f\|_*$ then (3.4) yields $g(Qf)=g(f)$. Thus we have

$$\mu(Q(B))=\int_{B\cap Q(A)\cap A} g(f)\,\lambda(df)\leqq\mu(B) \tag{3.6}$$

for any Borel set B. Setting $B=Q(C)$ we have $\mu(C)\leqq\mu(Q(C))$ for any Borel set C. Hence $\mu(Q(B))=\mu(B)$. This means that μ is orthogonally invariant. □

The condition (3.4) means that g depends on the norm of $\|f\|_*$. More precisely, let $X=\{\|f\|_*: f\in A\}$. Define $w:\mathbb{R}_+\to\mathbb{R}_+$ such that

$$w(x)=\begin{cases} g(f), & x\in X,\\ 0 & x\notin X\end{cases} \tag{3.7}$$

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

$$\begin{aligned}\mu(B)&=\int_{B\cap A} g(f)\,\lambda(df)=\int_{B\cap A} w(\|f\|_*)\,\lambda(df)\\ &=\int_B w(\|f\|_*)\,\lambda(df).\end{aligned}$$

Thus we have proven

Corollary 3.1. *The measure μ is orthogonally invariant* iff

$$\mu(B)=\int_B w(\|f\|_*)\,\lambda(df), \qquad \forall B-\text{Borel set}. \quad \square \tag{3.8}$$

The measures considered in [10] are of the form (3.8) and therefore they are orthogonally invariant.

(ii) We now turn to the general case $\dim(F_1) \leqq +\infty$. If $\dim(F_1) = +\infty$ then $T = S_\mu^{-1/2}$ is unbounded and for $f \notin S_\mu^{1/2}(F_1)$, Tf is not well defined. Therefore the results of subsection (i) do not hold.

We exhibit relations between orthogonal invariance of μ and orthogonal invariance of its finite dimensional projections. Let $\zeta_1, \zeta_2, \ldots$ be orthonormal eigenelements of the covariance operator S_μ, i.e.

$$S_\mu \zeta_i = \lambda_i \zeta_i, \tag{3.9}$$

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

$$P_m : F_1 \to X_m. \tag{3.10}$$

Let μ_m be the projection of the measure μ onto X_m, i.e.

$$\mu_m(B) = \mu(P_m^{-1} B) \tag{3.11}$$

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

Theorem 3.2. *The measure μ is orthogonally invariant* iff *the measures μ_m are orthogonally invariant for $m = 1, 2, \ldots$.* □

Proof. Assume that μ 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 μ_m and $h \in X_m$, $(S_m h, h) = 1$ or $h = 0$. First of all we show that $S_m x = S_\mu x$, $x \in X_m$. Indeed, for $x, y \in X_m$ we have

$$(S_m x, y) = \int_{X_m} (f, x)(f, y)\mu_m(df) = \int_{F_1} (P_m f, x)(P_m f, y)\mu(df)$$
$$= (S_\mu P_m x, P_m y) = (S_\mu x, y).$$

Since X_m is an invariant subspace of S_μ, $S_\mu x \in X_m$ and $S_m x = S_\mu x$, $\forall x \in X_m$, as claimed. Thus Q can be extended to the space F_1 with S_m replaced by S_μ. Let B be a Borel set in X_m. Note that

$$P_m^{-1} Q B = Q P_m^{-1} B. \tag{3.12}$$

Indeed, if $f \in P_m^{-1} Q B$ then $f = Qb + f_1$ where $b \in B$ and $f_1 \in X_m^\perp$. Since $Qf_1 = -f_1$, we have $f = Q(b - f_1) \in Q(P_m^{-1} B)$. Assume now that $f \in Q P_m^{-1} B$. Then $f = Q(b - f_1)$ where $b \in B$ and $f_1 \in X_m^\perp$. Thus $f = Qb - f_1 \in P_m^{-1} Q B$ as claimed.

From (3.11), (3.12) and orthogonal invariance of μ we have

$$\mu_m(QB) = \mu(P_m^{-1} QB) = \mu(Q P_m^{-1} B) = \mu(P_m^{-1} B) = \mu_m(B). \tag{3.13}$$

Thus μ_m is orthogonally invariant which completes this part of the proof.

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

$$Z = \{B : B \text{ is a Borel set in } F_1,\ \mu(QB) = \mu(B)\}. \tag{3.14}$$

Observe that Z is a σ-field. Indeed, if $B_i \in Z$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, then $QB_i \cap QB_j = \emptyset$ since Q is one-to-one. Then

$$\mu\Big(Q \bigcup_{i=1}^{\infty} B_i\Big) = \mu\Big(\bigcup_{i=1}^{\infty} QB_i\Big) = \sum_{i=1}^{\infty} \mu(QB_i) = \sum_{i=1}^{\infty} \mu(B_i) = \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big).$$

Thus $\bigcup_{i=1}^{\infty} B_i \in Z$. Of course $\emptyset \in Z$ and $B \in Z$ implies that $F_1 - B \in Z$. Hence Z is a σ-field as claimed.

We now show that each closed ball $B = \{f: \|f-a\| \leq r\}$ with $a \in X_{m_0}$ for some m_0, belongs to Z. Recall that $Qf = 2(f,h)S_u h - f$ where $(S_u h, h) = 1$ or $h = 0$. If $h \neq 0$ then take an index j such that $P_j h \neq 0$. Define $h_j = c P_j h$ where $c = (S_u P_j h, P_j h)^{-1/2}$. If $h = 0$, set $h_j = 0$. Then $h_j \in X_j$ and $(S_u h_j, h_j) = 1$ of $h_j = 0$.

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

$$Q_j(f) = 2(f, h_j) S_u h_j - f.$$

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

$$\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} P_j^{-1} Q_j(P_j B) \subset QB. \tag{3.15}$$

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

$$Q_{j_i} P_{j_i} x = P_{j_i} b_{j_i} \in P_{j_i} B.$$

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

$$\mu(QB) \geq \lim_{i \to \infty} \mu\Big(\bigcup_{j=i}^{\infty} P_j^{-1} Q_j P_j B\Big) \geq \lim_{i \to \infty} \mu(P_i^{-1} Q_i P_i B)$$
$$= \lim_{i \to \infty} \mu_i(Q_i P_i B).$$

Since μ_i is orthogonally invariant then

$$\mu_i(Q_i P_i B) = \mu_i(P_i B) = \mu(P_i^{-1} P_i B) \geq \mu(B).$$

Thus

$$\mu(QB) \geq \mu(B). \tag{3.16}$$

To prove the opposite inequality we show that

$$QB \subset \bigcup_{i=1}^{\infty} \bigcap_{j=i}^{\infty} P_j^{-1} Q_j P_j B. \tag{3.17}$$

where $B_\varepsilon = \{f: \|f-a\| \le r+\varepsilon\}$. Indeed, $x \in QB$ means that $x = Qb$ and $\|b-a\| \le r$. Note that Q_jP_jQb tends to $Q^2b = b$ as $j \to +\infty$. Thus there exists an index $j_0 = j_0(b)$ such that $\|Q_jP_jQb - a\| \le r+\varepsilon$ for $j \ge j_0$. Hence $Q_jP_jQb \in B_\varepsilon$. Since $Q_jX_j = X_j$ then $Q_jP_jQb = P_jQ_jP_jQb \in P_jB_\varepsilon$. Since $Q_j^2 = I$ we have $P_jQb \in Q_jP_jB_\varepsilon$ and $Qb \in P_j^{-1}Q_jP_jB_\varepsilon$ for $j \ge j_0$. Thus $x = Qb \in \bigcap_{j=j_0}^{\infty} P_j^{-1}Q_jP_jB_\varepsilon$ which completes the proof of (3.17). From this we have

$$\mu(QB) \le \lim_{i\to\infty} \mu\left(\bigcap_{j=i}^{\infty} P_j^{-1}Q_jP_jB_\varepsilon\right) \le \lim_{i\to\infty} \mu(P_i^{-1}Q_iP_iB_\varepsilon)$$
$$= \lim_{i\to\infty} \mu_i(Q_iP_iB_\varepsilon) = \lim_{i\to\infty} \mu_i(P_iB_\varepsilon). \tag{3.18}$$

We now show that

$$B_\varepsilon = \bigcap_{i=1}^{\infty} P_i^{-1}P_iB_\varepsilon. \tag{3.19}$$

Since $B_\varepsilon \subset P_i^{-1}P_iB_\varepsilon$, $\forall i$, it is enough take $x \in \bigcap_{i=1}^{\infty} P_i^{-1}P_iB_\varepsilon$ and show that $x \in B_\varepsilon$. We have $P_ix \in P_iB_\varepsilon$ and since $P_ia = a$ for $i \ge m_0$ we get $P_ix \in B_\varepsilon$. Note that P_ix tends to x and B_ε is closed which yields that $x \in B_\varepsilon$ as claimed. Since $P_{i-1}^{-1}P_{i-1}B_\varepsilon \subset P_i^{-1}P_iB_\varepsilon$ then (3.19) yields

$$\mu(B_\varepsilon) = \lim_{i\to\infty} \mu_i(P_iB_\varepsilon).$$

This and (3.18) yield

$$\mu(QB) \le \mu(B_\varepsilon). \tag{3.20}$$

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

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

for any closed ball with center lying in X_{m_0} for some m_0.

Thus $B \in Z$. Since any closed ball $A = \{f: \|f-a\| \le r\} = \bigcap_{k=1}^{\infty} \{f: \|f - P_ka\| \le r + \|(I-P_k)a\|\}$ and Z is a σ-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 μ is orthogonally invariant. This completes the proof. □

4. Properties of Orthogonally Invariant Measures

The proof of Theorem 2.1 depends on properties of the orthogonally invariant measure μ which will be obtained in this section.

Let

$$N^a(f) = [(f, g_1(f)), (f, g_2(f)), \dots, (f, g_n(f))] \tag{4.1}$$

be measurable adaptive information. This means that $g_1(\cdot), \dots, g_n(\cdot)$ are measurable and are of the form (2.2). Assume that $(S_u g_i(f), g_j(f)) = \delta_{i,j}$, $\forall f \in F_1$. (We show in Sect. 5 that this assumption is not restrictive.)

Define the mapping $D: F_1 \to F_1$ by

$$D(f) = 2 \sum_{i=1}^{n} (f, g_i(f)) S_u g_i(f) - f. \tag{4.2}$$

The mapping D plays an important role in our analysis. Observe that D is measurable. For nonadaptive information, i.e., $g_i(f) \equiv g_{0,i}$, D is linear. For adaptive information D is nonlinear. The mapping D has four important properties

$$N^a(D(f)) = N^a(f), \tag{4.3}$$

$$D^{-1} = D, \tag{4.4}$$

$$\|D(f)\|_* = \|f\|_*, \quad \forall f \in S_u(F_1), \tag{4.5}$$

$$D(f) = -\prod_{i=1}^{n} (I - 2 S_u g_i(f) \otimes g_i(f)) f, \quad \forall f \in F_1, \tag{4.6}$$

where $(x \otimes y)(f) = (f, y)x$. Indeed, observe that

$$(D(f), g_i(f)) = 2(f, g_i(f)) - (f, g_i(f)) = (f, g_i(f)), \quad i = 1, 2, \dots, n.$$

Since $g_i(f)$ is of the form (2.2) we have $g_1(D(f)) = g_1$,

$$g_2(D(f)) = g_2((Df, g_1)) = g_2((f, g_1)) = g_2(f)$$

and similarly $g_i(D(f)) = g_i(f)$. Thus $N^a(D(f)) = N^a(f)$ which proves (4.3). To show (4.4) observe that

$$\begin{aligned} D(D(f)) &= 2 \sum_{i=1}^{n} (D(f), g_i(D(f))) S_u g_i(D(f)) - D(f) \\ &= 2 \sum_{i=1}^{n} (f, g_i(f)) S_u g_i(f) = D(f) + f - D(f) \\ &= f. \end{aligned}$$

Thus $D^2(f) = f$ which implies that $D^{-1}(f) = D(f)$ as claimed.

To show (4.5) observe that $f \in S_u(F_1)$ implies $D(f) \in S_u(F_1)$ and $\|D(f)\|_*$ is well defined. We have

$$\begin{aligned} \|D(f)\|_*^2 &= (S_u^{-1} D(f), D(f)) = \left(2 \sum_{i=1}^{n} (f, g_i(f)) g_i(f) - S_u^{-1} f, D(f)\right) \\ &= 2 \sum_{i=1}^{n} (f, g_i(f))^2 - 2 \sum_{i=1}^{n} (f, g_i(f))(S_u^{-1} f, S_u g_i(f)) + (S_u^{-1} f, f) \\ &= (S_u^{-1} f, f) = \|f\|_*^2 \end{aligned}$$

as claimed. Finally observe that

$$(I - S_\mu g_i(f) \otimes g_i(f))(I - S_\mu g_j(f) \otimes g_j(f))f$$
$$= f - 2(f, g_i(f))S_\mu g_i(f) - 2(f, g_j(f))S_\mu g_j(f),$$

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 N^a. Property (4.4) means that D^2 is the identity operator. Property (4.5) means that D is orthogonal in the norm $\|\cdot\|_*$ and Property (4.6) states the factorization of the operator D.

We show that orthogonal invariance of the measure μ implies that the mapping D does not change the measure of a Borel set.

Theorem 4.1. *If μ is orthogonally invariant then*

$$\mu(D(B)) = \mu(B) \tag{4.7}$$

for any Borel set B. □

Proof. The elements $g_i(\cdot)$ which form the adaptive information N^a are of the form (2.2), i.e., $g_i: \mathbb{R}^{i-1} \to F_1$. For $y = [y_1, y_2, \ldots, y_{n-1}] \in \mathbb{R}^{n-1}$ denote $g_i(y) = g_i(y_1, \ldots, y_{i-1})$. Since g_i are measurable, they can be approximated by piecewise constant mappings,

$$g_i(y) = \lim_k g_{i,k}(y), \qquad \forall\, y \in \mathbb{R}^{n-1}, \tag{4.8}$$

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 $\mathbb{R}^{n-1}$, $j = 1, 2, \ldots, n_k$. Since $g_1(y) = g_1$ and $(S_\mu g_i(y), g_j(y)) = \delta_{i,j}$ we may assume the same properties for $g_{i,k}$, i.e.,

$$\begin{aligned} g_{1,k}(y) &= g_1, \\ (S_\mu g_{i,k}(y), g_{j,k}(y)) &= \delta_{i,j} \end{aligned} \tag{4.9}$$

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

Define the mapping

$$D_k(f) = 2 \sum_{i=1}^{n} (f, g_{i,k,j}) S_\mu g_{i,k,j} - f \tag{4.10}$$

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

$$D(f) = \lim_k D_k(f), \qquad \forall f \in F_1. \tag{4.11}$$

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

$$D_k(f) = -\prod_{i=1}^{n} (I - 2S_\mu g_{i,k,j} \otimes g_{i,k,j})f, \qquad N^a(f) \in A_{k,j}. \tag{4.12}$$

We now show that

$$D(B) \subset E \stackrel{\text{df}}{=} \bigcup_{i=1}^{\infty} \bigcap_{k=i}^{\infty} D_k^{-1}(B) \tag{4.13}$$

for any open set B of F_1. Indeed, let $x\in D(B)$. Then $x=D(f)$, $f\in B$. Since $D^2=I$, $f=D(x)$. Due to (4.11), $D_k(x)$ approaches $D(x)=f\in B$. Since B is open, $D_k(x)\in B$ for $k\geqq k_0$. Thus $x\in D_k^{-1}(B)$ for all $k\geqq k_0$. This proves (4.13).

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

$$\mu(D(B))\leqq\mu(E)=\lim_{i\to\infty}\mu\left(\bigcap_{k=i}^{\infty}D_k^{-1}(B)\right)\leqq\lim_{k\to\infty}\mu(D_k^{-1}(B)). \tag{4.14}$$

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

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

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

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

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

$$\begin{aligned}\mu(D_k^{-1}(B\cap B_{k,j}))&=\mu(D_{k,j}(B\cap B_{k,j}))\\ &=\mu\left(\prod_{i=1}^{n}(I-2S_\mu g_{i,k,j}\otimes g_{i,k,j})B\cap B_{k,j}\right)\\ &=\mu\left(\prod_{i=2}^{n}(I-2S_\mu g_{i,k,j}\otimes g_{i,k,j})B\otimes B_{k,j}\right)\\ &=\dots=\mu(B\cap B_{k,j}).\end{aligned}$$

Hence

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

Thus we have

$$\mu(D(B))\leqq\mu(B) \tag{4.15}$$

for any open set B.

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

$$\mu(D(B))\leqq\mu(D(B_s))\leqq\mu(B_s).$$

Thus $\mu(D(B))\leqq\lim_s\mu(B_s)=\mu(B)$. Hence (4.15) holds also for closed sets.

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

$$1-\mu(D(B))=\mu(D(F_1-B))\leqq\mu(F_1-B)=1-\mu(B).$$

Thus $\mu(B)\leqq\mu(D(B))$. This and (4.15) give

$$\mu(D(B))=\mu(B) \tag{4.16}$$

for any open set B. Since the set of B for which (4.16) holds is a σ-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

$$\int_B H(f)\mu(df) = \int_{D(B)} H(Df)\mu(df)$$

for any measurable function H and any Borel set B.

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

$$\mu_1(A, N^a) = \mu((N^a)^{-1}(A)) = \mu(\{f \in F_1 : N^a(f) \in A\}) \tag{4.17}$$

where A is a Borel set of $\mathbb{R}^n$. The measure μ_1, called the probability induced by N^a, tells us the probability that $N(f) \in A$.

We prove that the measure μ_1 is independent of N^a and μ_1 is orthogonally invariant with mean zero and the identity covariance operator.

Theorem 4.2. *There exists a probability measure μ_1 defined on Borel sets of $\mathbb{R}^n$ such that*

$$\mu_1(A, N^a) = \mu_1(A), \qquad \forall A \in \mathbb{B}(\mathbb{R}^n), \tag{4.18}$$

for any measurable adaptive information N^a of the form (4.1). ☐

Proof. We first consider nonadaptive information operators. Let

$$N_1(f) = [(f, \zeta_1), (f, \zeta_2), \ldots, (f, \zeta_n)],$$
$$N_2(f) = [(f, \eta_1), (f, \eta_2), \ldots, (f, \eta_n)]$$

where $(S_\mu \zeta_i, \zeta_j) = (S_\mu \eta_i, \eta_j) = \delta_{i,j}$. We prove

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

$$N_1 = N_2 Q \tag{4.19}$$

$$\mu(Q^{-1}B) = \mu(B), \qquad \forall B \in \mathbb{B}(F_1). \quad ☐ \tag{4.20}$$

Proof. Let $X = \mathrm{lin}\{S_\mu^{\frac12}\zeta_1, \ldots, S_\mu^{\frac12}\zeta_n, S_\mu^{\frac12}\eta_1, \ldots, S_\mu^{\frac12}\eta_n\}$. 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 $\{S_\mu^{\frac12}\eta_i\}_{i=1}^p$ and $\{S_\mu^{\frac12}\zeta_i\}_{i=1}^p$ are orthonormal bases of X. Define the mapping $H: F_1 \to F_1$,

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

Since $S_\mu^{\frac12}\eta_k = \sum_{i=1}^{p} (S_\mu^{\frac12}\eta_k, S_\mu^{\frac12}\zeta_i) S_\mu^{\frac12}\zeta_i$, we get $\eta_k = \sum_{i=1}^{p} (\eta_k, S_\mu \zeta_i)\zeta_i$ and

$$H\eta_k = \sum_{i=1}^{p} (\eta_k, S_\mu \eta_i)\zeta_i - \sum_{i=1}^{p} (\eta_k, S_\mu \zeta_i)\zeta_i - \eta_k = \zeta_k \tag{4.21}$$

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

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

To prove (4.19) note that $N_1 = N_2 Q$ 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 (4.21)).

To prove (4.20) we decompose H as

$$H = S_\mu^{-1/2} H_1 S_\mu^{1/2}$$

where $H_1 f = \sum_{i=1}^{p} (f, S_\mu^{1/2}(\eta_i+\zeta_i))S_\mu^{1/2}\zeta_i - f$. Note that $H_1 S_\mu^{1/2}(F_1) \subset S_\mu^{1/2}(F_1)$ and therefore $S_\mu^{-1/2}(H_1 S_\mu^{1/2})$ 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

$$H_1 f = -f, \quad \forall f \in X^\perp. \tag{4.22}$$

From (4.21) we have

$$H_1 S_\mu^{1/2}\eta_k = S_\mu^{1/2}\zeta_k, \quad k=1,2,\dots,p.$$

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

$$-H_1 f = D_1 D_2 \cdot \ldots \cdot D_p f, \quad \forall f \in X, \tag{4.23}$$

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

For $f \in X^\perp$ we have $(f,x_i) = 0$ and we get $D_1 D_2 \cdot \ldots \cdot D_p f = f$. Thus, (4.23) holds also for $f \in X^\perp$ due to (4.22). Hence we proved that $H_1 = -D_1 D_2 \cdot \ldots \cdot D_p$ and

$$\begin{aligned} H &= -S_\mu^{-1/2} D_1 D_2 \cdot \ldots \cdot D_p S_\mu^{1/2} \\ &= -(S_\mu^{-1/2} D_1 S_\mu^{1/2}) \cdot \ldots \cdot (S_\mu^{-1/2} D_p S_\mu^{1/2}) \\ &= -Q_1^* Q_2^* \ldots Q_p^* \end{aligned}$$

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

$$\begin{aligned} \mu(Q^{-1}B) &= \mu(-Q_1 \cdot \ldots \cdot Q_p B) = \mu(Q_1 \cdot \ldots \cdot Q_p B) = \mu(Q_2 \cdot \ldots \cdot Q_p B) \\ &= \ldots = \mu(B) \end{aligned}$$

which proves (4.20) and completes the proof of Lemma 4.1. □

Define the measure μ_1 as

$$\mu_1(A)=\mu_1(A,N_1), \qquad \forall A\in\mathbb{B}(\mathbb{R}^n). \tag{4.24}$$

From Lemma 4.1 we immediately get

$$\begin{aligned}\mu_1(A,N_2)&=\mu(N_2^{-1}A)=\mu(Q^{-1}N_2^{-1}A)\\&=\mu(N_1^{-1}A)=\mu_1(A,N_1)=\mu_1(A), \qquad \forall A\in\mathbb{B}(\mathbb{R}^n).\end{aligned}$$

Thus (4.18) holds for any nonadaptive information of the form (4.1).

Take now any measurable adaptive information N^a. Using (4.8) and (4.9) define

$$N_k(f)=[(f,g_{1,k,j}),(f,g_{2,k,j}),\dots,(f,g_{n,k,j})]$$

for $N^a(f)\in A_{k,j}$. Then

$$N^a(f)=\lim_k N_k(f), \qquad \forall f\in F_1.$$

Let A be an open set of $\mathbb{R}^n$. Then

$$(N^a)^{-1}(A)\subset E \overset{\mathrm{df}}{=} \bigcup_{i=1}^{\infty}\bigcap_{k=i}^{\infty} N_k^{-1}(A). \tag{4.25}$$

Indeed, if $f\in(N^a)^{-1}(A)$ then $y=N^a(f)\in A$. Let $y_k=N_k(f)$. Then $\lim_k y_k=y\in A$. Since A is open, $y_k\in A$ for $k\geq k_0$. Thus $f\in N_k^{-1}(y_k)\subset N_k^{-1}(A)$ for $k\geq k_0$. This means that $f\in E$ as claimed. From (4.25) we have

$$\begin{aligned}\mu_1(A,N^a)&=\mu((N^a)^{-1}(A))\leq\mu(E)\\&=\lim_{i\to\infty}\mu\left(\bigcap_{k=i}^{\infty}N_k^{-1}(A)\right)\leq\lim_{k\to\infty}\mu(N_k^{-1}(A)).\end{aligned}$$

Observe that

$$\mu(N_k^{-1}(A))=\sum_{j=1}^{n_k}\mu(N_k^{-1}(A\cap A_{k,j})).$$

Since N_k on each $A_{k,j}$ coincides with nonadaptive information, we have

$$\mu(N_k^{-1}(A\cap A_{k,j}))=\mu_1(A\cap A_{k,j})$$

and

$$\mu(N_k^{-1}(A))=\sum_{j=1}^{n_k}\mu_1(A\cap A_{k,j})=\mu_1(A).$$

Thus

$$\mu_1(A,N^a)\leq\mu_1(A) \tag{4.26}$$

for any open set A. Take now a closed set A and define $A_s=\{y\in\mathbb{R}^n: \operatorname{dist}(y,A)<1/s\}$, $s=1,2,\dots,n$. Then $A\subset A_{s+1}\subset A_s$, $A=\bigcap_{s=1}^{\infty}A_s$. Since A_s is open we have due to (4.26),

$$\mu_1(A,N^a)=\mu((N^a)^{-1}(A))\leq\mu((N^a)^{-1}(A_s))\leq\mu_1(A_s).$$

Thus

$$\mu_1(A\cdot N^a)\leq \lim_s \mu_1(A_s)=\mu_1(A).$$

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). □

Theorem 4.2 will be used to compute $\int_{F_1} H(Nf)\mu(df)$ for any measurable H and N of the form (4.1). 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 μ_1 of Theorem 4.2 is orthogonally invariant with mean zero and the identity covariance operator.* □

Proof. We first show that

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

Take $\zeta_1,\zeta_2,\ldots,\zeta_n$ from F_1 such that $(S_\mu\zeta_i,\zeta_j)=\delta_{i,j}$. Define

$$N(f)=[(f,\zeta_1),(f,\zeta_2),\ldots,(f,\zeta_n)]. \tag{4.28}$$

Let $g=\sum_{i=1}^n x_i\zeta_i$. Since $m_\mu=0$, we have

$$0=\int_{F_1}(f,g)\mu(df)=\int_{F_1}\sum_{i=1}^n x_i(f,\zeta_i)\mu(df).$$

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

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

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

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

$$\int_{\mathbb{R}^n}(y,x)(y,z)\mu_1(dy)=(x,z),\qquad \forall x,z\in\mathbb{R}^n. \tag{4.29}$$

For $g=\sum_{i=1}^n y_i\zeta_i$ and $h=\sum_{i=1}^n z_i\zeta_i$ 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 $(S_\mu g,h)=\sum_{i,j=1}^n x_i z_j(S_\mu\zeta_i,\zeta_j)=(x,z)$, (4.29) follows.

We now prove that μ_1 is orthogonally invariant, i.e.,

$$\mu_1(Q(B))=\mu_1(B),\qquad \forall B\in\mathbb{B}(\mathbb{R}^n)$$

where $Qy = 2(y, x)x - y$, $y \in \mathbb{R}^n$, $\|x\| = 1$ or $x = 0$. Define the mapping

$$Df = 2(f, g)S_\mu g - f, \quad \forall f \in F_1,$$

where, as before, $g = \sum_{i=1}^{n} x_i \zeta_i$. Then $(S_\mu g, g) = \sum_{i=1}^{n} x_i^2 = 1$ or $g = 0$. Observe that

$$N^{-1}QB = DN^{-1}B, \quad \forall B \in \mathbb{B}(\mathbb{R}^n). \tag{4.30}$$

Indeed, $f \in N^{-1}QB$ iff $Nf \in QB$ iff $QNf \in B$ since $Q^2 = I$. Similarly $f \in DN^{-1}B$ iff $NDf \in B$ since $D^2 = I$. Note that

$$QNf = 2(Nf, x)x - Nf = 2(f, g)x - Nf,$$
$$NDf = 2(f, g)N(S_\mu g) - Nf = 2(f, g)\, x - Nf$$

which proves (4.30). From Theorem 4.2, (4.30) and orthogonal invariance of μ, we have

$$\mu_1(Q(B)) = \mu(N^{-1}Q(B)) = \mu(DN^{-1}B) = \mu(N^{-1}B) = \mu_1(B)$$

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^a has minimal average error among all algorithms that use N^a. The second step is to estimate from below the average radius of information (or equivalently the average error of the spline algorithm).

Let

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

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

$$(S_\mu g_i(f), g_j(f)) = \delta_{i,j}. \tag{5.1}$$

Indeed, as in [16] let $\eta_1(f), \ldots, \eta_n(f)$ be an orthonormal basis of the linear space lin $(S_\mu g_1(f), \ldots, S_\mu g_n(f))$. Then there exists a nonsingular matrix M such that

$$[(f, \tilde{g}_1(f)), \ldots, (f, \tilde{g}_n(f))] = N^a(f)M \tag{5.2}$$

where $\tilde{g}_i(f) = S_\mu^{-\frac{1}{2}}\eta_i(f)$ and $(S_\mu \tilde{g}_i(f), \tilde{g}_j(f)) = \delta_{i,j}$. Thus, knowing $N^a(f)$ we can compute $(f, \tilde{g}_i(f))$. The mappings $\tilde{g}_i$ are also measurable. The elements $\tilde{g}_i(f)$ then play the role of $g_i(f)$. This explains (5.1).

Define

$$\sigma = \sigma(N^a(f)) = \sum_{i=1}^{n} (f, g_i(f))S_\mu g_i(f). \tag{5.3}$$

Note that $N^a(\sigma)=N^a(f)$, i.e., σ interpolates f. Furthermore, take $h\in S_a(F_1)$ such that $N^a(h)=N^a(f)$. Then

$$\|h\|_*^2=\|\sigma-h\|_*^2+\|\sigma\|_*^2-2(S_a^{-1}\sigma,h-\sigma).$$

Since $(h-\sigma,g_i(f))=0$ we have $(S_a^{-1}\sigma,h-\sigma)=0$ and $\|h\|_*\geq\|\sigma\|_*$. Thus σ has minimal norm $\|\cdot\|_*$ among elements which interpolate f and lie in $S_a(F_1)$. Such an element is called a spline interpolating f. Let

$$\varphi^s(N^a(f))=S\sigma(N^a(f))=\sum_{i=1}^{n}(f,g_i(f))SS_ag_i(f) \tag{5.4}$$

be the spline algorithm.

We say an algorithm φ is an optimal average error algorithm iff

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

Theorem 5.1. *If μ is orthogonally invariant then the spline algorithm φ^s is an optimal average error algorithm and*

$$e^{\text{avg}}(\varphi^s,N^a)^2=\int_{F_1}\|Sf\|^2\mu(df)-\int_{F_1}\|\varphi^s(N^a(f))\|^2\mu(df). \quad \square \tag{5.5}$$

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 μ and (4.3) yield

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

where φ is an algorithm and D is the mapping defined by (4.2); see Theorem 4.1. Thus

$$e^{\text{avg}}(\varphi,N^a)=\tfrac{1}{2}\int_{F_1}\{\|Sf-\varphi(N^a(f))\|^2$$
$$+\|SD(f)-\varphi(N^a(f))\|^2\}\mu(df). \tag{5.6}$$

Since

$$\|Sf-\varphi^s(N^a(f))\|^2=\tfrac{1}{4}\|Sf-SD(f)\|^2\leq\tfrac{1}{4}(\|Sf-\varphi(N^a(f))\|+\|SD(f)-\varphi(N^a(f))\|)^2$$
$$\leq\tfrac{1}{2}(\|Sf-\varphi(N^a(f))\|^2+\|SD(f)-\varphi(N^a(f))\|^2)$$

we get

$$e^{\text{avg}}(\varphi,N^a)\geq e^{\text{avg}}(\varphi^s,N^a).$$

This means that φ^s is an optimal average error algorithm. To prove (5.5) note that

$$a=\|Sf-\varphi^s(N^a(f))\|^2+\|SD(f)-\varphi^s(N^a(f))\|^2$$
$$=\|Sf\|^2+\|SD(f)\|^2-2\|\varphi^s(N^a(f))\|^2-2(Sf,\varphi^s(N^a(f)))-2(SD(f),\varphi^s(N^a(f))).$$

Since $D(f)=2\sigma(N^a(f))-f$, then $SD(f)=2\varphi^s(N^a(f))-f$ and

$$a=\|Sf\|^2+\|SD(f)\|^2-2\|\varphi^s(N^a(f))\|^2.$$

This and (5.6) with $\phi = \phi^s$ yield (5.5). $\square$

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

$$\int_{F_1} \|\phi^s(N^a(f))\|^2 \mu(df) = \int_{\mathbb{R}^n} \|\phi^s(y)\|^2 \mu_1(dy)$$

$$= \sum_{i,j=1}^{n} \int_{\mathbb{R}^n} y_i y_j (SS_u g_i(y), SS_u g_j(y)) \mu_1(dy) \tag{5.7}$$

where, as in Sect. 4, $g_i(y) = g_i(y_1, y_2, \ldots, y_{i-1})$. 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 $g_j(Qy) = g_j(y)$ for $j < i$. Since μ_1 is orthogonally invariant we have

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

$$= \int_{\mathbb{R}^n} y_i y_j (Qe_i, Qe_j)(SS_u g_i(Qy), SS_u g_j(Q(y)) \mu_1(dy)$$

$$= -a.$$

Hence $a = 0$ and (5.7) becomes

$$\int_{F_1} \|\phi^s(N^a(f))\|^2 \mu(df) = \sum_{i=1}^{n} \int_{\mathbb{R}^n} y_i^2 \|SS_u g_i(y)\|^2 \mu_1(dy). \tag{5.8}$$

For $i < n$ define the mapping

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

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

$$\int_{\mathbb{R}^n} y_i^2 \|SS_u g_i(y)\|^2 \mu_1(dy) = \int_{\mathbb{R}^n} y_n^2 \|SS_u g_i(y)\|^2 \mu_1(dy).$$

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

$$\int_{F_1} \|\phi^s(N^a(f))\|^2 \mu(df) = \int_{\mathbb{R}^n} y_n^2 \sum_{i=1}^{n} \|SS_u g_i(y)\|^2 \mu_1(dy)$$

$$\leq (\sup_{y \in \mathbb{R}^n} \|SS_u g_i(y)\|^2) \int_{\mathbb{R}^n} y_n^2 \mu_1(dy) = \sum_{i=1}^{n} \|SS_u g_i(f^*)\|^2. \tag{5.9}$$

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

$$\int_{F_1} \|\phi^s(N_{f^*}^{non}(f))\|^2 \mu(df) = \sum_{i,j=1}^{n} \int_{\mathbb{R}^n} y_i y_j (SS_u g_i^*, SS_u g_j^*) \mu_1(dy)$$

$$= \sum_{i=1}^{n} \|SS_u g_i^*\|^2. \tag{5.10}$$

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

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

This completes the proof. □

Acknowledgements. 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

1. Bakhvalov, N.S.: On the optimality of linear methods for operator approximation in convex classes of functions (in Russian). Zh. Vyčisl. Mat. i. Mat. Fiz. **11**, 1014–1018 (1971) (English transl.: U.S.S.R. Computational Math. and Math. Phys. **11**, 244–249 (1971))
2. Gal, S., Micchelli, A.C.: Optimal sequential and nonsequential procedures for evaluating a functional. Appl. Anal. **10**, 105–120 (1980)
3. Kacewicz, B.: On the optimal error of algorithms for solving a scalar autonomous ODE. BIT (in press)
4. Kuo, Hui-Hsuing: Gaussian measures in banach spaces. Lecture Notes in Mathematics 463. Berlin-Heidelberg-New York: Springer 1975
5. Ralston, A., Rabinowitz, P.: A first course in numerical analysis. New York: McGraw-Hill 1978
6. Skorohod, A.V.: Integration in Hilbert space. New York: Springer 1974
7. Sikorski, K.: Bisection is optimal. Numer. Math. **40**, 111–117 (1982)
8. Sikorski, K., Woźniakowski, H.: For which error criteria can we solve nonlinear equations? Dept. of Computer Science Report, Columbia University, 1983
9. Sukharev, A.G.: Optimal strategies of the search for an extremum (in Russian). Zh. Vyčisl. Mat. i. Mat. Fiz. 11, 910–924 (1971) (English transl.: U.S.S.R. Computational Math. and Math. Phys. 11, 119–137 (1971))
10. Traub, J.F., Wasilkowski, G.W., Woźniakowski, H.: Average case optimality for linear problems. Theor. Comput. Sci. (in press)
11. Traub, J.F., Wasilkowski, G.W., Woźniakowski, H.: Information, uncertainty, complexity. Reading, Mass.: Addison-Wesley 1983
12. Traub, J.F., Woźniakowski, H.: A general theory of optimal algorithms. New York: Academic Press 1980
13. Traub, J.F., Woźniakowski, H.: Information and computation. In: Advances in computers, Vol. 23 (M. Yovits, ed.). New York: Academic Press 1984
14. Wasilkowski, G.W.: Some nonlinear problems are as easy to approximate as the identity operator. Dept. of Computer Science Report, Columbia University 1983. Comput. Math. Appl. (in press)
15. Wasilkowski, G.W.: Local average error. Dept. of Computer Science Report, Columbia University 1983
16. Wasilkowski, G.W., Woźniakowski, H.: Average case optimal algorithms in Hilbert spaces. Dept. of Computer Science Report, Columbia University 1982
17. Zaliznyak, N.F., Ligun, A.A.: On optimum strategy in search of global maximum of functions (in Russian). Zh. Vyčisl. Mat. i. Mat. Fiz. **18**, 314–321 (1978) (English transl.: U.S.S.R. Computational Math. and Math. Phys. **18**, 31–38 (1978))

Received November 15, 1982 / November 21, 1983 / February 14, 1984