JOURNAL OF COMPLEXITY 8, 337–372 (1992)

Average Case Complexity of Linear Multivariate Problems

I. Theory

H. WOŹNIAKOWSKI

Department of Computer Science, Columbia University, 500 West 120th Street, New York, New York 10027 and Institute of Applied Mathematics, University of Warsaw, 00-325 Warsaw, Poland

Received November 1991

We study the average case complexity of linear multivariate problems, that is, the approximation of continuous linear operators on functions of d variables. The function spaces are equipped with Gaussian measures. We consider two classes of information. The first class Λ^{std} consists of function values, and the second class Λ^{all} consists of all continuous linear functionals. Tractability of a linear multivariate problem means that the average case complexity of computing an ε-approximation is $O((1/\varepsilon)^p)$ with p independent of d. The smallest such p is called the exponent of the problem. Under mild assumptions, we prove that tractability in Λ^{all} is equivalent to tractability in Λ^{std}, and that the difference of the exponents is at most 2. The proof of this result is not constructive. We provide a simple condition to check tractability in Λ^{all}. We also address the issue of how to construct optimal (or nearly optimal) sample points for linear multivariate problems. We use relations between average case and worst case settings. These relations reduce the study of the average case to the worst case for a different class of functions. In this way we show how optimal sample points from the worst case setting can be used in the average case. In Part II we shall apply the theoretical results to obtain optimal or almost optimal sample points, optimal algorithms, and average case complexity functions for linear multivariate problems equipped with the folded Wiener sheet measure. Of particular interest will be the multivariate function approximation problem.

1. INTRODUCTION

We study linear multivariate problems which are defined as approximating continuous linear operators on functions of d variables. We are

0885-064X/92 $5.00

particularly interested in the case of large d. Part I deals with the theory of linear multivariate problems. Part II will deal with applications of the theoretical results to concrete linear multivariate problems. Two important examples of such problems are *multivariate integration* and *multivariate function approximation* in which we wish to integrate or recover a function which depends on d variables.

Many linear multivariate problems are *intractable* in the worst case setting. That is, the worst case complexity of computing an ε-approximation is infinite or grows *exponentially*[1] with d; see e.g., TWW (1988).[2] For example, for multivariate integration and function approximation of r times continuously differentiable functions of d variables, the worst case complexity is of order $(1/\varepsilon)^{d/r}$, assuming that an ε-approximation is computed using function values. Thus, if only continuity of the functions is assumed, i.e., $r = 0$, then the worst case complexity is infinite. For positive r, if d is large relative to r, then the worst case complexity is huge even for modest ε. In either case, the problem cannot be solved in the worst case setting.

To break intractability of the worst case setting, we must switch to a different setting with a weaker guarantee of computing an ε-approximation. In this paper we choose to switch to an average case setting and we study liner multivariate problems on the average with respect to a Gaussian measure. The average case complexity is defined as the minimal average cost of computing an approximation with average error at most ε.

The average case complexity depends, in particular, on the class Λ of information operations. We consider two classes. The first class $\Lambda = \Lambda^{\text{std}}$ consists of function values, the second class $\Lambda = \Lambda^{\text{all}}$ consists of all continuous linear functionals. We are particularly interested in how the average case complexity depends on ε, d, and Λ.

We say that a linear multivariate problem is *tractable* in the average case setting iff there exists a nonnegative number p such that, for all d, its average case complexity is $O((1/\varepsilon)^p)$. The smallest such p is called the *exponent* of that linear multivariate problem. That is, tractability means that, no matter how large d, we can compute an average ε-approximation with an average cost which is a polynomial in $1/\varepsilon$ of fixed degree p. Obviously, we wish to have p as small as possible, and the smallest p is the exponent of the linear multivariate problem.

We stress that the concept of tractability ignores multiplicative factors which may, in particular, depend on d. In fact, most estimates presented in this paper are modulo a multiplicative factor which may depend on d.

[1] In this paper we will not distinguish between infinite complexity and complexity which grows exponentially with d. Both will be called intractable. Sometimes a distinction is made, see e.g., Traub and Woźniakowski (1991).

[2] By TWW (1988) we mean Traub, Wasilkowski, and Woźniakowski (1988).

Obviously, this dependence on d is very important in practical computations. Ideally, we would like to bound the average case complexity by $\alpha c(1/\varepsilon)^q$ for all d and ε, for some fixed (and hopefully small) nonnegative α and q which are independent of d. Here, c is the cost of computing a functional from Λ and may depend on d; this is the only dependence on d. We call this property *strong tractability*. We shall report on strong tractability in a future paper.

The first major subject of this paper is to study which linear multivariate problems are tractable in the average case setting. Under mild assumptions, we show that tractability in Λ^{std} is equivalent to tractability in Λ^{all}. The difference between their exponents is at most 2, and this is sharp. We provide a simple condition for checking tractability in Λ^{all}. We also show that tractability of multivariate function approximation for a particular measure implies tractability of all linear multivariate problems for that measure.

In this way we may check tractability of a particular linear multivariate problem in Λ^{all}, or equivalently in Λ^{std}. Clearly, all linear multivariate problems specified by a linear functional are tractable in Λ^{std} since they can be computed exactly with one information evaluation from Λ^{all} and thus are trivial in that class. Therefore, their average case complexity in Λ^{std} is at most of order $(1/\varepsilon)^2$.

In particular, this means that in the average case setting, multivariate integration is tractable in Λ^{std} and its exponent is at most 2. This is in a sharp contrast with the worst case setting where, even for $d = 1$, the worst case complexity in Λ^{std} can be infinite or an arbitrary increasing function of $1/\varepsilon$; see Werschulz (1985). Of course, intractability of multivariate integration in the worst case setting can be also broken by switching to the randomized setting and using the classical Monte Carlo algorithm.

In Λ^{all} it is known which information operations are optimal; see TWW (1988, p. 234). This fact is used in the proof of the theorem on tractability in Λ^{std} to conclude the existence of good sample points at which the function should be evaluated. Unfortunately, the proof is *not constructive*. Thus, although the theorem states that the average case complexity in Λ^{std} is bounded by a polynomial in $1/\varepsilon$ of fixed degree, its proof does not provide a constructive way to achieve this bound.

The optimal design problem of constructing sample points which achieve (or nearly achieve) the average case complexity in Λ^{std} is the second major subject of the paper. This problem has been long open, even for multivariate integration and function approximation. We address the construction of optimal (or nearly optimal) sample points for linear multivariate problems by utilizing relations between average case and worst case settings.

We first discuss the approximation of a continuous linear functional which corresponds to a multivariate *weighted* integration. In this case, the relation between average case and worst case settings is well known and used in many papers; see e.g., Kimeldorf and Wahba (1970a, b), Micchelli and Wahba (1981), Paskov (1991), Sacks and Ylvisaker (1966, 1968, 1970a, b), TWW (1988, Section 2.2 of Chap. 7), Wahba (1971), and Ylvisaker (1975). A thorough overview may be found in Wahba (1990). This relation states that the average error of a *linear* algorithm that uses n function values at *nonadaptive* sample points is equal to the worst error of the same algorithm over the unit ball of a reproducing kernel Hilbert space. The kernel of this Hilbert space is given by the covariance kernel of the measure defining the average case setting.

We compare the average case complexity to the worst case complexity of the corresponding problem defined on the reproducing kernel Hilbert space. We show that the average case complexity is bounded by the worst case complexity of the corresponding problem and they can differ only if adaption helps on the average. Using general results of Wasilkowski (1986), we show how to bound the average case complexity from below by the worst case complexity. Often the bounds differ only by a multiplicative factor. In this way, we reduce optimal design in the average case setting to optimal design of the corresponding problem in the worst case setting.

We note that the worst error of multivariate weighted integration is bounded by the worst error multivariate *function* approximation. Thus, it is enough to use optimal sample points of multivariate function approximation in the worst case setting to get good, and sometimes optimal, sample points for multivariate weighted integration in the average case setting.

For general multivariate problems, in which we approximate continuous linear operators, we may choose either of two approaches. The first one is to express a linear multivariate problem as a number of multivariate weighted integrals and apply the analysis performed for multivariate weighted integration. Hence, good sample points for multivariate weighted integration in the average case setting can be used for general multivariate problems in the average case setting.

The second approach is to notice that an arbitrary linear multivariate problem can be solved by solving the multivariate function approximation problem. Hence, it is enough to study the latter. Multivariate function approximation in the average case setting is related to a multivariate function approximation problem in the worst case setting for the reproducing kernel Hilbert space. The worst case problem assumes that the error is defined in the L_∞ norm. The worst error serves as an upper bound on the average error. In this way, good sample points for multivariate function approximation in the worst case setting can be used for general multivariate problems in the average case setting.

2. Linear Multivariate Problems

In this section we define a linear multivariate problem LMP as a sequence of linear multivariate problems indexed by d, LMP = $\{\text{LMP}_d\}$. Here, d represents the number of variables of the functions we are dealing with.

The linear multivariate problem LMP_d is specified by several parameters (F, μ, G, S, Λ) which may also depend on d. We now define them in turn.

Let D be a Lebesgue measurable subset of $\mathbb{R}^d$. By $\lambda(D)$ we denote the volume of D. We assume that $\lambda(D)$ is positive and finite. Let F be a separable Banach space of functions $f\colon D \to \mathbb{R}$. We assume that F is a subset of the space $L_2(D)$ of square integrable functions and that linear functionals $L(f) = f(x)$ for any $x \in D$ are continuous with respect to the norm of F.

The space F is equipped with a Gaussian measure μ with mean zero and covariance operator C_μ; for basic properties of Gaussian measures, see, e.g., Kuo (1975) and Vakhania (1981). Let

$$R_\mu(t, x) = \int_F f(t)f(x)\mu(df), \qquad t, x \in D, \tag{2.1}$$

be the covariance kernel of the measure μ. It is well defined since $f(t)$ and $f(x)$ are continuous linear functionals.

Consider a continuous linear operator S,

$$S\colon F \to G,$$

where G is a separable Hilbert space over the real field. Our aim is to approximate elements $S(f)$ for $f \in F$.

The last parameter of LMP_d is the class Λ which consists of certain continuous linear functionals $L\colon F \to \mathbb{R}$. We assume that Λ is either Λ^{std} or Λ^{all}. Here,

$$\Lambda^{\text{std}} = \{L\colon \text{there exists } x \in D \text{ such that } L(f) = f(x),\ \forall f \in F\},$$

which means that only function values are considered, and

$$\Lambda^{\text{all}} = F^*,$$

which means that all continuous linear functionals are considered.

This completes the definition of all parameters of the linear multivariate problem LMP_d. It is called *linear* to stress that we are approximating a *linear* operator S. As already mentioned, by a linear multivariate problem

LMP we mean a sequence $\{\text{LMP}_d\}$ with varying d. We are particularly interested in the case when d is large.

We now explain how we compute an approximation $U(f)$ to the element $S(f)$. Assume that information about the function f, $f \in F$, is gathered by computing a number of continuous linear functionals $L(f)$, where $L \in \Lambda$. Hence, if $\Lambda = \Lambda^{\text{std}}$ then we assume that only function values can be computed, and if $\Lambda = \Lambda^{\text{all}}$ then we assume that arbitrary continuous linear functionals can be computed.

Let

$$N(f) = [L_1(f), L_2(f), \ldots, L_n(f)], \qquad \forall f \in F,$$

denote the computed *information* about f. The choice of L_i, $L_i \in \Lambda$, may depend adaptively on the already computed information,

$$L_i = L_i(\cdot; y_1, \ldots, y_{i-1}) \qquad \text{with } y_i = L_i(f).$$

The number $n = n(f)$ is called the *cardinality* of the information at f, and, in general, depends on the computed y_i, see TWW (1988, Chap. 3).

Knowing $y = N(f)$, the approximation U is computed as $U(f) = \phi(y)$, where ϕ, ϕ: $N(F) \to G$, is an arbitrary mapping. Some restriction on the choice of ϕ are imposed by defining the cost of U and seeking U which computes an ε-approximation with minimal cost. In this way, ϕ with a high cost of computing $\phi(y)$ will be automatically eliminated.

We define error and cost of the approximation U. Since we deal with the average case setting, error and cost are both defined on the average. The *average error* of U is defined as

$$e^{\text{avg}}(U) = \left(\int_F \|S(f) - U(f)\|^2 \mu(df)\right)^{1/2}.$$

The average cost of U is defined as follows. Assume that each evaluation of $L(f)$, $L \in \Lambda$ and $f \in F$, cost $c = c(d)$, where $c > 0$. Assume that we can perform arithmetic operations and comparisons on real numbers as well as the basic operations in the space G with cost taken as unity. By the basic operations we mean adding two elements $g + h$, and multiplying by a scalar αg for $g, h \in G$ and $\alpha \in \mathbb{R}$. Usually the cost of computing $L(f)$ is much larger than unity, $c \gg 1$.

Let cost(N, f) denote the information cost of computing $y = N(f)$. Clearly, we have cost$(N, f) \geq cn(f)$. Let $n_1(f)$ denote the number of operations needed to compute $\phi(y)$ given $y = N(f)$. (It may happen that $n_1(f) = +\infty$.) The *average cost* of U is then given as

$$\text{cost}^{\text{avg}}(U) = = \int_F (\text{cost}(N, f) + n_1(f))\mu(df).$$

We are ready to define the average case complexity of LMP_d as the minimal cost of computing ε-approximations,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon) = \inf\{\mathrm{cost}^{\mathrm{avg}}(U) \colon U \text{ such that } e^{\mathrm{avg}}(U) \le \varepsilon\}.$$

We stress that the average case complexity $\mathrm{comp}^{\mathrm{avg}}(\varepsilon)$ depends on all parameters of LMP_d,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon) = \mathrm{comp}^{\mathrm{avg}}(\varepsilon; \mathrm{LMP}_d) = \mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, F, \mu, G, S, \Lambda).$$

To stress the dependence on certain parameters, we will sometimes list only just those. Hence, if we write

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon; d) \qquad \text{or} \qquad \mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, \Lambda),$$

then the role of d, or d and Λ is stressed. As mentioned before, the dependence on d is also present in some other parameters. For example, the measure μ and the operator S both depend on d. Sometimes we write $\mu = \mu_d$ and $S = S_d$ to stress this dependence.

We will be particularly interested in how the average case complexity depends on ε and d as well as how it depends on Λ. Since $\Lambda^{\mathrm{std}} \subset \Lambda^{\mathrm{all}}$, we have

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, \Lambda^{\mathrm{all}}) \le \mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, \Lambda^{\mathrm{std}}).$$

The average case complexity functions in Λ^{all} and Λ^{std} are usually closely related and $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, \Lambda^{\mathrm{std}})$ cannot be much larger than $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; d, \Lambda^{\mathrm{all}})$, as we see in the next section.

3. Tractability of Linear Multivariate Problems

Consider a linear multivariate problem $\mathrm{LMP} = \{\mathrm{LMP}_d\}$. Let $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \mathrm{LMP}_d)$ denote the average case complexity of LMP_d. Suppose we know that

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \mathrm{LMP}_d) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{p(d)}\right),$$

where the multiplicative factor in the Θ notation may depend on d. Obviously, the cost c of one functional evaluation also depends on d, i.e., $c = c(d)$.

If $\lim_{d\to+\infty} p(d) = +\infty$ then for large d, the average case complexity is huge even for moderate ε. Thus, LMP_d cannot be solved. In this case we say that the linear multivariate problem LMP is *intractable*.

We wish to investigate for which linear multivariate problems LMP the exponents $p(d)$ do not tend to infinity and can be uniformly bounded, i.e., $p(d) \leq p$. This motivates the following definition.

A linear multivariate problem LMP $= \{\text{LMP}_d\}$ is called *tractable* in the average case setting if there exists a nonnegative number p such that for all d,

$$\text{comp}^{\text{avg}}(\varepsilon;\, \text{LMP}_d) = O\left(c\left(\frac{1}{\varepsilon}\right)^p\right). \tag{3.1}$$

As before, the multiplicative factor in the big O notation may depend on d. Thus, tractability means that the average case complexity is asymptotically in ε bounded by a polynomial in $1/\varepsilon$ whose degree does not exceed p for all d.

Obviously, (3.1) does not uniquely define p. Furthermore, (3.1) allows us to ignore a polynomial factor in $\log 1/\varepsilon$ whose degree may depend on d. Indeed, if for all d

$$\text{comp}^{\text{avg}}(\varepsilon;\, \text{LMP}_d) = O\left(c\left(\frac{1}{\varepsilon}\right)^{p_1}\left(\log\frac{1}{\varepsilon}\right)^{p_2(d)}\right)$$

then we can absorb $(\log 1/\varepsilon)^{p_2(d)}$ by taking $p > p_1$.

From a practical point of view, we would like to have p in (3.1) as small as possible. This motivates the definition of the *exponent* $p^* = p^*(\text{LMP})$ of a linear multivariate problem LMP which is given by

$$p^* = \begin{cases} \inf\{p\colon p \in P\} & \text{if } P \neq \varnothing, \\ +\infty & \text{if } P = \varnothing, \end{cases}$$

where P is the set of all nonnegative p for which (3.1) holds. Thus, the exponent P^* of LMP is, roughly, the smallest p for which (3.1) holds. If for all d

$$\text{comp}^{\text{avg}}(\varepsilon;\, \text{LMP}_d) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{p_1}\left(\log\frac{1}{\varepsilon}\right)^{p_2(d)}\right)$$

then, obviously, $p^* = p_1$.

Tractability of a linear multivariate problem depends on all its parameters. In particular, it depends on Λ. To stress the role of Λ, we say that a linear multivariate problem is *tractable in* Λ iff (3.1) holds for Λ.

We are ready to study tractability in Λ^{std} and Λ^{all}. Obviously, tractability in Λ^{std} implies tractability in Λ^{all}. We now show that, under mild assumptions, the converse is true, i.e., tractability in Λ^{all} implies tractability in Λ^{std} and the difference between their exponents is at most 2. We also present a simple condition to check tractability in Λ^{all}.

Let $\|\cdot\|_d$ denote the $L_2(D)$ norm,

$$\|f\|_d = \left(\int_D f^2(t)\, dt\right)^{1/2}.$$

We assume that for all d there exist two nonnegative constants $K_1 = K_1(d)$ and $K_2 = K_2(d)$ such that

$$\text{(A.1):} \quad \|S_d(f)\| \leq K_1\|f\|_d, \qquad \forall f \in F,$$

$$\text{(A.2):} \quad \|R_\mu(\cdot, \cdot)\|_{L_\infty(D)} \leq K_2.$$

Assumption (A.1) means that the linear operator $S = S_d$ maps f into $S_d(f)$ whose G norm does not exceed a multiple of the $L_2(D)$ norm of f. Since S_d is continuous, (A.1) holds, in particular, if the embedding of F into $L_2(D)$ is continuous.

Assumption (A.2) means that the values $R_\mu(t, t)$ of the covariance kernel of the measure $\mu = \mu_d$ are bounded in the norm of the space $L_\infty(D)$.

We now relate the average case complexity functions of LMP in Λ^{all} and in Λ^{std}.

THEOREM 3.1. *Let* (A.1) *and* (A.2) *hold. Suppose that for all d there exists a nonnegative* $K_3 = K_3(d)$ *such that*

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) \leq K_3(c + 2)\left(\frac{1}{\varepsilon}\right)^{p(d)} \qquad \forall \varepsilon \in [0, 1] \tag{3.2}$$

Then

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}}) \leq K_4(c + 2)\left(\frac{1}{\varepsilon}\right)^{p(d)+2}, \qquad \forall \varepsilon \in [0, 1], \tag{3.3}$$

where

$$K_4 = \lambda(D)K_1^2K_2K_3(1 + p(d)/2)(1 + 2/p(d))^{p(d)/2}$$

with the convention that $\infty^0 = 1$.

Proof. Let $\nu = \mu S^{-1}$ be the Gaussian measure on the Hilbert space G with mean zero and covariance operator C_ν. Let

$$C_\nu \eta_j = \lambda_j \eta_j, \qquad j = 1, 2, \ldots,$$

where $\{\eta_j\}$ form a complete orthonormal system of G and $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ with trace$(C_\nu) = \sum_{i=1}^{+\infty} \lambda_i < +\infty$. It is known, see TWW (1988, p. 254), that

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) = cn^*(\varepsilon)(1 + a), \tag{3.4}$$

where $a \in [-1/n^*(\varepsilon), 2/c]$ and

$$n^*(\varepsilon) = \min \left\{ n : \sum_{i=n+1}^{+\infty} \lambda_i \leq \varepsilon^2 \right\}. \tag{3.5}$$

From the assumed form (3.2) of $\text{comp}^{\text{avg}}(d, \Lambda^{\text{all}})$ we know that

$$n^*(\varepsilon) \leq K_3 \varepsilon^{-p(d)}. \tag{3.6}$$

Let n be a given positive integer. It is known that

$$U(f) = \sum_{j=1}^{n} \langle S(f), \eta_j \rangle \, \eta_j = S(f) - \sum_{j=n+1}^{+\infty} \langle S(f), \eta_j \rangle \, \eta_j$$

has average error $\sqrt{\sum_{j=n+1}^{+\infty} \lambda_j}$. Here $\langle \cdot, \cdot \rangle$ denotes the inner product of G.

Consider a linear functional L_j, $L_j : F \to \mathbb{R}$, given by

$$L_j(f) = \langle S(f), \eta_j \rangle. \tag{3.7}$$

Note that $|L_j(f)| \leq \|S(f)\| \, \|\eta_j\| \leq K_1 \|f\|_d$ due to (A.1). Since F is a subset of $L_2(D)$, the mapping L_j can be treated as a continuous linear functional defined on a linear subspace of $L_2(D)$.

Let $\overline{F}$ denote the closure of F in the $L_2(D)$ norm. If $\overline{F} \neq F$ then we can extend the domain of S by setting $S(f^*) = \lim_{i \to +\infty} S(f_i)$, where $f^* = \lim_{i \to +\infty} f_i$. It is a well defined extension due to (A.1) and the completeness of G. This defines L_j on $\overline{F}$.

If $\overline{F} \neq L_2(D)$ then we can use the Hahn–Banach theorem to extend the functional L_j in (3.7) to $L_2(D)$ and to preserve its norm. Applying Riesz's theorem there exists a function a_j in $L_2(D)$ such that

$$L_j(f) = \int_D a_j(t) f(t) \, dt, \qquad \forall f \in L_2(D),$$

and $\|a_j\|_d \leq K_1$. Using this representation of L_j we may rewrite $U(f)$ as

$$U(f) = \sum_{j=1}^{n} L_j(f) \eta_j, \qquad \text{with } L_j(f) = \int_D a_j(t) f(t) \, dt. \tag{3.8}$$

Observe that the approximation U uses information which consists of weighted integrals of f. This type of information is allowed in Λ^{all}.

We now turn to Λ^{std} in which only function values can be used. We approximate the weighted integrals in (3.8) by the integrand values at some points. More precisely, let $\boldsymbol{t} = [t_1, t_2, \ldots, t_k] \in D^k$ denote k arbitrary points from D and let

$$N^{\text{std}}(f; \boldsymbol{t}) = [f(t_1), f(t_2), \ldots, f(t_k)]$$

denote the information which consists of k functionals from Λ^{std}. We approximate $L_j(f)$ by

$$U_j(f; \boldsymbol{t}) = \frac{\lambda}{k} \sum_{i=1}^{k} a_j(t_i) f(t_i),$$

where $\lambda = \lambda(D)$. Define the approximation $U(\cdot; \boldsymbol{t})$ which uses information from Λ^{std} as

$$U(f; \boldsymbol{t}) = \sum_{j=1}^{n} U_j(f; \boldsymbol{t}) \eta_j.$$

Consider the average error of $U(\cdot; \boldsymbol{t})$,

$$\begin{aligned} e^{\text{avg}}(U(\cdot; \boldsymbol{t}))^2 &= \int_F \|S(f) - U(f; \boldsymbol{t})\|^2 \, \mu(df) \\ &= \sum_{j=1}^{n} \int_F (\langle S(f), \eta_j \rangle - U_j(f; \boldsymbol{t}))^2 \, \mu(df) \\ &\quad + \sum_{j=n+1}^{+\infty} \int_F \langle S(f), \eta_j \rangle^2 \, \mu(df) \\ &= \sum_{j=1}^{n} \int_F (L_j(f) - U_j(f; \boldsymbol{t}))^2 \, \mu(df) + \sum_{j=n+1}^{+\infty} \lambda_j. \end{aligned} \tag{3.9}$$

The average error of $U(\cdot; \boldsymbol{t})$ depends on $\boldsymbol{t}$, i.e., on the points t_i used in the information $N^{\text{std}}(\cdot; \boldsymbol{t})$. We now integrate both sides of (3.9) with respect to $\boldsymbol{t}$,

$$\frac{1}{\lambda^k} \int_{D^k} e^{\text{avg}}(U(\cdot; \boldsymbol{t}))^2 \, d\boldsymbol{t} = \sum_{j=1}^{n} \int_F \left(\frac{1}{\lambda^k} \int_{D^k} (L_j(f) - U_j(f; \boldsymbol{t}))^2 \, d\boldsymbol{t} \right) \mu(df) + \sum_{j=n+1}^{+\infty} \lambda_j.$$

Since

$$\frac{1}{\lambda^k}\int_{D^k}(L_j(f) - U_j(f; \boldsymbol{t}))^2\,d\boldsymbol{t} = \frac{\lambda}{k}\left(\int_D a_j^2(t)f^2(t)\,dt - \frac{1}{\lambda}\left(\int_D a_j(t)f(t)\,dt\right)^2\right)$$

is the square of the error of the classic Monte Carlo algorithm, we obtain

$$\frac{1}{\lambda^k}\int_{D^k} e^{\text{avg}}(U(\cdot, \boldsymbol{t}))^2\,d\boldsymbol{t} \le \frac{\lambda}{k}\sum_{j=1}^{n}\int_D a_j^2(t)\int_F f^2(t)\mu(df)\,dt + \sum_{j=n+1}^{+\infty}\lambda_j$$

$$= \frac{\lambda}{k}\sum_{i=1}^{n}\int_D a_j^2(t)R_\mu(t, t)\,dt + \sum_{j=n+1}^{+\infty}\lambda_j.$$

Due to (A.2) and the bound on a_j we finally get

$$\frac{1}{\lambda^k}\int_{D^k} e^{\text{avg}}(U(\cdot, \boldsymbol{t}))^2\,d\boldsymbol{t} \le \lambda(D)K_1^2K_2\frac{n}{k} + \sum_{j=n+1}^{+\infty}\lambda_j. \tag{3.10}$$

Applying the mean value theorem to the left side of (3.10), we conclude that there exists a vector $\boldsymbol{t}^*$, i.e., k points $t_1^*, t_2^*, \ldots, t_k^*$ which form the information $N^{\text{std}}(\cdot; \boldsymbol{t}^*)$, such that the average error of $U^* = U(\cdot; \mathbf{t}^*)$ satisfies the inequality

$$e^{\text{avg}}(U^*)^2 = \int_F \|S(f) - U^*(f)\|^2\,\mu(df) \le \lambda(D)K_1^2K_2\frac{n}{k} + \sum_{j=n+1}^{+\infty}\lambda_j.$$

Let $p = p(d)$ and let x_p minimize the function $g(x) = (x^p(1 - x^2))^{-1}$ for $x \in [0, 1]$. Then $g(x_p) = (1 + p/2)(1 - 2/(p + 2))^{-p/2}$.

Take now $n = n^*(x_p\varepsilon)$ and $k = \lambda(D)K_1^2K_2n^*(x_p\varepsilon)/((1 - x_p^2)\varepsilon^2)$. From the definition of $n^*(\cdot)$, see (3.5), and (3.6), we obtain $\sum_{j=n^*(x_p\varepsilon)+1}^{+\infty}\lambda_j \le x_p^2\varepsilon^2$ and $k \le K_4\varepsilon^{-p(d)-2}$. This yields

$$e^{\text{avg}}(U^*) \le \varepsilon.$$

Thus U^* computes ε-approximations and uses only function values. To estimate the average cost of U^* note that $U^*(f)$ can be rewritten as

$$U^*(f) = \sum_{i=1}^{k} f(t_i^*)g_i, \qquad \text{where } g_i(t) = \frac{\lambda}{k}\sum_{j=1}^{n} a_j(t_i^*)\eta_j(t).$$

Since the functions g_i can be precomputed, the average cost of U^* is bounded by

$$\text{cost}^{\text{avg}}(U^*) \leq (c + 2)k \leq K_4(c + 2)\varepsilon^{-p(d)-2}.$$

Obviously, $\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}}) \leq \text{comp}^{\text{avg}}(U^*)$, and (3.3) follows as claimed. ■

Theorem 3.1 states that the exponents of the average case complexity functions in Λ^{all} and Λ^{std} may differ by at most 2. The constant 2 cannot be improved as can be proven by considering the integration problem, $S(f) = \int_D f(t)\, dt$. Thus, $K_1 = 1$. Obviously, $S \in \Lambda^{\text{all}}$ and $\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) \leq c$. Hence, $p(d) = 0$ and $K_3 = 1$ in (3.2).

Wasilkowski (1991) proves that for the Wiener isotropic measure the average case complexity in Λ^{std} is of order $\varepsilon^{-2/(1+1/d)}$. Thus, in this case the exponent in (3.3) is $2/(1 + 1/d)$ and cannot be replaced by a number which is smaller than 2 for all d.

The presence of 2 in the exponent of (3.3) can also be expected in view of related results on the average case complexity of integration with respect to a worst probability measure due to Novak (1988) and Mathé (1991).

We stress that the proof of Theorem 3.1 is *not* constructive. Indeed, we use the mean value theorem to conclude the existence of sample points t_i^* at which the function f should be evaluated to solve the problem with average cost of order $(1/\varepsilon)^{p(d)+2}$. We address the issue of constructing sample points t_i^* in Section 4.

Remark 3.1. We stress that the relation between the complexity functions in Λ^{all} and Λ^{std} holds in the *average case* setting. Such a relation does *not* hold, in general, if we switch to the worst case setting. To see this, consider the integeration problem. Then the worst case complexity in Λ^{all} remains constant whereas the worst case complexity in Λ^{std} can be an essentially arbitrary increasing function of $1/\varepsilon$ as proven by Werschulz (1985).

From Theorem 3.1 we immediately conclude that tractability of LMP in Λ^{all} and Λ^{std} coincide. Indeed, if $p(d)$ is bounded by p for all d in Λ^{all} then $p(d) + 2$ is bounded by $p + 2$ for all d in Λ^{std}. This also shows that the difference between the exponents of LMP in Λ^{all} and Λ^{std} can be at most 2. We summarize this in the following corollary.

COROLLARY 3.1. *Let* (A.1) *and* (A.2) *hold.*

(i) *A linear multivariate problem LMP is tractable in* Λ^{all} *iff this linear multivariate problem LMP is tractable in* Λ^{std}.

(ii) *If* $p^*(\Lambda)$ *is the exponent of LMP in* Λ *then*

$$p^*(\Lambda^{\text{std}}) \leq p^*(\Lambda^{\text{all}}) + 2.$$

As a simple application, consider a linear multivariate problem LMP for which S is a continuous linear functional. Since $S \in \Lambda^{\text{all}}$, we have

$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) \le c$ and $p(d) = 0$. Thus, such a LMP is tractable in Λ^{all} with exponent $p^* = 0$ and we have the following corollary.

COROLLARY 3.2. *Let* (A.1) *and* (A.2) *hold. If* $S \in \Lambda^{\text{all}}$, *then LMP is tractable in* Λ^{std} *with exponent at most* 2, *i.e.*,

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}}) = O(c\varepsilon^{-2}).$$

Hence, the average case complexity of any continuous linear functional in Λ^{std} is at most of order ε^{-2}. As mentioned before, the proof of this fact is not constructive and to find sample points which achieve such a bound can be a challenging problem.

We now provide a simple condition to check tractability of LMP = $\{\text{LMP}_d\}$ in Λ^{all}, or in Λ^{std} if (A.1) and (A.2) hold. As in the proof of Theorem 3.1, let

$$\nu_d = \mu_d S_d^{-1}$$

be the Gaussian measure on G with mean zero and covariance operator $C_{\nu,d}$ with eigenvalues $\lambda_i = \lambda_i(d)$ such that $\lambda_1(d) \ge \lambda_2(d) \ge \cdots \ge 0$, and $\text{trace}(C_{\nu,d}) = \sum_{i=1}^{+\infty} \lambda_i(d) < +\infty$.

THEOREM 3.2. *A linear multivariate problem LMP* = $\{LMP_d\}$ *is tractable in* Λ^{all} *iff there exists a positive number* α *such that for all* d,

$$\sum_{i=n+1}^{+\infty} \lambda_i(d) = O\left(\left(\frac{1}{n}\right)^{2\alpha}\right), \qquad \textit{as } n \to +\infty. \tag{3.11}$$

The exponent of LMP is $p^* = +\infty$ *if there exists no* α *satisfying* (3.11); *otherwise*

$$p^* = 1/\sup\{\alpha\text{: } \alpha \textit{ satisfies } (3.11)\}.$$

If α *satisfies* (3.11) *then*

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) = O\left(c\left(\frac{1}{\varepsilon}\right)^{1/\alpha}\right),$$

and if also (A.1) *and* (A.2) *hold then*

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}}) = O\left(c\left(\frac{1}{\varepsilon}\right)^{2+1/\alpha}\right).$$

Proof. From (3.4) and (3.6) of the proof of Theorem 3.1 we know that

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) = \Theta(cn^*(\varepsilon)).$$

Suppose that (3.11) holds. Then (3.5) yields $n^*(\varepsilon) = O(\varepsilon^{-1/\alpha})$ which proves tractability of LMP with $p = 1/\alpha$. This also yields that $p^* \le 1/\sup A$, where A is the set of α's which satisfy (3.11).

Suppose now that LMP is tractable. Then $n^*(\varepsilon) = O(\varepsilon^{-p})$. Hence, (3.5) yields

$$\sum_{i=n^*(\varepsilon)+1}^{+\infty} \lambda_i(d) = O(\varepsilon^2), \qquad \text{as } \varepsilon \to 0.$$

Setting $n = n^*(\varepsilon)$ we obtain

$$\sum_{i=n+1}^{+\infty} \lambda_i(d) = O(n^{-2/p}), \qquad \text{as } n \to +\infty.$$

Thus, (3.11) hold with $\alpha = 1/p$. Since p can be arbitrarily close to p^*, we conclude that $\sup A \ge 1/p^*$. Hence, $p^* = 1/\sup A$, as claimed for tractable problems.

On the other hand, if LMP is not tractable then the set A is the empty and $p^* = +\infty$. The rest of the proof easily follows from Theorem 3.1. ∎

Theorem 3.2 states that tractability of LMP in either class depends on how fast the truncated trace of the covariance operator $C_{\nu,d}$ tends to zero. Tractability holds iff the speed of this convergence is of the form $n^{-2\alpha}$ with α independent of d. Theorem 3.2 solves Problem 3 in Traub and Woźniakowski (1991) for linear operators and Gaussian measures.

We now discuss *adaptive* information for tractable linear multivariate problems. If $\Lambda = \Lambda^{\text{all}}$ then it is known that adaption does not help; see Wasilkowski (1986) and TWW(1988, p. 247) where this result is also reported.

If $\Lambda = \Lambda^{\text{std}}$ then adaptive information may be more powerful than nonadaptive information. However, as we now show, for tractable problems adaption can help only by a multiplicative constant which depends on the exponent p^* of LMP.

As in TWW (1988, p. 249), let $n^{\text{avg}}(\varepsilon)$ denote the minimal number of *nonadaptive* evaluations which are needed to compute ε-approximations. Let

$$N = [L_1, L_2, \ldots, L_n], \qquad L_i \in \Lambda^{\text{std}}, \; n = n^{\text{avg}}(\varepsilon), \qquad (3.12)$$

be such information. As explained in TWW (1988, p. 225) we may assume that $L_i(C_\mu L_j) = \delta_{i,j}$. Define the linear approximation

$$U(f) = \sum_{j=1}^{n} L_j(f) S(C_\mu L_j). \tag{3.13}$$

Then $e^{\text{avg}}(U) \le \varepsilon$ and $\text{cost}^{\text{avg}}(U) \le (c + 2)n^{\text{avg}}(\varepsilon)$ since $S(C_\mu L_j)$ can be precomputed.

From Theorems 5.7.1 and 5.7.2 of TWW (1988, pp. 248–249) we know that for any $x > 1$,

$$c \min \left\{ n^{\text{avg}}(x\varepsilon), \frac{x^2 - 1}{x^2} n^{\text{avg}}(\varepsilon) \right\} \le \text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}}) \le (c + 2)n^{\text{avg}}(\varepsilon).$$

Since $n^{\text{avg}}(x\varepsilon) \le n^{\text{avg}}(\varepsilon)$, the minimum of the left hand side is at least $(x^2 - 1)/x^2 n^{\text{avg}}(x\varepsilon)$. Hence,

$$c \frac{x^2 - 1}{x^2} n^{\text{avg}}(x\varepsilon) \le \text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}}) \le (c + 2)n^{\text{avg}}(\varepsilon), \qquad \forall x > 1. \tag{3.14}$$

Assume now that LMP is tractable and

$$\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}}) \le Kc \left(\frac{1}{\varepsilon}\right)^p, \qquad \forall \varepsilon > 0. \tag{3.15}$$

Then (3.14) yields that

$$n^{\text{avg}}(\varepsilon) \le \frac{x^{p+2}}{x^2 - 1} K \left(\frac{1}{\varepsilon}\right)^p, \qquad \forall \varepsilon > 0.$$

Take x which minimizes the function $x^{p+2}/(x^2 - 1)$. That is, $x = +\infty$ for $p = 0$, and $x = \sqrt{(p + 2)/p}$ for $p > 0$. The minimum is equal to a_p with

$$a_p = \begin{cases} 1 & \text{if } p = 0, \\ \frac{1}{2}p(1 + 2p^{-1})^{(p+2)/2} & \text{if } p \ne 0. \end{cases} \tag{3.16}$$

Observe that a_p is an increasing function of p and

$$a_1 = 3\sqrt{3}/2, \quad a_2 = 4, \quad a_4 = 6\tfrac{3}{4}, \text{ and}$$
$$a_p = \tfrac{1}{2}pe(1 + o(1)), \quad \text{as } p \to +\infty.$$

Hence,

$$n^{\text{avg}}(\varepsilon) \le a_p K \left(\frac{1}{\varepsilon}\right)^p,$$

and the approximation U given by (3.13) that uses *nonadaptive* N given by (3.12) computes ε-approximations with

$$\text{cost}^{\text{avg}}(U) \leq (c+2)n^{\text{avg}}(\varepsilon) \leq a_p K(c+2)\left(\frac{1}{\varepsilon}\right)^p.$$

That is, the average cost U is at most a_p times larger than the bound on the average case complexity in (3.15) since, usually, $c \gg 1$ and the difference between c and $c + 2$ is negligible. Note that p in (3.15) can be taken arbitrarily close to the exponent p^* of LMP. This leads to the following corollary.

COROLLARY 3.3. *For a tractable linear multivariate problem with exponent p^* in Λ^{std}, adaption helps at most by a multiplicative factor a_{p^*}.*

Tractability of a linear multivariate problem LMP depends, in particular, on the linear operators S_d. We now show that sometimes it is enough to ascertain tractability of one specific linear multivariate problem and conclude tractability of other LMPs.

This specific problem is called *multivariate function approximation* and is defined as a linear multivariate problem APP = $\{\text{APP}_d\}$, where $S_d = I_d$ is a continuous embedding operator, $I_d: F \to L_2(D)$ with $I_d(f) = f$ and $\|I_d(f)\|_d \leq K^*\|f\|_F$, $\forall f \in F$, for some constant $K^* = K^*(d)$. The other parameters F, μ, Λ of APP_d are defined as in Section 2. As always, $\Lambda \in \{\Lambda^{\text{all}}, \Lambda^{\text{std}}\}$. Note that (A.1) trivially holds for APP with $K_1 = 1$.

THEOREM 3.3. *Suppose that multivariate function approximation APP is tractable in Λ with exponent p^*. Consider a linear multivariate problem LMP which differs from APP by the choice of the operator S_d, $S_d: F \to G$. If (A.1) holds for S_d then LMP is tractable in Λ with exponent at most equal to p^*.*

Proof. Take a linear U^* which solves APP_d with average error at most ε/K_1,

$$U^*(f) = \sum_{j=1}^{n} L_j(f) I_d(C_\mu L_j),$$

where $L_j(C_\mu L_i) = \delta_{i,j}$. Due to tractability of APP in Λ and Corollary 3.3 we can take $n = O(\varepsilon^{-p^*-\delta})$ for a positive δ. The average cost of U^* is at most $(c+2)n = O(c\varepsilon^{-p^*-\delta})$ since $I_d(C_\mu L_j)$ can be precomputed.

For LMP = $\{\text{LMP}_d\}$ with a linear operator S_d which satisfy (A.1), define

$$U(f) = \sum_{j=1}^{n} L_j(f) S_d(C_\mu L_j).$$

Observe that U is well defined since $C_\mu L_j \in F$ and $U(f) \in G$. From (A.1) we have

$$\|S_d(f) - U(f)\| \leq \|S_d\| \|I_d(f) - U^*(f)\| \leq K_1 \|I_d(f) - U^*(f)\|$$

and therefore

$$e^{\text{avg}}(U) \leq K_1 e^{\text{avg}}(U^*) \leq \varepsilon.$$

Furthermore $\text{cost}^{\text{avg}}(U) \leq (c + 2)n = O(c\varepsilon^{-p^*-\delta})$ since $S(C_\mu L_j)$ can be precomputed. This yields

$$\text{comp}^{\text{avg}}(\varepsilon; \text{LMP}_d) = O(c\varepsilon^{-p^*-\delta}).$$

Hence, LMP is tractable in Λ and its exponent $p^*(\text{LMP})$ is bounded by $p^* + \delta$. Since δ arbitrary, $p^*(\text{LMP}) \leq p^*$, as claimed. ∎

4. Construction of Sample Points

We analyze the construction of optimal (or nearly optimal) sample points for linear multivariate problems in the average case setting by utilizing relations between average case and worst case settings. In this way, we reduce the construction of optimal sample points in the average case setting to the same problem in the worst case setting. The construction of optimal sample points in the worst case setting is known for a number of cases. This will enable us in Part II to exhibit optimal (or nearly optimal) sample points for multivariate integration and function approximation in the average case setting.

In Section 4.1 we consider the approximation of continuous linear functionals, while in Section 4.2 we consider the general case of approximating continuous linear operators.

4.1. *Continuous Linear Functionals*

We analyze linear multivariate problems $\text{LMP}_d = \{F, \mu, G, S, \Lambda^{\text{std}}\}$ for which $G = \mathbb{R}$. That is, S is a continuous linear functional, $S: F \to \mathbb{R}$. We assume that S satisfies (A.1) which now means that S is also continuous with respect to the norm of $L_2(D)$. Since F is a subspace of $L_2(D)$ we can claim, as in the proof of Theorem 3.1, that

$$S(f) = \int_D \rho(t) f(t)\, dt, \qquad \forall f \in F \subset L_2(D).$$

Here, ρ is a fixed function from the space $L_2(D)$.

A linear multivariate problem with $G = \mathbb{R}$ will be called *multivariate weighted integration* and denoted by $\rho\mathrm{INT} = \{\rho\mathrm{INT}_d\}$. If $\rho \equiv 1$ then we call such a problem *multivariate integration* and denote by $\mathrm{INT} = \{\mathrm{INT}_d\}$.

Obviously, ρINT is tractable in Λ^{all} with the exponent $p^* = 0$. Since $\rho \in L_2(D)$, S satisfies (A.1) with $K_1 = \|\rho\|_d$. Hence, assuming (A.2), ρINT is tractable in Λ^{std} with exponent $p^* \leq 2$, see Corollary 3.2.

We now show that the *average* case complexity of ρINT is closely related to the *worst* case complexity of the same S restricted to a specific subset of F. This specific subset of F is the unit ball BH_μ of a reproducing kernel Hilbert space H_μ which is defined in terms of the Gaussian measure μ of the space F.

The space H_μ is the completion of finite dimensional spaces of the form

$$\mathrm{span}(R_\mu(\cdot, x_1), R_\mu(\cdot, x_2) \, . \, . \, . \, , R_\mu(\cdot, x_k))$$

for any integer k and any points x_i from D. As before, R_μ is the covariance kernel of μ, see (2.1). The completion is with respect to the norm $\|\cdot\|_\mu = \langle \cdot, \cdot \rangle_\mu^{1/2}$, where the inner product is defined by

$$\langle f, g \rangle_\mu = \sum_{j=1}^{k} \sum_{i=1}^{m} a_j b_i R_\mu(t_i, x_j)$$

for any $f = \sum_{j=1}^{k} a_j R_\mu(\cdot, x_j)$ and $g = \sum_{i=1}^{m} b_i R_\mu(\cdot, t_i)$.

The space H_μ is a subset of $C_\mu(F^*) \subset F$ since $R_\mu(\cdot, x) = C_\mu L_x \in C_\mu(F^*)$, where C_μ is the covariance operator of μ and $L_x(f) = f(x)$. In the reproducing kernel Hilbert space H_μ we have

$$f(x) = \langle f, R_\mu(\cdot, x) \rangle_\mu, \qquad \forall f \in H_\mu, \forall x \in D.$$

We define a linear multivariate problem $\overline{\rho\mathrm{INT}}_d = (BH_\mu, \mathbb{R}, S, \Lambda^{\mathrm{std}})$. This problem will be considered in the worst case setting. Its worst case complexity, $\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho\mathrm{INT}}_d)$, is defined as the minimal cost over all approximations U whose error does not exceed ε. Here, cost and error of U are defined as in Section 2 with the integrals replaced by the supremum over the unit ball BH_μ, see TWW (1988, Chap. 3).

We now show that the average case complexity of multivariate weighted integration in the class F,

$$\rho\mathrm{INT}_d = (F, \mu, G, S, \Lambda^{\mathrm{std}}),$$

is closely related to the worst case complexity of multivariate weighted integration in the unit ball BH_μ,

$$\overline{\rho\mathrm{INT}}_d = (BH_\mu, \mathbb{R}, S, \Lambda^{\mathrm{std}}).$$

THEOREM 4.1. (i) *For any* $x > 1$ *we have*

$$\frac{c}{c+2}\,\frac{x^2-1}{x^2}\,\mathrm{comp}^{\mathrm{wor}}(x\varepsilon;\overline{\rho INT_d}) \le \mathrm{comp}^{\mathrm{avg}}(\varepsilon;\rho INT_d) \le \frac{c+2}{c}\,\mathrm{comp}^{\mathrm{wor}}(\varepsilon;\overline{\rho INT_d}).$$

(ii) *If*

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon;\rho INT_d) \le Kc\left(\frac{1}{\varepsilon}\right)^p$$

then

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon;\overline{\rho INT_d}) \le a_p K(c+2)\left(\frac{1}{\varepsilon}\right)^p,$$

where a_p *is given by* (3.16).

Proof. The basic step of the proof is to use a known relation between the average and worst errors of linear algorithms that use nonadaptive information. This relation has been used in many papers as indicated in the introduction.

Consider nonadaptive information

$$N(f) = [f(x_1), f(x_2), \ldots, f(x_n)]$$

with fixed x_i from D, and a linear approximation

$$U(f) = \sum_{j=1}^{n} a_j f(x_j), \qquad a_j \in \mathbb{R}. \tag{4.1}$$

To stress the dependence on the class F, let

$$e^{\mathrm{avg}}(U;F) = \left(\int_F (S(f)-U(f))^2\mu(df)\right)^{1/2}$$

denote the average error of U in F. Let

$$e^{\mathrm{wor}}(U;BH_\mu) = \sup_{f\in BH_\mu} |S(f)-U(f)|$$

denote the worst error of U in the unit ball BH_μ of H_μ. Then it is known that the average error of U in F coincides with the worst error of U in BH_μ,

$$e^{\text{avg}}(U; F) = e^{\text{wor}}(U; BH_\mu) = \|h^*\|_\mu, \tag{4.2}$$

where

$$h^*(x) = S(R_\mu(x, \cdot)) - U(R_\mu(x, \cdot)) = \int_D \rho(t)R_\mu(x, t)\, dt - \sum_{j=1}^n a_j R_\mu(x, x_j).$$

Let $n^{\text{avg}}(\varepsilon)$ be, as in Section 3, the minimal cardinality of nonadaptive information needed to compute an ε-approximation. Since μ is Gaussian there exists a linear U that uses nonadaptive information of cardinality $n^{\text{avg}}(\varepsilon)$ such that $e^{\text{avg}}(U) \le \varepsilon$ and $\text{cost}^{\text{avg}}(U) \le (c + 2)n^{\text{avg}}(\varepsilon)$. This linear U is of the form (4.1), with a_j and x_j chosen to minimize the average error.

Let

$$e_n = \inf_{x_1,\ldots,x_n \in D}\ \inf_{a_1,\ldots,a_n \in \mathbb{R}} \left\| \int_D \rho(t)R_\mu(\cdot, t)\, dt - \sum_{j=1}^n a_j R_\mu(\cdot, x_j) \right\|_\mu \tag{4.3}$$

be the minimal norm of the function h^* in (4.2). Due to (4.2), e_n is also the minimal average error of ρINT which can be achieved after n *nonadaptive* function evaluations. We thus have

$$n^{\text{avg}}(\varepsilon) = \min\{n\colon e_n \le \varepsilon\};$$

see also TWW (1988, p. 304). From (3.14) we know that $n^{\text{avg}}(\varepsilon)$ is closely related to the average case complexity by

$$c\,\frac{x^2 - 1}{x^2}\, n^{\text{avg}}(x\varepsilon) \le \text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d) \le (c + 2)n^{\text{avg}}(\varepsilon), \quad \forall x > 1. \tag{4.4}$$

We now consider the worst case complexity of $\overline{\rho\text{INT}}_d$. It is known, see Bakhvalov (1971), that adaption does not help in the worst case setting for linear functionals S. Let $n^{\text{wor}}(\varepsilon)$ be the minimal cardinality of *nonadaptive* information which allows us to compute ε-approximations in the worst case setting for $\overline{\rho\text{INT}}_d$; see TWW (1988, p. 101). Due to Smolyak's theorem, see e.g., TWW (1988, p. 76), the worst error of algorithms that use nonadaptive information is minimized by a linear U. Due to (4.2), e_n is thus the minimal worst error of $\overline{\rho\text{INT}}$ which can be achieved after n *adaptive* function evaluations. Thus

$$n^{\text{wor}}(\varepsilon) = n^{\text{avg}}(\varepsilon).$$

It is known that

$$cn^{\mathrm{wor}}(\varepsilon) \le \mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho\mathrm{INT}_d}) \le (c+2)n^{\mathrm{wor}}(\varepsilon). \tag{4.5}$$

From this and (4.4), (i) of Theorem 4.1 easily follows.

To show (ii), note that the bound on $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \rho\mathrm{INT}_d)$ and the left hand side of (4.4) imply that $n^{\mathrm{avg}}(\varepsilon) = n^{\mathrm{wor}}(\varepsilon) \le x^{p+2}/(x^2-1)K\varepsilon^{-p}$. Taking $x = +\infty$ for $p = 0$ and $x = \sqrt{(p+2)/p}$ for $p > 0$, (ii) follows from the right-hand side of (4.5). ■

Since $c + 2 \simeq c$, Theorem 4.1 states that the average case complexity of $\rho\mathrm{INT}_d$ is no greater than the worst case complexity of $\overline{\rho\mathrm{INT}_d}$. These two complexity functions can differ only if adaption helps in the average case setting. However, as (ii) of Theorem 4.1 states, if the average case complexity is bounded by $Kc\varepsilon^{-p}$ then adaption can help at most by a multiplicative factor a_p. Since now $p \le 2$, the factor a_p is bounded by 6.75.

How much adaption can help depends on the sequence $\{e_n^2\}$, see (4.3), as analyzed by Wasilkowski (1986). In particular, if $\{e_n^2\}$ is convex, i.e., $e_n^2 \le (e_{n-1}^2 + e_{n+1}^2)/2$, $\forall n > 2$, adaption does not help and $\mathrm{comp}^{\mathrm{avg}}(\varepsilon) \ge c(n^{\mathrm{avg}}(\varepsilon) - 1)$. If $\{e_n^2\}$ is semiconvex, i.e., $\alpha^2\alpha_n \le e_n^2 \le \beta^2\beta_n$ for positive α and β and convex sequences α_n and β_n, then $\mathrm{comp}^{\mathrm{avg}}(\varepsilon) \ge c(n^{\mathrm{avg}}(\varepsilon\beta/\alpha) - 1)$. As we shall see in Part II, e_n is often of the form $\Theta(n^{-1/p}(\log n)^q)$ for some positive p and q. In this case, the sequence $\{e_n^2\}$ is semiconvex. We summarize this in the following corollary.

COROLLARY 4.1. *Let $\{e_n^2\}$ be given by* (4.3).

(i) *If $\{e_n^2\}$ is convex then*

$$\frac{c}{c+2}\,\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho INT_d}) - c \le \mathrm{comp}^{\mathrm{avg}}(\varepsilon; \rho INT_d) \le \frac{c+2}{c}\,\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho INT_d}).$$

Thus, for large c,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \rho INT_d) \simeq \mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho INT_d}) \simeq cn^{\mathrm{avg}}(\varepsilon) = cn^{\mathrm{wor}}(\varepsilon) = c\min\{n\colon e_n \le \varepsilon\}.$$

(ii) *If $\{e_n^2\}$ is semiconvex, $\alpha^2\alpha_n \le e_n^2 \le \beta^2\beta_n$, then*

$$\frac{c}{c+2}\,\mathrm{comp}^{\mathrm{wor}}\Big(\varepsilon\frac{\beta}{\alpha}; \overline{\rho INT_d}\Big) - c \le \mathrm{comp}^{\mathrm{avg}}(\varepsilon; \rho INT_d) \le \frac{c+2}{c}\,\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \overline{\rho INT_d}).$$

In particular, if $e_n = \Theta(n^{-1/p}(\log n)^q)$, *where* $p = p(d) > 0$ *and* $q = (d)$, *then*

$$\text{comp}^{\text{avg}}(\varepsilon; \rho INT_d) = \Theta(\text{comp}^{\text{wor}}(\varepsilon; \overline{\rho INT_d})) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^p \left(\log\frac{1}{\varepsilon}\right)^{pq}\right),$$

and the exponent p^* *of* ρINT *is given by*

$$p^* = \sup\{p(d): d = 1, 2, \ldots\} \le 2.$$

We now show how to obtain optimal (or nearly optimal) sample points for ρINT_d in the average case setting. Let $n = n^{\text{wor}}(\varepsilon)$. Without loss of generality, we may assume that the infima in (4.3) are attained for x_j^* and a_j^*. Let

$$U_n^*(f) = \sum_{j=1}^{n} a_j^* f(x_j^*). \tag{4.6}$$

Then (4.2) yields

$$e^{\text{wor}}(U_n^*; BH_\mu) = e_n \le \varepsilon, \qquad \text{and} \qquad \text{cost}^{\text{wor}}(U_n^*) \le (c + 2)n^{\text{wor}}(\varepsilon).$$

For large c, from (4.5) we conclude that

$$\text{comp}^{\text{wor}}(\varepsilon; \overline{\rho INT_d}) \simeq c\,\text{cost}^{\text{wor}}(U_n^*).$$

This means that the sample points $\{x_1^*, x_2^*, \ldots, x_n^*\}$ and U_n^* are *optimal* for $\overline{\rho INT_d}$ in the worst case setting.

Since $e^{\text{avg}}(U_n^*; F) = e_n \le \varepsilon$ and $\text{cost}^{\text{avg}}(U_n^*) \le (c + 2)n$, proceeding as in Corollary 4.1, we conclude that the sample points $\{x_1^*, x_2^*, \ldots, x_n^*\}$ and U_n^* also enjoy optimality properties for ρINT_d in the average case setting. More precisely we have the following corollary.

COROLLARY 4.2. *Let* $\{e_i^2\}$ *be given by* (4.3).

(i) *Let* $n = n^{\text{wor}}(\varepsilon)$. *If* $\{e_i^2\}$ *is convex then for large* c *the sample points* $\{x_1^*, x_2^*, \ldots, x_n^*\}$ *and* U_n^* *given by* (4.6) *are optimal for* ρINT_d *in the average case setting,*

$$\text{comp}^{\text{avg}}(\varepsilon; \rho INT_d) \simeq \text{cost}^{\text{avg}}(U_n^*) \simeq cn^{\text{wor}}(\varepsilon).$$

(ii) *Let* $e_i = \Theta(i^{-1/p}(\log i)^q)$, *where* $p = p(d) > 0$ *and* $q = q(d)$. *Let* n *be the smallest integer for which* $e_n \le \varepsilon$. *Then, modulo a multiplicative factor which may depend on* d, *the sample points* $\{x_1^*, x_2^*, \ldots, x_n^*\}$ *and* U_n^* *given by* (4.6) *are optimal for* ρINT_d *in the average case setting,*

$$\text{comp}^{\text{avg}}(\varepsilon; \rho INT_d) = \Theta(\text{cost}^{\text{avg}}(U_n^*)) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^p\left(\log\frac{1}{\varepsilon}\right)^{pq}\right).$$

The essence of Theorem 4.1 and Corollaries 4.1, 4.2 is that the average case complexity of multivariate integration ρINT_d, optimal sample points and U_n^* can be found by using the known results about the worst case of multivariate integration $\overline{\rho\text{INT}}_d$. In Part II we shall provide a number of examples.

Remark 4.1: Worst Case in Reproducing Kernel Hilbert Spaces. Suppose that (A.2) holds. Then (ii) of Theorem 4.1 holds with $p = 2$. This means that the worst case complexity of $\overline{\rho\text{INT}}$ is at most of order $(1/\varepsilon)^2$. This fact can be proven directly without using the relation to the average case setting. Indeed, consider an *arbitrary* reproducing kernel Hilbert space H of functions defined on D with kernel R. For $\rho \in L_2(D)$ assume that

$$S(f) = \int_D \rho(t) f(t)\, dt, \qquad \forall f \in H,$$

is a continuous linear functional and that $\|R(\cdot, \cdot)\|_{L_\infty(D)} < +\infty$.

Define a linear approximation $U(f) = \lambda(D) n^{-1} \sum_{j=1}^n \rho(t_j) f(t_j)$. The second equality of (4.2) states that the worst error, $e^{\text{wor}}(U; BH)$, of U in the unit ball BH is equal to the norm of h^* with

$$h^*(x; \mathbf{t}) = \int_D \rho(t) R(x, t)\, dt - \frac{\lambda(D)}{n} \sum_{j=1}^n \rho(t_j) R(x, t_j),$$

with $\mathbf{t} = [t_1, t_2, \ldots, t_n]$. Let $\lambda = \lambda(D)$. Integrating the square norm of h^* with respect to t_j we obtain

$$\frac{1}{\lambda^n}\int_{D^n} \|h^*(\cdot; \mathbf{t})\|^2\, dt_1 \cdots dt_n$$
$$= \frac{\lambda}{n}\left(\int_D \rho^2(t) R(t, t)\, dt - \frac{1}{\lambda}\int_D \int_D \rho(t)\rho(x) R(x, t)\, dt\, dx\right).$$

Due to the mean value theorem, there exist points $t_1, t_2, \ldots, t_n$ such that

$$e^{\text{wor}}(U; BH) = \|h^*(\cdot; \mathbf{t})\|$$
$$\leq \sqrt{\frac{\lambda(D)}{n}}\left(\int_D \rho^2(t) R(t, t)\, dt - \frac{1}{\lambda(D)}\int_D \int_D \rho(t)\rho(x) R(x, t)\, dt\, dx\right)^{1/2}$$
$$\leq \sqrt{\frac{\lambda(D)}{n}}\, \|\rho\|_{L_2(D)} \|R(\cdot, \cdot)\|_{L_\infty(D)}^{1/2}.$$

This implies that the worst case complexity is at most of order $(1/\varepsilon)^2$. We stress, however, that the proof is not constructive.

We now exhibit a relation between multivariate weighted integration ρINT in the average case setting and multivariate function approximation $\mathrm{APP}^{\mathrm{wor},2} = \{\mathrm{APP}_d^{\mathrm{wor},2}\}$ in the worst case setting. Multivariate approximation $\mathrm{APP}_d^{\mathrm{wor},2}$ is defined by $(BH_\mu, L_2(D), I_d, \Lambda^{\mathrm{std}})$, where I_d is an embedding between BH_μ and $L_2(D)$. Since BH_μ is a subset of $L_2(D)$, I_d is well defined. Recall that $\|\cdot\|_d$ denotes the norm of $L_2(D)$.

For multivariate weighted integration $\rho\mathrm{INT}_d$, consider nonadaptive information $N(f) = [f(x_1), f(x_2), \ldots, f(x_n)]$ and a linear approximation $U_\rho(f) = \sum_{j=1}^n a_j f(x_j)$. Then (4.2) states that

$$e^{\mathrm{avg}}(U_\rho; F) = e^{\mathrm{wor}}(U_\rho; BH_\mu).$$

It is known that if we choose a_j to minimize the worst error $e^{\mathrm{wor}}(U_\rho; BH_\mu)$ then

$$e^{\mathrm{wor}}(U_\rho; BH_\mu) = \sup_{f(x_j)=0, \|f\|_\mu \le 1} \left| \int_D \rho(t) f(t)\, dt \right|.$$

From this we have (see also Novak, 1988)

$$e^{\mathrm{wor}}(U_\rho; BH_\mu) \le \|\rho\|_d \sup_{f(x_j)=0, \|f\|_\mu \le 1} \|f\|_d,$$

$$\sup_{\|\rho\|_d \le 1} e^{\mathrm{wor}}(U_\rho; BH_\mu) = \sup_{f(x_j)=0, \|f\|_\mu \le 1} \|f\|_d. \tag{4.7}$$

Let $e_n^{\mathrm{avg}}(\rho\mathrm{INT}_d)$ denote the minimal average error of any algorithm that uses nonadaptive information of cardinality n for multivariate weighted integration $\rho\mathrm{INT}_d$. Let $e_n^{\mathrm{wor}}(\overline{\rho\mathrm{INT}}_d)$ be the corresponding minimal worst error for $\overline{\rho\mathrm{INT}}_d$. Since linear algorithms minimize the errors, (4.2) yields

$$e_n^{\mathrm{avg}}(\rho\mathrm{INT}_d) = e_n^{\mathrm{wor}}(\overline{\rho\mathrm{INT}}_d) \tag{4.8}$$

Consider now multivariate function approximation $\mathrm{APP}^{\mathrm{wor},2} = \{\mathrm{APP}_d^{\mathrm{wor},2}\}$ in the worst case setting. Let

$$e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},2}) = \inf_{x_1,\ldots,x_n} \inf_{\phi} \sup_{\|f\|_\mu \le 1} \|f - \phi(f(x_1) \ldots, f(x_n))\|_d$$

be the minimal worst error of any algorithm that uses nonadaptive information of cardinality n for $\mathrm{APP}_d^{\mathrm{wor},2}$. It is also known (see Micchelli and

Rivlin, [1977]; this result can also be found in TWW, 1988, p. 80) that linear algorithms minimize the error and

$$e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2}) = \inf_{x_1,\ldots,x_n} \sup_{f(x_j)=0, \|f\|_\mu \le 1} \|f\|_d.$$

This, (4.7), and (4.8) yield

$$\sup_{\|\rho\|_d \le 1} e_n^{\text{avg}}(\rho\text{INT}_d) \le e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2}). \tag{4.9}$$

This means that the average error of multivariate weighted integration ρINT_d with an arbitrary weight ρ, $\|\rho\|_d \le 1$, is at most equal to the worst error of multivariate function approximation $\text{APP}_d^{\text{wor},2}$.

This permits us to relate the average case complexity of ρINT_d to the worst case complexity $\text{comp}^{\text{wor}}(\varepsilon; \text{APP}_d^{\text{wor},2})$. Since adaption does not help in the worst case setting for $\text{APP}_d^{\text{wor},2}$, see e.g., TWW (1988, p. 59), we have

$$\text{comp}^{\text{wor}}(\varepsilon; \text{APP}_d^{\text{wor},2}) = (c + a) \min\{n: e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2}) \le \varepsilon\}, \qquad a \in [0, 2].$$

Repeating the second part of the proof of Theorem 4.1, we conclude from (4.9) the following corollary.

COROLLARY 4.3. (i) *We have*

$$\sup_{\|\rho\|_d \le 1} \text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d) \le \frac{c+2}{c} \text{comp}^{\text{wor}}(\varepsilon; \text{APP}_d^{\text{wor},2}).$$

(ii) *If* $e_n^{\text{wor},2}(\text{APP}_d^{\text{wor},2}) = O(n^{-1/p}(\log n)q)$, *where* $p = p(d) > 0$ *and* $q = q(d)$, *then*

$$\sup_{\|\rho\|_d \le 1} \text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d) = O\left(c \left(\frac{1}{\varepsilon}\right)^p \left(\log \frac{1}{\varepsilon}\right)^{pq}\right),$$

and the exponent p^* *of* ρINT *is bounded by*

$$p^* \le \sup\{p(d): d = 1, 2, \ldots\}.$$

Corollary 4.3 states that the weight ρ, $\|\rho\|_d \le 1$, may increase the average case complexity of ρINT_d up to at most the worst case complexity of multivariate function approximation of $\text{APP}_d^{\text{wor},2}$.

Remark 4.2: "Easy Weights." One may ask the rather theoretical question as to whether there are weights ρ, $\|\rho\|_d = 1$, for which the average case complexity of ρINT_d is essentially easier than

$$\sup_{\|\rho\|_d \leq 1} \text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d).$$

Indeed, it may happen that for some weights ρ, multivariate weighted integration is trivial since $\text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d) \leq c + 1$. To show this, assume that R_μ is continuous at a point $(t^*, t^*) \in D^2$. For a given positive ε, choose δ such that

$$|R_\mu(t, x) - R_\mu(t^*, t^*)| \leq \varepsilon^2 \qquad \text{for } t, x \in D_\delta,$$

and such that $\alpha = \int_{t \in D_\delta} dt \leq \frac{1}{3}$. Here, $D_\delta = \{t \in D: \|t - t^*\|_\infty \leq \delta\}$.

Define the weight

$$\rho(t) = \begin{cases} \alpha^{-1/2} & \text{if } t \in D_\delta, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\|\rho\|_d = 1$. Define the linear approximation

$$U(f) = \sqrt{\alpha}\, f(t^*).$$

The average error of U is given by

$$\begin{aligned} e^{\text{avg}}(U)^2 &= \frac{1}{\alpha} \int \int_{t,x \in D_\delta} R_\mu(t, x)\, dt\, dx - 2 \int_{t \in D_\delta} R_\mu(t, t^*)\, dt + \alpha R_\mu(t^*, t^*) \\ &= \frac{1}{\alpha} \int \int_{t,x \in D_\delta} (R_\mu(t, x) - R_\mu(t^*, t^*))\, dt\, dx \\ &\quad + 2 \int_{t \in D_\delta} (R_\mu(t, t^*) - R_\mu(t^*, t^*))\, dt \\ &\leq \varepsilon^2 \alpha + 2\varepsilon^2 \alpha = 3\alpha\varepsilon^2 \leq \varepsilon^2. \end{aligned}$$

Hence, $e^{\text{avg}}(U) \leq \varepsilon$ and the cost of U is at most $c + 1$, as claimed.

Observe that if there exists a continuity point (t^*, t^*) of R_μ for which $R_\mu(t^*, t^*) = 0$ then we can set $U(f) = 0$ since $f(t^*) = 0$ with probability 1. Then $\text{comp}^{\text{avg}}(\varepsilon; \rho\text{INT}_d) = 0$. This is the case for the folded Wiener sheet measure with $t^* = 0$; see Part II.

From the analysis presented above it is clear how to use sample points and algorithms, that are optimal for multivariate function approximation in the worst case setting, for multivariate weighted integration in the average case setting, see also Sacks and Ylvisaker (1970b). Indeed, we may choose the sample points x_j^* and the functions h_j^* such that

$$T_n^*(f, t) = \sum_{j=1}^{n} f(x_j^*) h_j^*(t)$$

minimizes the worst case error for $\mathrm{APP}_d^{\mathrm{wor},2}$,

$$e^{\mathrm{wor}}(T_n^*; BH_\mu) = e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},2}).$$

For multivariate weighted integration $\rho\mathrm{INT}_d$, define

$$U_\rho(f) = \int_D \rho(t) T_n^*(f, t)\, dt = \sum_{j=1}^{n} f(x_j^*) \int_D \rho(t) h_j^*(t)\, dt. \tag{4.10}$$

Then (4.2) yields

$$\begin{aligned} e^{\mathrm{avg}}(U_\rho; F) = e^{\mathrm{wor}}(U_\rho; BH_\mu) &= \sup_{\|f\|_\mu \le 1} \left| \int_D \rho(t)(f(t) - T_n(f, t))\, dt \right| \\ &\le \|\rho\|_d \|f - T_n(f)\|_d \le \|\rho\|_d\, e^{\mathrm{wor}}(T_n; BH_\mu). \end{aligned}$$

Thus, if we choose n such that $e^{\mathrm{wor}}(T_n^*; BH_\mu) \le \varepsilon/\|\rho\|_d$ then $e^{\mathrm{avg}}(U_\rho; F) \le \varepsilon$. Since the integrals of ρh_j^* can be precomputed, the cost of U is at most $(c + 2)n$ which is essentially the same as the cost of T_n^*. From this and Corollary 4.3 we conclude the following.

COROLLARY 4.4. *Let* $\|\rho\|_d \le 1$. *Multivariate weighted integration* $\rho\mathrm{INT}_d$ *can be solved in the average case setting by using sample points* x_j^* *and functions* h_j^* *of* (4.10), *which are optimal for multivariate function approximation in the worst case setting, with average cost at most* $(1 + 2/c)\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \mathrm{APP}_d^{\mathrm{wor},2})$.

4.2. *Continuous Linear Operators*

We now study general linear multivariate problems $\mathrm{LMP}_d = \{F, \mu, G, S, \Lambda^{\mathrm{std}}\}$. We assume that S satisfies (A.1). As already mentioned, we may attack LMP_d by two approaches. The first one is to express LMP_d as a number of continuous linear functionals and apply the analysis of Section 4.1. This is done in Section 4.2.1. The second approach is to estimate S_d by a multiple of I_d and switch to the multivariate function approximation problem $\mathrm{APP}_d = \{F, \mu, L_2(D), I_d, \Lambda^{\mathrm{std}}\}$. We show that APP_d is related to a multivariate approximation problem in the worst case setting. This is done in Section 4.2.2.

4.2.1. *Approach 1: Weighted Integrals.* As in Section 3, see (3.8), let

$$U(f) = \sum_{j=1}^{k} \langle S(f), \eta_j \rangle \eta_j = \sum_{j=1}^{k} L_j(f)\eta_j, \quad L_j(f) = \int_D a_j(t)f(t)\,dt. \quad (4.11)$$

Then U has average error $\sqrt{\sum_{j\ge k+1} \lambda_j}$ and $\|a_j\|_d \le K_1$. The approximation U uses information from Λ^{all} which is now not allowed. Since we can use only function values, we must replace $L_j(f)$ by appropriate approximation composed of, say, n function values.

As in Section 4.1, we replace $L_j(f)$ by $U_{a_j}(f)$ of (4.10) and we get

$$U_n^{\text{std}}(f) = \sum_{j=1}^{k} L_j(T_n^*(f))\eta_j = \sum_{j=1}^{k} f(x_j^*)g_j, \quad (4.12)$$

where

$$g_j = \sum_{i=1}^{k} \left(\int_D a_i(t)h_j^*(t)\,dt\right)\eta_i.$$

Observe that g_j's do not depend on f and they can be precomputed.

Thus, U_n^{std} is linear, uses information from Λ^{std}, and can be computed at cost $(c + 2)n$. It is easy to estimate the average error U_n^{std}. Indeed, using the estimates of Section 4.1, we obtain

$$e^{\text{avg}}(U_n^{\text{std}})^2 = e^{\text{avg}}(U_n^{\text{std}}; \text{LMP}_d)^2 \le K_1^2 k e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2})^2 + \sum_{j=k+1}^{+\infty} \lambda_j.$$

We now choose k to minimize the estimate of $e^{\text{avg}}(U_n^{\text{std}})$. Assume that

$$\sqrt{\sum_{j\ge k+1} \lambda_j} = \Theta\left(\frac{(\log k)^q}{k^{1/p}}\right),$$

$$e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2}) = O\left(\frac{(\log n)^{q_1}}{n^{1/p_1}}\right), \quad (4.13)$$

where $p = p(d)$ and $p_1 = p_1(d)$ are positive, and $q = q(d)$ and $q_1 = q_1(d)$. Then taking

$$k = \Theta(n^{2p/(p_1(p+2))}\,(\log n)^{2p(q-q_1)/(p+2)})$$

we obtain

$$e^{\text{avg}}(U_n^{\text{std}}) = O\left(\frac{(\log n)^{(pq+2q_1)/(p+2)}}{n^{2/(p_1(p+2))}}\right).$$

Obviously,

$$e^{\text{avg}}(U_n^{\text{std}}) \geq \sqrt{\sum_{j \geq n+1} \lambda_j} = \Theta\left(\frac{(\log n)^q}{n^{1/p}}\right).$$

This proves that $1/p_1 \leq 1/p + 1/2$, and if $1/p_1 = 1/p + 1/2$ then $q \leq q_1$.

In general, there is no further relation between (p, q) and (p_1, q_1). Indeed, it may happen that the weights a_j are "easy" and the difference between $L_j(f)$ and $L_j(T_n^*(f))$ is much smaller on the average than $e^{\text{wor}}(\text{APP}_d^{\text{wor},2})$. (The extreme case is to take $S = 0$. Then $a_j = 0$ and $p = q = 0$, although (p_1, q_1) could be positive.)

On the other hand, if $1/p_1 = 1/p + 1/2$ then

$$e^{\text{avg}}(U_n^{\text{std}}) = O\left(\frac{(\log n)^{q+2(q_1-q)/(p+2)}}{n^{1/p}}\right).$$

Thus, modulo a power of $\log n$, the average error of U_n^{std} is minimal. To guarantee that $e^{\text{avg}}(U_n^{\text{std}}) \leq \varepsilon$ we take $n = \Theta(\varepsilon^{-p}(\log 1/\varepsilon)^{pq+(q_1-q)/(1/p+1/2)})$. Then the average cost of U_n^{std} is, modulo a power of $\log 1/\varepsilon$, equal to the average case complexity $\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{all}})$. This shows that Λ^{std} is almost as powerful as Λ^{all}. We summarize this discussion in the following theorem.

THEOREM 4.2. *Suppose that S satisfies* (A.1) *and that in* (4.13) $1/p_1 = 1/p + 1/2$. *Then the average case complexity functions* $\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}})$ *and* $\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{all}})$ *of* LMP_d *differ at most by a power of* $\log 1/\varepsilon$,

$$\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}}) = O\left(c\left(\frac{1}{\varepsilon}\right)^p \left(\log\frac{1}{\varepsilon}\right)^{pq+(q_1-q)/(1/p+1/2)}\right),$$

$$\text{comp}^{\text{avg}}(\varepsilon; \Lambda^{\text{std}}) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^p \left(\log\frac{1}{\varepsilon}\right)^{pq}\right).$$

For $n = \Theta(\varepsilon^{-p}(\log 1/\varepsilon)^{pq+(q_1-q)/(1/p+1/2)})$, *the sample points* x_j^* *and* U_n^{std} *given by* (4.12) *are, modulo a power of* $\log 1/\varepsilon$, *optimal in the average case setting for* LMP_d.

If $1/p_1(d) = 1/p(d) + 1/2$, $\forall d$, *then the exponents of* LMP *in* Λ^{std} *and* Λ^{all} *are the same and are equal to*

$$p^* = \sup\{p(d)\colon d = 1, 2, \ldots\}.$$

In Part II we shall show that the assumptions of Theorem 4.2 are satisfied for some multivariate function approximation problems.

4.2.2. *Approach 2: Worst Case Multivariate Function Approximation.* As explained in Section 3, we can use the results on multivariate function approximation to linear multivariate problems LMP_d which differ from APP_d only in the definition of the operator S. Hence, it is enough to analyze multivariate function approximation $\text{APP} = \{\text{APP}_d\}$ with $\text{APP}_d = \{F, \mu, L_2(D), I_d, \Lambda^{\text{std}}\}$.

In Section 4.1 we have already used relations between multivariate weighted integration in the average case and worst case settings. In this section we obtain similar relations for multivariate function approximation.

Consider a linear U which uses sample points x_j,

$$U(f) = \sum_{j=1}^{n} f(x_j)g_j, \qquad g_j \in L_2(D). \tag{4.14}$$

We now show that the average error of U is equal to

$$e^{\text{avg}}(U) = e^{\text{avg}}(U; \text{APP}_d) = \left(\int_D \|h^*(\cdot, x)\|_\mu^2 \, dx\right)^{1/2}, \tag{4.15}$$

where

$$h^*(\cdot, x) = R_\mu(\cdot, x) - \sum_{j=r}^{n} g_j(x)R_\mu(\cdot, x) \in H_\mu.$$

As always, R_μ is the covariance kernel of μ and $\|\cdot\|_\mu$ is the norm in the reproducing kernel space H_μ, see Section 4.1. Indeed, observe that

$$\begin{aligned} e^{\text{avg}}(U; \text{APP}_d)^2 &= \int_F \int_D \left(f(x) - \sum_{j=1}^{n} f(x_j)g_j(x)\right)^2 dx \, \mu(df) \\ &= \int_D \left(R_\mu(x, x) - 2\sum_{j=1}^{n} g_j(x)R_\mu(x, x_j) \right. \\ &\quad \left. + \sum_{i,j=1}^{n} g_i(x)g_j(x)R_\mu(x_i, x_j)\right) dx. \end{aligned}$$

On the other hand, since $\langle R_\mu)\cdot, x), R_\mu(\cdot, t)\rangle_\mu = R_\mu(t, x)$, we have

$$\int_D \|h^*(\cdot, x)\|_\mu^2 \, dx = \int_D \left\| R_\mu(\cdot, x) - \sum_{j=1}^n g_j(x) R_\mu(\cdot, x_j) \right\|_\mu^2 dx$$

$$= \int_D \Big(R_\mu(x, x) - 2 \sum_{j=1}^n g_j(x) R_\mu(x, x_j)$$

$$+ \sum_{i,j=1}^n g_j(x) g_i(x) R_\mu(x_i, x_j) \Big) \, dx,$$

which proves (4.15).

Consider now the same linear U for multivariate function approximation

$$\mathrm{APP}_d^{\mathrm{wor},2} = \{BH_\mu, L_2(D), I_d, \Lambda^{\mathrm{std}}\}$$

in the worst case setting. Then for the worst error U we have

$$e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor},2})^2 = \sup_{\|f\|_\mu \le 1} \int_D \Big(f(x) - \sum_{j=1}^n f(x_j) g_j(x) \Big)^2 dx$$

$$= \sup_{\|f\|_\mu \le 1} \int_D \Big\langle f, R_\mu(\cdot, x) - \sum_{j=1}^n g_j(x) R_\mu(\cdot, x_j) \Big\rangle_\mu^2 dx$$

$$\le \sup_{\|f\|_\mu \le 1} \|f\|_\mu^2 \int_D \|h^*(\cdot, x)\|_\mu^2 \, dx = \int_D \|h^*(\cdot, x)\|_\mu^2 \, dx.$$

From (4.15) we thus obtain

$$e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor},2}) \le e^{\mathrm{avg}}(U; \mathrm{APP}_d). \tag{4.16}$$

Finally, consider the same U for multivariate function approximation in the $L_\infty(D)$ norm

$$\mathrm{APP}_d^{\mathrm{wor},\infty} = \{BH_\mu, L_\infty(D), I_d, \Lambda^{\mathrm{std}}\}$$

in the worst case setting. We now assume that H_μ is a subset of $L_\infty(D)$ and that the embedding I_d maps H_μ into $L_\infty(D)$. We also assume that the functions g_j of (4.14) belong to $L_\infty(D)$. The worst error of U is now equal to

$$e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor},\infty}) = \sup_{\|f\|_\mu \le 1} \left\| f - \sum_{j=1}^n f(x_j) g_j \right\|_{L_\infty(D)}$$

$$= \sup_{\|f\|_\mu \le 1} \operatorname*{ess\,sup}_{x \in D} \left| f(x) - \sum_{j=1}^n f(x_j) g_j(x) \right|.$$

It is easy to show that

$$e^{\text{wor}}(U, \text{APP}_d^{\text{wor},\infty}) = \operatorname*{ess\,sup}_{x \in D} \|h^*(\cdot, x)\|_\mu. \tag{4.17}$$

Indeed,

$$e^{\text{wor}}(U; \text{APP}_d^{\text{wor},\infty}) = \sup_{\|f\|_\mu \le 1} \operatorname*{ess\,sup}_{x \in D} \left| \left\langle f, R_\mu(\cdot, x) - \sum_{j=1}^{n} g_j(x) R_\mu(\cdot, x_j) \right\rangle_\mu \right|$$

$$\le \sup_{\|f\|_\mu \le 1} \operatorname*{ess\,sup}_{x \in D} \|f\|_\mu \|h^*(\cdot, x)\|_\mu = \operatorname*{ess\,sup}_{x \in D} \|h^*(\cdot, x)\|_\mu.$$

To prove the reverse inequality, assume without loss of generality that there exists x_0 such that

$$\|h^*(\cdot, x_0)\|_\mu = \operatorname*{ess\,sup}_{x \in D} \|h^*(\cdot, x)\|_\mu.$$

Then taking $f = h^*(\cdot, x_0)/\|h^*(\cdot, x_0)\|_\mu$ we get $e^{\text{wor}}(U; \text{APP}_d^{\text{wor},\infty}) \ge \|h^*(\cdot, x_0)\|_\mu$ which implies (4.17). Since

$$\operatorname*{ess\,sup}_{x \in D} \|h^*(\cdot, x)\|_\mu \ge \frac{1}{\sqrt{\lambda(D)}} \left(\int_D \|h^*(\cdot, x)\|_\mu^2 \, dx \right)^{1/2},$$

we conclude from (4.17) and (4.15) that

$$e^{\text{avg}}(U; \text{APP}_d) \le \sqrt{\lambda(D)}\, e^{\text{wor}}(U; \text{APP}_d^{\text{wor},\infty}).$$

This and (4.16) yield the following lemma.

LEMMA 4.1. *For multivariate function approximation, the average error of a linear U of* (4.14) *is bounded from below by the worst error in the L_2 norm and, if $g_j \in L_\infty(D)$, it is bounded from above by a multiple of the worst error in L_∞ norm,*

$$e^{\text{wor}}(U; \text{APP}_d^{\text{wor},2}) \le e^{\text{avg}}(U; \text{APP}_d) \le \sqrt{\lambda(D)}\, e^{\text{wor}}(U; \text{APP}_d^{\text{wor},\infty}).$$

In general, as we see later, the first inequality of Lemma 4.1 is not sharp. The second inequality is, modulo a power of log n, sharp for some multivariate function approximation problems.

We now relate the average case complexity of APP_d to the worst case complexity of $\text{APP}_d^{\text{wor},2}$ and $\text{APP}_d^{\text{wor},\infty}$. Let $e_n^{\text{wor}}(\text{APP}_d^{\text{wor},2})$, $e_n^{\text{wor}}(\text{APP}_d^{\text{wor},\infty})$ and $e_n^{\text{avg}}(\text{APP}_d)$ denote the minimal worst (or average) error of any algorithm that uses nonadaptive information of cardinality n for the corresponding problem $\text{APP}_d^{\text{wor},2}$, $\text{APP}_d^{\text{wor},\infty}$ or APP_d. Since linear algorithms minimize the error for all three problems, Lemma 4.1 implies

$$e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},2}) \le e_n^{\mathrm{avg}}(\mathrm{APP}_d) \le \sqrt{\lambda(D)}\, e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},\infty}).$$

Since adaption does not help for $\mathrm{APP}_d^{\mathrm{wor},2}$ and $\mathrm{APP}_d^{\mathrm{wor},\infty}$, we have following formulas for their worst case complexity functions:

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon;\, \mathrm{APP}_d^{\mathrm{wor},2}) = (c + a_1) \min\{n\colon e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},2}) \le \varepsilon\}, \qquad a_1 \in [0, 2],$$

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon;\, \mathrm{APP}_d^{\mathrm{wor},\infty}) = (c + a_2) \min\{n\colon e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},\infty}) \le \varepsilon\}, \qquad a_2 \in [0, 2],$$

Proceeding as in the second part of the proof of Theorem 4.1, the following theorem is obtained from Lemma 4.1.

THEOREM 4.3. (i) *For any $x > 1$ we have*

$$a\,\frac{x^2 - 1}{x^2}\, \mathrm{comp}^{\mathrm{wor}}(x\varepsilon;\, \mathrm{APP}_d^{\mathrm{wor},2}) \le \mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \mathrm{APP}_d) \le a^{-1}\, \mathrm{comp}^{\mathrm{wor}}(\varepsilon b;\, \mathrm{APP}_d^{\mathrm{wor},\infty}),$$

where $a = c/(c + 2)$ and $b = 1/\sqrt{\lambda(D)}$.

(ii) *Suppose that*

$$e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},2}) = \Omega\left(\frac{(\log n)^{q_1}}{n^{1/p_1}}\right),\quad e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},\infty}) = O\left(\frac{(\log n)^{q_2}}{n^{1/p_2}}\right),$$

where $p_1 = p_1(d)$ and $p_2 = p_2(d)$ are positive, and $q_1 = q_1(d)$ and $q_2 = q_2(d)$. Then

$$\Omega\left(c\left(\frac{1}{\varepsilon}\right)^{p_1}\left(\log\frac{1}{\varepsilon}\right)^{p_1q_1}\right) = \mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \mathrm{APP}_d) = O\left(c\left(\frac{1}{\varepsilon}\right)^{p_2}\left(\log\frac{1}{\varepsilon}\right)^{p_2q_2}\right).$$

In general, the lower bound in Theorem 4.3 on $\mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \mathrm{APP}_d)$ is not sharp. We can sometimes obtain a better lower bound on $\mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \mathrm{APP}_d)$ by using the corresponding average case complexity for Λ^{all}. Indeed, assume that for multivariate function approximation $(F, \mu, L_2(D), I_d, \Lambda^{\mathrm{all}})$ we have

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \Lambda^{\mathrm{all}}) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{p}\left(\log\frac{1}{\varepsilon}\right)^{pq}\right).$$

for some p and q. If $p = p_2$, then the upper bound in Theorem 4.3 on $\mathrm{comp}^{\mathrm{avg}}(\varepsilon;\, \mathrm{APP}_d)$ is sharp modulo a power of $\log(1/\varepsilon)$. In this case, we can find optimal sample points and an optimal linear U. More precisely, let the sample points x_j^* and the function $h_j^*, j = 1, 2, \ldots, n$, be chosen such that

$$U_n^*(f) = \sum_{j=1}^{n} f(x_j^*) h_j^*, \qquad h_j^* \in L_\infty(D), \tag{4.18}$$

minimizes the worst case error for $\mathrm{APP}_d^{\mathrm{wor},\infty}$,

$$e^{\mathrm{wor}}(U_n^*; \mathrm{APP}_d^{\mathrm{wor},\infty}) = e_n^{\mathrm{wor}}(\mathrm{APP}_d^{\mathrm{wor},\infty}).$$

Then from Theorem 4.3 we conclude the following corollary.

COROLLARY 4.5. *Suppose that* $p_2 = p$.

(i) *Then the average case complexity functions* $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \Lambda^{\mathrm{std}})$ *and* $\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \Lambda^{\mathrm{all}})$ *of multivariate function approximation* APP_d *in the average case setting as well as the worst case complexity* $\mathrm{comp}^{\mathrm{wor}}(\varepsilon; \mathrm{APP}_d^{\mathrm{wor},\infty})$ *differ at most by a power of* $\log 1/\varepsilon$, *and*

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon; \Lambda^{\mathrm{std}}) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^p \left(\log\frac{1}{\varepsilon}\right)^{p(q+\alpha(\varepsilon))}\right),$$

where $\alpha \in [0, q_2 - q]$.

(ii) *For* $n = \Theta(\varepsilon^{-p}(\log 1/\varepsilon)^{pq_2})$, *the sample points* x_j^* *and* U_n^* *given by* (4.18) *are, modulo a power of* $\log 1/\varepsilon$, *optimal for multivariate function approximation* APP_d *in the average case setting.*

(iii) *If* $p_2(d) = p(d), \forall d$, *then the exponents of* APP *in* Λ^{std} *and* Λ^{all} *are the same and equal to*

$$p^* = \sup\{p(d): d = 1, 2, \ldots\}.$$

In Part II we shall show that the assumption $p(d) = p_2(d), \forall d$, holds for some multivariate function approximation problems.

ACKNOWLEDGMENTS

I thank A. Papageorgiou, S. Paskov, L. Plaskota, J. F. Traub, G. W. Wasilkowski, and A. G. Werschulz for valuable comments.

REFERENCES

BAKHVALOV, N. S. (1971), On the optimality of linear methods for operator approximation in convex classes of functions, *USSR Comput. Math. Math. Phys.* **11**, 244–249.

KIMELDORF, G. S., AND WAHBA, G. (1970a), A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41,** 495–502.

KIMELDORF, G. S., AND WAHBA, G., (1970b), Spline functions and stochastic processes, *Sankhya Ser. A* **32,** 173–180.

KUO, H.-H. (1975), "Gaussian Measures in Banach Spaces," Lectures Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin.

MATHÉ, P. (1991), "A Minimax Principle in Information-Based Complexity," in progress.

MICCHELLI, C. A., AND RIVLIN, T. J. (1977), A survey of optimal recovery, *in* "Optimal Estimation in Approximation Theory" (C. A. Micchelli and T. J. Rivlin, Eds.), pp 1–54, Plenum, New York.

MICCHELLI, C. A., AND WAHBA, G., (1981), Design problems for optimal surface interpolation, *in* "Approximation Theory and Applications" (Z. Ziegler, Ed.), pp. 329–347 Academic Press, New York.

NOVAK, E. (1988), "Deterministic and Stochastic Error Bounds in Numerical Analysis," Lectures Notes in Math., Vol. 1349, Springer-Verlag, Berlin.

PASKOV, S. (1991), "Average Case Complexity of Multivariate Integration for Smooth Functions," Dept. of Computer Science, Columbia University, to appear in J. Complexity.

SACKS, J., AND YLVISAKER, D. (1966), Designs for regression with correlated errors, *Ann. Math. Statist.* **37,** 68–89.

SACKS, J., AND YLVISAKER, D. (1968), Designs for regression problems with correlated errors; many parameters, *Ann. Math. Statist.* **39,** 49–69.

SACKS, J., AND YLVISAKER, D. (1970a), Designs for regression problems with correlated errors III, *Ann. Math. Statist.* **41,** 2057–2074.

SACKS, J., AND YLVISAKER, D. (1970b), Statistical design and integral approximation, Proc. 12th Bienn. Semin. Can. Math. Cong." pp. 115–136.

TRAUB, J. F., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1988), "Information-Based Complexity," Academic Press, New York.

TRAUB, J. F., AND WOŹNIZKOWSKI, H., (1991), Information-based complexity: New questions for mathematicians, *Math. Intell.* **13,** 34–43.

VAKHANIA, N. N. (1981), "Probability Distributions on Linear Spaces," North-Holland, New York.

WAHBA, G. (1971), On the regression design problem of Sacks and Ylvisaker, *Ann. Math. Statist.* **42,** 1035–1043.

WAHBA, G., (1990), "Spline Models for Observational Data," CBMS 59, SIAM.

WASILKOWSKI, G. W., (1986), Information of varying cardinality, *J. Complexity* **2,** 204–228.

WASILKOWSKI, G. W. (1991), "Integration and Approximation of Multivariate Functions: Average Case Complexity with Isotropic Wiener measure," Dept. of Computer Science, University of Kentucky, to appear in BAMS.

WERSCHULZ, A. G. (1985), Counterexamples in optimal quadratures, *Aequationes Math.* **29,** 183–202.

YLVISAKER, D. (1975), Designs on random fields, *in* " A Survey of Statistical Design and Linear Models" (J. Srivastava, Ed.), pp. 593–607, North-Holland, Amsterdam.