

THE POWER OF STANDARD INFORMATION FOR MULTIVARIATE APPROXIMATION IN THE RANDOMIZED SETTING

G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI

ABSTRACT. We study approximating multivariate functions from a reproducing kernel Hilbert space with the error between the function and its approximation measured in a weighted L_2 -norm. We consider functions with an arbitrarily large number of variables, d , and we focus on the *randomized setting* with algorithms using *standard* information consisting of function values at randomly chosen points.

We prove that standard information in the randomized setting is *as powerful* as *linear* information in the worst case setting. Linear information means that algorithms may use arbitrary continuous linear functionals, and by the power of information we mean the speed of convergence of the n th minimal errors, i.e., of the minimal errors among all algorithms using n function evaluations. Previously, it was only known that standard information in the randomized setting is *no more powerful* than the linear information in the worst case setting.

We also study (*strong*) *tractability* of multivariate approximation in the randomized setting. That is, we study when the minimal number of function evaluations needed to reduce the initial error by a factor ε is polynomial in ε^{-1} (strong tractability), and polynomial in d and ε^{-1} (tractability). We prove that these notions in the randomized setting for standard information are equivalent to the same notions in the worst case setting for linear information. This result is useful since for a number of important applications only standard information can be used and verifying (strong) tractability for standard information is in general difficult, whereas (strong) tractability in the worst case setting for linear information is known for many spaces and is relatively easy to check.

We illustrate the tractability results for weighted Korobov spaces. In particular, we present necessary and sufficient conditions for strong tractability and tractability. For product weights independent of d , we prove that strong tractability is equivalent to tractability.

We stress that all proofs are constructive. That is, we provide randomized algorithms that enjoy the maximal speed of convergence. We also exhibit randomized algorithms which achieve strong tractability and tractability error bounds.

Received by the editor December 8, 2005.

2000 *Mathematics Subject Classification.* Primary 41A63, 65C05, 65D15, 65Y20.

Key words and phrases. Weighted multivariate approximation, randomized algorithms, Monte Carlo methods, tractability.

1. INTRODUCTION

Multivariate approximation is one of the most commonly studied problems in information-based complexity. The main reason is that many other continuous problems are intrinsically related to multivariate approximation. Examples include multivariate integration, solution of partial differential or integral equations, quasi-linear problems, and some non-linear problems; see for example [25, 30] as well as Section 4.

In this paper, by *multivariate approximation* we mean approximation of functions defined on $D \subseteq \mathbb{R}^d$ that belong to a reproducing kernel Hilbert space. We stress that the number of variables, d , can be arbitrarily large. The distance between the function f and its approximation $A(f)$ provided by an algorithm A is measured in the ρ -weighted L_2 -norm,

$$\|f - A(f)\| = \left(\int_D |f(\mathbf{x}) - A(f)(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x} \right)^{1/2}$$

for a probability density function ρ .

The algorithm errors can be studied in the worst case, average case, asymptotic and/or randomized settings; see e.g., [13, 19, 20, 21, 22, 23, 29] for more details. We focus here on the randomized setting and show relations between the randomized and worst case settings.

We consider algorithms that use a finite number of information evaluations about the function f being approximated, where one information evaluation corresponds to the evaluation of one linear functional. In a number of applications, information evaluation is constrained to be a function value at some point; this corresponds to the class Λ^{std} of *standard information*. If evaluation of an arbitrary continuous linear functional is allowed, then this corresponds to the class Λ^{all} of *linear information*.

These two classes, Λ^{std} and Λ^{all} , have often been studied for multivariate approximation. We measure their power by the speed of convergence of the n th minimal errors, i.e., of the minimal errors among all algorithms using n information evaluations. More precisely, we are looking for the largest exponents p^{all} and p^{std} for which we can find algorithms using n information evaluations from Λ^{all} and Λ^{std} whose errors are of order $n^{-p^{\text{all}}}$ and $n^{-p^{\text{std}}}$, respectively. Obviously, $p^{\text{all}} \geq p^{\text{std}}$ which means that Λ^{std} is at most as powerful as Λ^{all} . Due to the practical significance of Λ^{std} , it is important to verify for which spaces and settings the class Λ^{std} is as powerful as the class Λ^{all} , i.e., when

$$p^{\text{all}} = p^{\text{std}}.$$

The analysis needed to determine the maximal speed of convergence is usually much easier for the class Λ^{all} and is *constructive*. For example, for the multivariate approximation problem studied in this paper, it is known that the algorithms with minimal worst case errors are fully determined by the eigenpairs of a certain compact operator W which depends only on the reproducing kernel and the weight ρ of the L_2 -norm; see Section 2.1. Furthermore, the n th minimal errors are the square-roots of the eigenvalues λ_{n+1} of the operator W . Hence, the power of the class Λ^{all} in the worst case setting is fully determined by the speed of convergence of $\sqrt{\lambda_n}$ to zero, and $p^{\text{all}} = p$ if $\lambda_n = \Theta(n^{-2p})$. A similar characterization holds in the average case setting.

The analysis for the class Λ^{std} is much harder. For general reproducing kernel Hilbert spaces, there are a few *non-constructive* results relating the power of standard information to the power of linear information; see e.g., [5, 28] for the worst case setting and [8, 28] for the average case setting. In the worst case setting, it was shown that $p^{\text{std}} \geq p^{\text{all}}(1 - \beta)$ for some positive β depending on the space, whereas in the average case setting it was shown that $p^{\text{std}} = p^{\text{all}}$. These results were obtained under some assumptions on the Hilbert spaces. In particular, in the average case settings it was assumed that eigenfunctions of the operator W normalized in the L_2 -norm are uniformly bounded in the L_∞ -norm. The last assumption holds for weighted Korobov spaces; see Section 5. For some specific spaces there are a number of *constructive* results for multivariate approximation in the worst case and average case settings (see e.g., [9, 10, 27]); however, with the exception of [27], it is still open if the algorithms proposed there achieve the speed of convergence n^{-p} with the optimal exponent $p = p^{\text{std}}$.

There are so far no constructive results showing that $p^{\text{std}} = p^{\text{all}}$ for all d either in the worst case, randomized, or average case setting. In this paper, we address this question for multivariate approximation in the randomized setting.

Before we describe the main results of this paper, we briefly recall what has been known so far for multivariate approximation in the randomized setting; see, e.g., [6, 11, 13, 15, 21, 26]. It was proven in [15, 26] that linear information in the randomized setting is *no more powerful* than the linear information in the worst case setting. That is, if we denote the maximal exponent of convergence in the worst case setting for the class Λ^{all} by $p^{\text{all-wor}}$, and the maximal exponents of convergence in the randomized setting for the classes Λ^{all} and Λ^{std} by $p^{\text{all-ran}}$ and $p^{\text{std-ran}}$, respectively, then

$$p^{\text{std-ran}} \leq p^{\text{all-ran}} = p^{\text{all-wor}}.$$

The main result of this paper, Theorem 1, states that standard information in the randomized setting is as powerful as linear information in the worst case setting and, hence, both of them are equally powerful as linear information in the randomized setting, i.e.,

$$(1) \quad p^{\text{std-ran}} = p^{\text{all-ran}} = p^{\text{all-wor}}.$$

We stress that this holds when multivariate approximation is defined over reproducing kernel Hilbert spaces and the error is also measured in a Hilbert space. It is proved in [6, 11] that (1) is not true if multivariate approximation and its error are defined over some Banach spaces.

The proof of (1) is *constructive*. That is, we provide randomized algorithms and distributions of their sample points with the errors achieving the speed of convergence with the exponents arbitrarily close to $p^{\text{all-wor}}$. More precisely, for any integer k , we construct randomized algorithms using nk function values at randomly chosen sample points with the randomized error of order n^{-p_k} , where

$$p_k = p^{\text{all-wor}} \left[1 - \left(\frac{2p^{\text{all-wor}}}{2p^{\text{all-wor}} + 1} \right)^k \right].$$

We also show that by taking k of order $\ln \ln n$ we obtain an algorithm that uses $m = O(n \ln(\ln(n)))$ function values and whose error is of order $m^{-p^{\text{all-wor}}}$ times $(\ln(\ln(m)))^{p^{\text{all-wor}}+0.5}$. Hence, modulo a power of the double logarithm of m we achieve the best possible speed of convergence.

For given k , the randomized algorithm uses k different distributions each for n sample points, and all distributions depend on the eigenpairs of the operator W from the worst case setting. If such distributions are hard to implement, one can use only one simpler distribution of the sample points under an additional assumption. Indeed, the same speed of convergence can be achieved by using independent and identically distributed sample points whose distribution uses the weight ρ as its probability density function, providing that the normalized in the ρ -weighted L_2 -norm eigenfunctions of the operator W are uniformly bounded in the L_∞ -norm; see Theorem 2. This assumption is the same as in [8], where it was used for the analysis of the average case setting, and it holds for weighted Korobov spaces; see Section 5. If, however, the last assumption is not satisfied, and instead the values of the kernel of the Hilbert space are uniformly bounded, then we provide randomized algorithms using nk function values with convergence of order n^{-p_k} with p_k arbitrarily close to $1/2$. Hence, in this case, we may lose the speed of convergence if the eigenvalues of W tend to zero faster than $n^{-1/2}$. Note, however, that the speed of convergence $n^{-1/2}$ is independent of d which parallels the well-known and celebrated speed of convergence of the classical Monte Carlo algorithm for multivariate integration.

As already mentioned, results for multivariate approximation can be applied for other problems. To illustrate this point, we apply them in Section 4 to derive efficient randomized algorithms for approximating weighted integrals $\int_D f(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x}$. More precisely, we apply the classical Monte Carlo algorithm with the well-known variance reduction technique utilizing the multivariate approximation algorithms of this paper. When we do this with the three multivariate approximation algorithms mentioned above, we obtain three integration algorithms whose randomized errors are \sqrt{n} times smaller than the randomized errors of the corresponding multivariate approximation algorithms. In particular, our algorithms guarantee the speed of convergence arbitrarily close to n^{-1} (or higher) for all uniformly bounded reproducing kernels of the Hilbert spaces. Needless to say this is much faster convergence than the convergence of the classical Monte Carlo method.

We also study (strong) tractability of multivariate approximation in the randomized setting for the class Λ^{std} . That is, we analyze when the minimal number of randomized function values, $n^{\text{ran}}(\varepsilon, d)$, needed to reduce the initial error by a factor ε depends polynomially on d and ε^{-1} . If this holds, then multivariate approximation is *tractable* in the randomized setting for the class Λ^{std} . If $n^{\text{ran}}(\varepsilon, d)$ can be bounded by a polynomial only in ε^{-1} independent of d , then multivariate approximation is *strongly* tractable in the randomized setting for the class Λ^{std} .

Based on Theorem 1, we prove in Theorem 4 that strong tractability and tractability in the randomized setting for the class Λ^{std} are respectively equivalent to strong tractability and tractability in the worst case setting for the class Λ^{all} . Furthermore, the estimates of $n^{\text{ran}}(\varepsilon, d)$ are practically the same as its counterparts in the worst case setting. This is important since the conditions on strong tractability and tractability in the worst case setting for the class Λ^{all} are known for many spaces and, in general, are easy to obtain. We illustrate this point for weighted Korobov spaces of periodic functions with general weights which moderate how functions depend on the groups of variables. We provide in Theorem 5 necessary and sufficient conditions on strong tractability and tractability as well as bounds on the minimal number of information evaluations in the worst case setting for the class Λ^{all} and in the randomized setting for the class Λ^{std} in terms of general

weights of this space. In particular, we show that for product weights independent of d the notions of strong tractability and tractability coincide.

We end this introduction by a note on future research. We believe that the proof technique presented in Theorem 1 can also be applied for multivariate approximation in the worst case and average case settings, and will permit improving the existing results on the power of standard information. In particular, in the worst case setting for reproducing kernel Hilbert spaces, we would like to verify whether we always have $p^{\text{all-wor}} = p^{\text{std-wor}}$. In the average case setting, we believe that the assumption used in [8] on the uniformly bounded eigenfunctions in the L_∞ -norm can be eliminated and we always have $p^{\text{all-avg}} = p^{\text{std-avg}}$. This will be reported in a forthcoming paper.

2. MULTIVARIATE APPROXIMATION

We begin with basic definitions and facts concerning multivariate approximation. Let H be a separable Hilbert space of real functions f defined on a domain $D \subseteq \mathbb{R}^d$. The inner product and the norm in H are denoted by $\langle \cdot, \cdot \rangle_H$ and $\|\cdot\|_H$, respectively. Since we approximate f based on its function values, we need to assume that $f(\mathbf{x})$ is a continuous functional for any $\mathbf{x} \in D$. This is equivalent to assuming that H is a reproducing kernel Hilbert space whose kernel is denoted by K . For general properties of reproducing kernels, we refer the reader to, e.g., [1, 24]. Here we only recall that

$$K(\cdot, \mathbf{x}) \in H \quad \text{and} \quad f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_H \quad \forall f \in H, \mathbf{x} \in D.$$

In particular, we have

$$|f(\mathbf{x})| \leq \|f\|_H \cdot \|K(\cdot, \mathbf{x})\|_H \quad \text{and} \quad \|K(\cdot, \mathbf{x})\|_H = \sqrt{K(\mathbf{x}, \mathbf{x})}.$$

We are interested in approximating functions f with errors measured in a weighted L_2 -norm. That is, let ρ be a given probability density function,

$$\rho : D \rightarrow \mathbb{R}_+ \quad \text{and} \quad \int_D \rho(\mathbf{x}) \, d\mathbf{x} = 1.$$

Without loss of generality, we can assume that $\rho(\mathbf{x}) > 0$ for almost all \mathbf{x} . Then

$$G = G_\rho := \left\{ g : D \rightarrow \mathbb{R} : \int_D |g(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x} < \infty \right\}$$

is a well defined Hilbert space whose inner-product and norm are given by

$$\langle f, g \rangle_G := \int_D f(\mathbf{x}) g(\mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x} \quad \text{and} \quad \|g\|_G^2 := \int_D |g(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x}.$$

Multivariate approximation considered in this paper is defined as the problem of approximating f with the error measured in $\|\cdot\|_G$. Without loss of generality, we can restrict our attention to linear algorithms of the form

$$(2) \quad A(f) = \sum_{i=1}^n L_i(f) g_i$$

where n is a non-negative integer, the functions $g_i = g_{i,n}$ belong to G , and $L_i = L_{i,n}$ are continuous linear functionals. For $n = 0$ we set $A = 0$. For the settings of this paper we can assume that the number n of functional evaluations is fixed, i.e.,

is deterministic and does not depend on f ; see also Remark 1. It is called the *cardinality* of A and is denoted by

$$n = \text{card}(A).$$

We consider two classes of L_i . The first class is $\Lambda = \Lambda^{\text{all}}$ which consists of all continuous linear functionals. That is, $L_i \in \Lambda^{\text{all}}$ iff there is an element of $g_i \in H$ such that $L_i(f) = \langle f, g_i \rangle_H$ for all $f \in H$. The second class is the class of *standard information* $\Lambda = \Lambda^{\text{std}}$ which consists of function evaluations. That is, $L_i \in \Lambda^{\text{std}}$ iff there is a point $t_i \in D$ such that $L_i(f) = f(t_i)$ for all $f \in H$.

2.1. Deterministic Worst Case Setting. We now briefly recall what is known for multivariate approximation in the deterministic worst case setting with $\Lambda = \Lambda^{\text{all}}$. In this setting, the functions g_i 's from G and L_i 's from Λ^{all} in (2) are fixed, and the (*deterministic*) *worst case error* of an algorithm A is defined by

$$e^{\text{wor}}(A; H, \rho) := \sup_{\|f\|_H \leq 1} \|f - A(f)\|_G.$$

For a fixed value of n , let $e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}})$ denote the n th *minimal error*, i.e.,

$$\begin{aligned} & e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}}) \\ & := \inf \{ e^{\text{wor}}(A; H, \rho) : A \text{ given by (2) with } L_i \in \Lambda^{\text{all}} \text{ and } \text{card}(A) \leq n \}. \end{aligned}$$

In another words, $e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}})$ is the smallest worst case error among all algorithms that use at most n continuous linear functionals. An algorithm that achieves this smallest error is said to be an n th *optimal algorithm*.

We now recall well-known results (see e.g., [23, Thm. 5.3]), that characterize n th optimal algorithms via eigenpairs of the following operator W ,

$$(3) \quad W : H \rightarrow H \quad \text{and} \quad Wg := \int_D g(\mathbf{t}) K(\cdot, \mathbf{t}) \rho(\mathbf{t}) \, d\mathbf{t}.$$

One of the results states that

$$\lim_{n \rightarrow \infty} e^{\text{wor}}(n, H, \rho, \Lambda^{\text{all}}) = 0 \quad \text{iff} \quad W \text{ is compact.}$$

This is why we assume throughout this paper the compactness of W . The operator W is also self-adjoint and non-negative definite. Moreover,

$$(4) \quad \langle f, g \rangle_H = \langle f, Wg \rangle_G \quad \forall f, g \in H.$$

This implies, in particular, that H is continuously embedded in G , and

$$\|f\|_H \leq \|W^{1/2}\| \|f\|_G \quad \text{with} \quad e^{\text{wor}}(0; H, \rho, \Lambda) = \|W^{1/2}\|.$$

The operator W has eigenpairs (λ_k, η_k) ,

$$(5) \quad W\eta_k = \lambda_k \eta_k \quad k = 1, 2, \dots,$$

with orthonormalized η_k 's and ordered λ_k , i.e.,

$$\langle \eta_j, \eta_k \rangle_H = \delta_{j,k} \quad \text{and} \quad \lambda_k \geq \lambda_{k+1} \quad \forall k \quad \text{with} \quad \lim_{k \rightarrow \infty} \lambda_k = 0,$$

where $\delta_{j,k}$ denotes the Kronecker delta. We can also assume that all the eigenvalues are positive since otherwise the problem would be trivial. Another result states that

$$(6) \quad A_n^*(f) := \sum_{k=1}^n \langle f, \eta_k \rangle_H \eta_k$$

is an n th optimal algorithm and

$$(7) \quad e^{\text{wor}}(A_n^*; H, \rho) = e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}}) = \sqrt{\lambda_{n+1}}.$$

In particular, for $n = 0$ this means that $\|W^{1/2}\| = \sqrt{\lambda_1}$.

Observe that for any $f \in H$ and $k \geq 1$, we have

$$(8) \quad \lambda_k \langle f, \eta_k \rangle_H = \langle f, W\eta_k \rangle_H = \int_D f(\mathbf{x})\eta_k(\mathbf{x})\rho(\mathbf{x}) \, d\mathbf{x} = \langle f, \eta_k \rangle_G.$$

This implies that

$$(9) \quad \langle \eta_j, \eta_k \rangle_G = \lambda_k \delta_{j,k}.$$

Therefore, if we set

$$(10) \quad \overline{\eta}_k = \lambda_k^{-1/2} \eta_k,$$

then the functions $\overline{\eta}_k$ are orthonormal in G ,

$$(11) \quad \langle \overline{\eta}_j, \overline{\eta}_k \rangle_G = \delta_{j,k}.$$

This is why A_n^* can be rewritten as

$$A_n^*(f) = \sum_{k=1}^n \langle f, \overline{\eta}_k \rangle_G \overline{\eta}_k.$$

2.2. Randomized Setting. In the randomized setting, we consider algorithms A of the form (2) for which both the g_i 's and L_i 's could be randomly selected according to some probability measures. We explain in Remark 1 that randomization does not help for the class Λ^{all} . That is why we restrict our attention to the class Λ^{std} and consider the following class of simplified random algorithms that use function evaluations at randomly chosen points. They are of the form

$$(12) \quad A_{\vec{t}}(f) = \sum_{i=1}^n f(\mathbf{t}_i) g_{i,\vec{t}},$$

where $\vec{t} = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ for random points $\mathbf{t}_1, \dots, \mathbf{t}_n$ from D , which are independent, and each \mathbf{t}_i is distributed according to some probability whose density function is denoted by ω_i . The functions $g_{i,\vec{t}}$ may depend on the selected points \mathbf{t}_j 's. For any f , we can view $A_{\cdot}(f)$ as a random process, and $A_{\vec{t}}(f)$ is its specific realization.

The *randomized error* of A is defined as

$$e^{\text{ran}}(A; H, \rho) := \sup_{\|f\|_H \leq 1} \left(\int_D \cdots \int_D \|f - A_{\vec{t}}(f)\|_G^2 \omega_1(\mathbf{t}_1) \cdots \omega_n(\mathbf{t}_n) \, d\mathbf{t}_1 \dots d\mathbf{t}_n \right)^{1/2}.$$

To simplify the notation, we will write $\mathbb{E}_{\vec{t}}\|f - A_{\vec{t}}(f)\|_G^2$ to denote

$$\mathbb{E}_{\vec{t}}\|f - A_{\vec{t}}(f)\|_G^2 = \int_D \cdots \int_D \|f - A_{\vec{t}}(f)\|_G^2 \omega_1(\mathbf{t}_1) \cdots \omega_n(\mathbf{t}_n) \, d\mathbf{t}_1 \dots d\mathbf{t}_n.$$

It is easy to see that for algorithms (12) we have

$$\left(\mathbb{E}_{\vec{t}}\|f - A_{\vec{t}}(f)\|_G^2\right)^{1/2} \leq \|f\|_H e^{\text{ran}}(A; H, \rho) \quad \forall f \in H.$$

Similarly to the worst case setting, by $e^{\text{ran}}(n; H, \rho, \Lambda^{\text{std}})$ we will denote the smallest randomized error among all such algorithms A with cardinality at most n ,

$$e^{\text{ran}}(n; H, \rho, \Lambda^{\text{std}}) := \inf \{e^{\text{ran}}(A; H, \rho) : A \text{ is given by (12) with } \text{card}(A) \leq n\}.$$

Here, the infimum is taken with respect to all possible choices of distributions of sample points t_i as well as all possible choices of functions $g_{i,t}$.

There are a number of results showing that, essentially, Λ^{all} is equally powerful in the deterministic worst case and the randomized case settings for multivariate approximation defined over Hilbert spaces; see, e.g., [6, 11, 13, 15, 21, 26]. For instance, we know from [15] (see also [26]), that

$$e^{\text{ran}}(n; H, \rho, \Lambda^{\text{all}}) \leq e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}}) \leq \sqrt{2} e^{\text{ran}}(\lfloor (n+1)/2 \rfloor; H, \rho, \Lambda^{\text{all}}).$$

One of the main contributions of the current paper is to relate the power of Λ^{all} with Λ^{std} . Indeed, in Section 3, we provide conditions under which $e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}})$ and $e^{\text{ran}}(n; H, \rho, \Lambda^{\text{std}})$ have similar rates of convergence to zero as n tends to infinity. This and the following remark are the reasons why, without loss of generality, we can restrict our attention to simplified randomized algorithms of the form (12).

Remark 1. In the randomized setting, one may consider a very general class of randomized algorithms of the form

$$A_\omega(f) = \phi_\omega(L_{1,\omega}(f), L_{2,\omega}(f), \dots, L_{n_\omega,\omega}(f))$$

with a random element ω distributed according to some probability measure. Here, $L_{i,\omega}$ can be from the class Λ^{std} or Λ^{all} , the random mapping ϕ_ω can be linear or non-linear, and we may use adaption; see e.g., [4, 6, 7, 12, 13, 15, 18, 21, 26]. Moreover, the number $n_\omega = n_\omega(f)$ of functional evaluations may be random and may depend adaptively on f ; in such a case $\text{card}(A)$ is defined by $\sup_{\|f\|_H \leq 1} \mathbb{E}_\omega n_\omega(f)$.

Let $e^{\text{gen-ran}}(n; H, \rho, \Lambda^{\text{all}})$ denote the n th minimal error among all such generalized randomized algorithms with $L_{i,\omega} \in \Lambda^{\text{all}}$. Clearly,

$$e^{\text{gen-ran}}(n; H, \rho, \Lambda^{\text{all}}) \leq e^{\text{ran}}(n; H, \rho, \Lambda^{\text{all}}) \leq e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}}).$$

The standard proof technique to obtain a lower bound on $e^{\text{gen-ran}}(n; H, \rho, \Lambda^{\text{all}})$ is the reduction to the average case with some probability measure on the unit ball of H which was first applied by Bakhvalov in 1959 (see [2]) and used in many papers; see e.g., [6, 7, 11, 13, 15, 18, 21, 26]. By choosing an appropriate probability measure on the unit ball of H one obtains an estimate in terms of the worst case setting; see e.g., [15, 26] for multivariate approximation. More precisely, from [14] we know that varying cardinality of n_ω can only help by a fixed factor, and using again the results from [15] we conclude that even such generalized randomization does not help for Λ^{all} since

$$e^{\text{wor}}(n; H, \rho, \Lambda^{\text{all}}) \leq 2\sqrt{2} e^{\text{gen-ran}}(\lfloor (n+1)/4 \rfloor; H, \rho, \Lambda^{\text{all}}).$$

In particular, this means that under the conditions of the next section, the simplified randomized algorithms (12) of this paper are almost as powerful as very general (perhaps complicated and expensive to implement) randomized algorithms.

3. MAIN RESULT

Throughout this section we assume that the eigenpairs of the compact operator W satisfy

$$(13) \quad \sqrt{\lambda_n} \leq C_0 \frac{C(n)}{n^p} \quad \forall n = 1, 2, \dots$$

for positive numbers p and C_0 , and a monotonically non-decreasing function $C(n)$ for which $C(1) = 1$. Let

$$(14) \quad p_k := p \frac{2p_{k-1} + 1}{2p + 1} \quad \text{with} \quad p_0 := 0 \quad \text{and} \quad m_k := \left\lfloor n^{(2p_{k-1} + 1)/(2p + 1)} \right\rfloor.$$

It is easy to check that

$$(15) \quad p_k = p \left(1 - \left(\frac{2p}{2p + 1} \right)^k \right).$$

Hence, p_k 's are increasing and $\lim_k p_k = p$. Moreover, the sequence $\{m_k\}_k$ is non-decreasing and $m_k \leq n$. Let

$$(16) \quad \omega_{m_k}(\mathbf{t}) := \rho(\mathbf{t}) \cdot u_{m_k}(\mathbf{t}) \quad \text{with} \quad u_{m_k}(\mathbf{t}) := m_k^{-1} \sum_{j=1}^{m_k} [\bar{\eta}_j(\mathbf{t})]^2,$$

where the $\bar{\eta}_j$'s are given by (10). Clearly, due to orthonormality of the eigenfunctions $\bar{\eta}_j$, the function ω_{m_k} is a probability density function.

For given $n \geq 1$, we consider the following family of randomized algorithms $\{A_{n,k}\}_{k=1}^\infty$. The algorithm $A_{n,k}$ will have cardinality at most nk , and will sample the functions at points whose collection is denoted by $\vec{\mathbf{t}}_{n,k}$. To stress the role of these points we write $A_{n,k} = A_{n,k,\vec{\mathbf{t}}_{n,k}}$.

For $k = 1$, we define

$$(17) \quad A_{n,1,\vec{\mathbf{t}}_{n,1}}(f) := \sum_{j=1}^{m_1} \left[\frac{1}{n} \sum_{\ell=1}^n \frac{f(\boldsymbol{\tau}_\ell) \bar{\eta}_j(\boldsymbol{\tau}_\ell)}{u_{m_1}(\boldsymbol{\tau}_\ell)} \right] \bar{\eta}_j$$

with $\vec{\mathbf{t}}_{n,1} = [\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_n]$ for independent and identically distributed (i.i.d. for short) points $\boldsymbol{\tau}_\ell$'s that are distributed according to the probability density function ω_{m_1} .

For $k \geq 2$, we define

$$(18) \quad \begin{aligned} A_{n,k,\vec{\mathbf{t}}_{n,k}}(f) &:= A_{n,k-1,\vec{\mathbf{t}}_{n,k-1}}(f) \\ &+ \sum_{j=1}^{m_k} \left[\frac{1}{n} \sum_{\ell=1}^n (f - A_{n,k-1,\vec{\mathbf{t}}_{n,k-1}}(f))(\boldsymbol{\tau}_\ell) \frac{\bar{\eta}_j(\boldsymbol{\tau}_\ell)}{u_{m_k}(\boldsymbol{\tau}_\ell)} \right] \bar{\eta}_j \end{aligned}$$

with

$$\vec{\mathbf{t}}_{n,k} = [\vec{\mathbf{t}}_{n,k-1}, \vec{\boldsymbol{\tau}}_n], \quad \text{where} \quad \vec{\boldsymbol{\tau}}_n = [\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_n]$$

for i.i.d. $\boldsymbol{\tau}_\ell$'s that are distributed according to the probability density function ω_{m_k} . Of course, the whole vector $\vec{\mathbf{t}}_{n,k}$ consists of components distributed according to the probability density functions $\omega_{m_1}, \dots, \omega_{m_k}$, respectively. Clearly,

$$\text{card}(A_{n,k,\vec{\mathbf{t}}_{n,k}}) \leq kn \quad \forall \vec{\mathbf{t}}_{n,k} \in D^{nk}.$$

Theorem 1. *Let (13) hold. Then for every n and k ,*

$$(19) \quad e^{\text{ran}}(A_{n,k}; H, \rho) \leq \frac{C_0}{n^{p_k}} \sqrt{k[C(n+1)]^2 + 1}.$$

Proof. The proof is by induction with respect to k . Since $A_{n,k,\vec{\mathbf{t}}_{n,k}}$ is a linear combination of $\bar{\eta}_j$ for $j = 1, 2, \dots, m_k$, we have for every n and k ,

$$(20) \quad \left\langle A_{n,k,\vec{\mathbf{t}}_{n,k}}(f), \bar{\eta}_j \right\rangle_G = 0 \quad \forall \vec{\mathbf{t}}_{n,k} \in D^{kn}, \quad \forall f \in H, \quad \forall j \geq m_k + 1.$$

We begin with $k = 1$. Consider now functions f with $\|f\|_H \leq 1$. Due to (8), (10), (11), and the fact that the eigenfunctions η_j form an orthonormal system in H , we have

$$\begin{aligned} \|f - A_{n,1,\vec{\tau}}(f)\|_G^2 &= E_1(f; \vec{t}) + \sum_{j=m_1+1}^{\infty} \langle f, \eta_j \rangle_H^2 \|\eta_j\|_G^2 \\ &= E_1(f; \vec{\tau}) + \sum_{j=m_1+1}^{\infty} \lambda_j \langle f, \eta_j \rangle_H^2 \leq E_1(f; \vec{\tau}) + \lambda_{m_1+1}, \end{aligned}$$

where

$$E_1(f; \vec{t}) := \sum_{j=1}^{m_1} \left| \langle f, \bar{\eta}_j \rangle_G - n^{-1} \sum_{\ell=1}^n f(\tau_\ell) \bar{\eta}_j(\tau_\ell) / u_{m_1}(\tau_\ell) \right|^2.$$

It is well known, and can be easily verified, that then

$$\begin{aligned} \mathbb{E}_{\vec{\tau}}(E_1(f; \vec{\tau})) &= \int_D \cdots \int_D E_1(f; \vec{\tau}) \omega_{m_1}(\tau_1) \cdots \omega_{m_1}(\tau_n) \, d\tau_1 \cdots d\tau_n \\ &\leq \frac{1}{n} \sum_{j=1}^{m_1} \int_D \frac{|f(\mathbf{x}) \bar{\eta}_j(\mathbf{x})|^2 \rho(\mathbf{x})}{u_{m_1}(\mathbf{x})} \, d\mathbf{x} = \frac{m_1}{n} \int_D f^2(\mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x} \\ &\leq \frac{m_1}{n} \lambda_1 \leq \frac{m_1}{n} C_0^2. \end{aligned}$$

Since $m_1 \leq n^{1/(2p+1)} \leq m_1 + 1 \leq n + 1$ and $C(n)$ is monotonic, we obtain

$$e^{\text{ran}(A_{n,1}; H, \rho)^2} \leq \lambda_{m_1+1} + \frac{m_1}{n} C_0^2 \leq \frac{C_0^2}{n^{2p/(2p+1)}} [(C(n+1))^2 + 1].$$

This completes the proof for $k = 1$.

Consider now $k \geq 2$ and the algorithm $A_{n,k}$. For functions f with $\|f\|_H \leq 1$ we have, as for the case of $k = 1$,

$$(21) \quad \|f - A_{n,k,\vec{t}_{n,k}}(f)\|_G^2 \leq \lambda_{m_k+1} + E_k(f; \vec{t}_{n,k})$$

with

$$\begin{aligned} E_k(f; \vec{t}_{n,k}) &:= \sum_{j=1}^{m_k} \left| \left\langle f - A_{n,k-1,\vec{t}_{n,k-1}}(f), \bar{\eta}_j \right\rangle_G \right. \\ &\quad \left. - \frac{1}{n} \sum_{\ell=1}^n [f - A_{n,k-1,\vec{t}_{n,k-1}}(f)](\tau_\ell) \frac{\bar{\eta}_j(\tau_\ell)}{u_{m_k}(\tau_\ell)} \right|^2. \end{aligned}$$

Now taking the expectation of $E_k(f; \vec{t}_{n,k})$ with respect to $\vec{t}_{n,k}$ and using the independence of $\vec{t}_{n,k-1}$ and $\vec{\tau}_n$ we obtain

$$\begin{aligned} \mathbb{E}_{\vec{t}_{n,k}}(E_k(f; \vec{t}_{n,k})) &= \mathbb{E}_{\vec{t}_{n,k-1}} \mathbb{E}_{\vec{\tau}_n}(E_k(f; [\vec{t}_{n,k-1}, \vec{\tau}_n])) \\ &\leq \mathbb{E}_{\vec{t}_{n,k-1}} \left(\frac{1}{n} \sum_{j=1}^{m_k} \int_D |(f - A_{n,k-1,\vec{t}_{n,k-1}}(f))(\mathbf{x})|^2 \frac{(\bar{\eta}_j(\mathbf{x}))^2}{u_{m_k}(\mathbf{x})} \rho(\mathbf{x}) \, d\mathbf{x} \right) \\ &= \frac{m_k}{n} \cdot \mathbb{E}_{\vec{t}_{n,k-1}} \|f - A_{n,k-1,\vec{t}_{n,k-1}}(f)\|_G^2 \leq \frac{m_k}{n} e^{\text{ran}(A_{n,k-1}; H, \rho)^2} \\ &\leq m_k n^{-2p_{k-1}-1} C_0^2 ((k-1) [C(n+1)]^2 + 1). \end{aligned}$$

This and (21) imply that

$$\begin{aligned} e^{\text{ran}}(A_{n,k}; H, \rho)^2 &\leq C_0^2 \frac{[C(m_k + 1)]^2}{(m_k + 1)^{2p}} + \frac{m_k C_0^2 ((k - 1) [C(n + 1)]^2 + 1)}{n^{2p_{k-1} + 1}} \\ &\leq \frac{C_0^2}{n^{2p_k}} (k [C(n + 1)]^2 + 1), \end{aligned}$$

with the last inequality due to monotonicity of $C(n)$ and the fact that

$$m_k \leq n^{(2p_{k-1} + 1)/(2p + 1)} \leq m_k + 1 \leq n + 1.$$

This completes the proof. □

We already remarked that the exponents p_k converge to p . In fact, due to (15), we have an exponentially fast convergence

$$p - p_k = \frac{2p}{2p + 1} (p - p_{k-1}) = \left(\frac{2p}{2p + 1} \right)^k p.$$

Hence, it is enough to perform a few steps to get p_k very close to p . More precisely, we have the following proposition.

Proposition 1. *For*

$$k = \left\lceil \frac{\ln(1/\delta)}{\ln(1 + 1/(2p))} \right\rceil$$

we have

$$p_k \geq p(1 - \delta).$$

Theorem 1 states that we can achieve nearly the same speed of convergence in the randomized setting in the class Λ^{std} as in the worst case setting in the class Λ^{all} . Due to Remark 1, it is the best possible result. Furthermore, the factors multiplying n^{-p_k} in the randomized case are roughly only \sqrt{k} larger than in the worst case. Again, since k is relatively small, this extra factor is not very important.

Assume now that $C(n) \equiv 1$ in (13) and take

$$k = k^* = \left\lceil \frac{\ln(\ln(n))}{\ln(1 + 1/(2p))} \right\rceil$$

for n such that $\ln(\ln(n)) > 1$. Then it is easy to show that

$$e^{\text{ran}}(A_{n,k^*}; H, \rho) \leq e C_0 n^{-p} \sqrt{2 + \frac{\ln(\ln(n))}{\ln(1 + 1/(2p))}} = O\left(n^{-p} \sqrt{\ln(\ln(n))}\right).$$

This yields the following proposition.

Proposition 2. *Let $m = k^* n = \Theta(n \ln(\ln(n)))$. If $C(n) \equiv 1$ in (13), then the algorithm $\bar{A}_m = A_{n,k^*}$ uses at most m function values and*

$$e^{\text{ran}}(\bar{A}_m; H, \rho) = O\left(\frac{(\ln(\ln(m)))^{p+0.5}}{m^p}\right)$$

with the factor in the big O notation independent of d and m and depending only on C_0 and p . Hence, modulo a power of $\ln(\ln(m))$, we obtain the same speed of convergence as for the optimal algorithm A_m^ .*

We now relate the speed of convergence in the worst case setting for the class Λ^{all} with the speed of convergence in the randomized setting for the class Λ^{std} using the concept of the optimal rate of convergence.

The optimal rate of convergence in the worst case setting for a class Λ is defined as

$$p^{\text{wor}}(H, \rho, \Lambda) := \sup \left\{ r > 0 : \lim_{n \rightarrow \infty} e^{\text{wor}}(n; H, \rho, \Lambda) n^r = 0 \right\}$$

with the convention that the supremum of the empty set is zero. When $\Lambda = \Lambda^{\text{all}}$, then we have, due to (7),

$$p^{\text{wor}}(H, \rho, \Lambda^{\text{all}}) = \sup \left\{ r > 0 : \lim_{n \rightarrow \infty} \sqrt{\lambda_{n+1}} n^r = 0 \right\}.$$

The optimal rate of convergence in the randomized setting for a class Λ is defined analogously as

$$p^{\text{ran}}(H, \rho, \Lambda) := \sup \left\{ r > 0 : \lim_{n \rightarrow \infty} e^{\text{ran}}(n; H, \rho, \Lambda) n^r = 0 \right\}.$$

Then Theorem 1 and Remark 1 yield the following corollary.

Corollary 1.

$$p^{\text{ran}}(H, \rho, \Lambda^{\text{std}}) = p^{\text{ran}}(H, \rho, \Lambda^{\text{all}}) = p^{\text{gen-ran}}(H, \rho, \Lambda^{\text{all}}) = p^{\text{wor}}(H, \rho, \Lambda^{\text{all}}).$$

The algorithms $A_{n,k}$ achieve nearly the optimal speed of convergence. They use randomly generated points $\mathbf{t}_1, \dots, \mathbf{t}_{nk}$ with different distributions. Indeed, ω_{m_i} is the probability density function of \mathbf{t}_ℓ with $i = \lceil \ell/k \rceil$. Note that ω_{m_i} given by (16) depends on the eigenfunctions $\overline{\eta}_j$ in a way that might be sometimes too difficult to implement. Therefore, it is natural to ask if a similar result could be obtained with the same distribution for all the points and whether this distribution could be simplified and be independent of the eigenfunctions $\overline{\eta}_j$. It turns out that this can be achieved under an additional assumption on the eigenfunctions $\overline{\eta}_j$ which was used in a different context in [8], where multivariate approximation has been studied in the average case setting.

We consider therefore the following algorithms $A'_{n,k}$ that are modifications of the algorithms $A_{n,k}$. They are given by (17) and (18), however, with the i.i.d. points τ_ℓ distributed according to the density function ρ , i.e., $\omega_i = \rho$ for every i .

Theorem 2. *Suppose that the eigenpairs of the operator W satisfy (13) and*

$$(22) \quad \sup_n \|\overline{\eta}_n\|_\infty \leq C_2$$

for a positive constant C_2 . Then the algorithms $A'_{n,k}$ satisfy

$$(23) \quad e^{\text{ran}}(A'_{n,k}; H, \rho) \leq \frac{C_0 C_2}{n^{pk}} \sqrt{k [C(n+1)]^2 + 1}.$$

Proof. The proof of this theorem is very similar to the proof of Theorem 1 as well as to the proof in [8], and hence is omitted. □

We illustrate the above results by applying them to the space H of d -variate functions with mixed partial derivatives of order one bounded in the L_2 -norm. This space has been considered in many papers dealing with the integration problem as well as the approximation in the worst case and average case settings.

Example 1. Consider $D = [0, 1]^d$ and $\rho \equiv 1$. The space H is the reproducing kernel Hilbert space of functions $f : [0, 1]^d \rightarrow \mathbb{R}$ with the kernel

$$K(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^d (1 + \min(x_j, y_j)).$$

It follows from [17, Thm. 2.1] that the eigenvalues of the operator W satisfy

$$\lambda_n = \frac{1}{(d-1)! \pi^{4d}} \left(\frac{\ln^{d-1}(n+e)}{n} \right)^2 (1 + o(1)) \quad \text{as } n \rightarrow \infty,$$

i.e., the assumption (13) holds with $p = 1$, $C_0 = 1/(\pi^{2d} \sqrt{(d-1)!})$, and $C(n) = \ln^{d-1}(n+e)(1 + o(1))$ as $n \rightarrow \infty$. Then Theorem 1 yields

$$e^{\text{ran}}(A_{n,k}) \leq \frac{\sqrt{k}}{\pi^{2d} \sqrt{(d-1)!}} \frac{\ln^{d-1}(n+e)}{n^{pk}} (1 + o(1)) \quad \text{as } n \rightarrow \infty.$$

In particular, for $k = 4$, we have $p_4 \geq 0.8$ and

$$e^{\text{ran}}(A_{n,4}) \leq \frac{4}{\pi^{2d} \sqrt{(d-1)!}} \frac{\ln^{d-1}(n+e)}{n^{0.8}} (1 + o(1)) \quad \text{as } n \rightarrow \infty.$$

Theorem 2 yields a weaker result:

$$e^{\text{ran}}(A'_{n,k}) \leq \frac{\sqrt{k} 2^{d/2}}{\pi^{2d} \sqrt{(d-1)!}} \frac{\ln^{d-1}(n+e)}{n^{pk}} (1 + o(1)) \quad \text{as } n \rightarrow \infty$$

since the L_∞ -norms of the eigenfunctions $\bar{\eta}_i$ are equal to $2^{d/2}(1 + o(1))$ as n tends to infinity; see again [17].

The assumption (22) may be difficult to verify. We now show that this assumption can be replaced by another one that is easy to verify; however, at the possible expense of the speed of convergence. We now assume that

$$(24) \quad C_3 := \text{ess sup}_{\mathbf{x} \in D} \sqrt{K(\mathbf{x}, \mathbf{x})} < \infty$$

instead of (22). As shown in the following example, (24) does not imply (22).

Example 2. Let $D = (0, 1]$, $\rho \equiv 1$, $\eta_n = \mathbf{1}_{(1/(n+1), 1/n]}$, and $\lambda_n = 1/(n(n+1))$. Clearly $K(x, y) = \sum_{n=1}^\infty \eta_n(x) \eta_n(y)$ is a well-defined kernel and $\bar{\eta}_n = \eta_n \sqrt{n(n+1)}$ are orthonormal in L_2 . Moreover, $\bar{\eta}_n$ are the eigenvectors of the operator W , $\|\bar{\eta}_n\|_\infty = \sqrt{n(n+1)}$, yet $K(x, x) \equiv 1$, i.e., $C_3 = 1$.

The assumption (24) implies that

$$\|f\|_\infty = \text{ess sup}_{\mathbf{x} \in D} |\langle f, K(\mathbf{x}, \cdot) \rangle_H| \leq \|f\|_H \text{ess sup}_{\mathbf{x} \in D} \sqrt{K(\mathbf{x}, \mathbf{x})} \leq C_3 \|f\|_H \quad \forall f \in H.$$

Furthermore, the eigenvalues of W are summable. Indeed,

$$K(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^\infty \eta_j(\mathbf{x}) \eta_j(\mathbf{y})$$

and (9) yield

$$\sum_{j=1}^\infty \lambda_j = \sum_{j=1}^\infty \int_D \eta_j(\mathbf{x})^2 \rho(\mathbf{x}) \, d\mathbf{x} = \int_D K(\mathbf{x}, \mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x} \leq C_3^2.$$

This implies that $n\lambda_n \leq \sum_{j=1}^\infty \lambda_j \leq C_3^2$ and

$$\sqrt{\lambda_n} \leq C_3 n^{-1/2}.$$

Hence, (13) holds with $p \geq 1/2$, and $C_0 \leq C_3$.

Consider now a family of algorithms $A''_{n,k}$ that differ from $A'_{n,k}$ only by the definition of the values m_k . We now define m_k , for $k = 1, 2, \dots$, such that

$$\lambda_{m_k+1} \leq \frac{C_3 C_4(k-1)}{n^{1-2^{-k}}} \leq \lambda_{m_k},$$

where

$$C_4(k) := 2C_3 \left(\frac{C_0}{2C_3} \right)^{2^{-k}} \leq 2C_3.$$

Theorem 3. *Let (24) hold. Then the algorithms $A''_{n,k}$ satisfy*

$$e^{\text{ran}}(A''_{n,k}; H, \rho) \leq C_4(k) n^{-1/2+2^{-k-1}}.$$

Proof. As in the proof of the previous theorems, consider first $k = 1$ and the corresponding term E_1 . Its expectation with respect to the points τ_ℓ for $\|f\|_H \leq 1$ is bounded by

$$\mathbb{E}_{\vec{\tau}}(E_1(f; \vec{\tau})) \leq \frac{1}{n} \sum_{j=1}^{m_1} \int_D [f(\mathbf{x}) \bar{\eta}_j(\mathbf{x})]^2 \rho(\mathbf{x}) \, d\mathbf{x} = \frac{1}{n} \int_D f^2(\mathbf{x}) \rho(\mathbf{x}) \sum_{j=1}^{m_1} \bar{\eta}_j(\mathbf{x})^2 \, d\mathbf{x}.$$

Note that

$$\sum_{j=1}^{m_1} \bar{\eta}_j(\mathbf{x})^2 = \sum_{j=1}^{m_1} \lambda_j^{-1} \eta_j(\mathbf{x})^2 \leq \lambda_{m_1}^{-1} \sum_{j=1}^{m_1} \eta_j(\mathbf{x})^2 \leq \lambda_{m_1}^{-1} K(\mathbf{x}, \mathbf{x}) \leq \lambda_{m_1}^{-1} C_3^2.$$

Hence

$$\mathbb{E}_{\vec{\tau}}(E_1(f; \vec{\tau})) \leq \frac{C_3^2}{n \lambda_{m_1}} \int_D f^2(\mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x} = \frac{C_0^2 C_3^2}{n \lambda_{m_1}}.$$

This means that

$$e^{\text{ran}}(A''_{n,1}; H, \rho)^2 \leq \lambda_{m_1+1} + \frac{C_0^2 C_3^2}{n \lambda_{m_1}} \leq \frac{2C_0 C_3}{\sqrt{n}}$$

due to the definition of m_1 . Hence we have that

$$e^{\text{ran}}(A''_{n,1}; H, \rho) \leq C_4(1) n^{-s_1}$$

since $C_4(1) = \sqrt{2C_0 C_3}$ and with $s_1 = 1/4$, as claimed.

Suppose by induction that

$$e^{\text{ran}}(A''_{n,k}; H, \rho) \leq C_4(k) n^{-s_k} \quad \text{with} \quad s_k = \frac{1}{2} - 2^{-k-1}.$$

For the algorithm $A''_{n,k+1}$ we have

$$\begin{aligned} & \mathbb{E}_{\vec{\tau}_{n,k+1}}(E_{k+1}(f; \vec{\tau}_{n,k+1})) \\ & \leq \mathbb{E}_{\vec{\tau}_{n,k}} \left(\frac{1}{n} \int_D |(f - A''_{n,k, \vec{\tau}_{n,k}}(f))(\mathbf{x})|^2 \rho(\mathbf{x}) \sum_{j=1}^{m_{k+1}} \eta_j^2(\mathbf{x}) / \lambda_j \, d\mathbf{x} \right) \\ & \leq \frac{C_3^2}{\lambda_{m_{k+1}} n} e^{\text{ran}}(A''_{n,k}; H, \rho)^2 \leq \frac{C_3^2 C_4^2(k)}{\lambda_{m_{k+1}} n^{1+2s_k}}. \end{aligned}$$

Due to the definition of m_{k+1} we have

$$e^{\text{ran}}(A''_{n,k+1}; H, \rho) \leq \left(\lambda_{m_{k+1}+1} + \frac{C_3^2 C_4^2(k)}{\lambda_{m_{k+1}} n^{1+2s_k}} \right)^{1/2} \leq C_4(k+1)n^{-s_{k+1}}$$

with

$$C_4(k+1) = \sqrt{2 C_3 C_4(k)}.$$

It is easy to show that

$$C_4(k) = 2 C_3 \left(\frac{C_0}{2 C_3} \right)^{2^{-k}} \leq 2 C_3,$$

as claimed. □

As already pointed out, (24) implies that $\lambda_n = O(n^{-1/2})$. Since the convergence rate of the errors of the algorithms $A''_{n,k}$ approaches $n^{-1/2}$, they are almost optimal only when the eigenvalues λ_n do not converge to zero much faster than $n^{-1/2}$.

4. MULTIVARIATE INTEGRATION

Based on the results of Section 3, we now derive efficient randomized algorithms for multivariate integration. For this problem, we want to approximate

$$I_\rho(f) := \int_D f(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x}$$

by randomized algorithms of the form

$$Q(f; \omega) = \sum_{j=1}^n f(\mathbf{t}_{j,\omega}) a_{j,\omega}$$

for some random element ω with $\mathbf{t}_{j,\omega} \in D$ and $a_{j,\omega} \in \mathbb{R}$. The randomized error of Q is given by

$$e^{\text{ran}}(Q; H, I_\rho) := \sqrt{\mathbb{E}_\omega (I_\rho(f) - Q(f; \omega))^2}.$$

We now use a well-known variance reduction technique for deriving a randomized algorithm $\text{Quad}_{n,k,\vec{\mathbf{t}},\vec{\mathbf{y}}}$ from a randomized algorithm $\text{Alg}_{n,k,\vec{\mathbf{t}}}$ for multivariate approximation:

$$(25) \quad \text{Quad}_{n,k,\vec{\mathbf{t}},\vec{\mathbf{y}}}(f) := I_\rho \left(\text{Alg}_{n,k,\vec{\mathbf{t}}}(f) \right) - \frac{1}{n} \sum_{\ell=1}^n \left[f - \text{Alg}_{n,k,\vec{\mathbf{t}}}(f) \right] (\mathbf{y}_\ell)$$

for i.i.d. points \mathbf{y}_ℓ whose distribution has ρ as its density. It is well known and easy to verify that

$$\mathbb{E}_{\vec{\mathbf{y}}} \left(I_\rho(f) - \text{Quad}_{n,k,\vec{\mathbf{t}},\vec{\mathbf{y}}}(f) \right)^2 \leq \frac{1}{n} \|f - \text{Alg}_{n,k,\vec{\mathbf{t}}}(f)\|_G^2 \quad \forall \vec{\mathbf{t}}.$$

This and the results of the previous section yield the following proposition.

Proposition 3. *Let $Q_{n,k,\vec{\mathbf{t}}}$, $Q'_{n,k,\vec{\mathbf{t}}}$, and $Q''_{n,k,\vec{\mathbf{t}}}$ be the randomized algorithms obtained by applying (25) with Alg equal to $A_{n,k,\vec{\mathbf{t}}}$, $A'_{n,k,\vec{\mathbf{t}}}$, and $A''_{n,k,\vec{\mathbf{t}}}$ from Section 3, respectively. Each of them uses at most $n(k+1)$ function values and, under the corresponding assumptions from Section 3, their randomized errors are bounded from above, respectively, by*

$$\frac{C_0}{n^{1/2+p_k}} \sqrt{k[C(n+1)]^2 + 1}, \quad \frac{C_0 C_2}{n^{1/2+p_k}} \sqrt{k[C(n+1)]^2 + 1}, \quad \text{and} \quad \frac{C_4(k)}{n^{1-2^{-k-1}}}.$$

We stress the increase by 1/2 in the order of convergence for multivariate integration.

5. APPLICATION TO WEIGHTED KOROBOV SPACES

In this section we consider multivariate approximation for the weighted Korobov space of periodic functions defined over $D = [0, 1]^d$; see e.g., [3, 8, 9, 16]. We take $\rho \equiv 1$ and then $G = L_2([0, 1]^d)$. To stress the dependence on the number d of variables, we will write H_d and G_d instead of H and G , and we will drop ρ from all the notation.

Let α be a given number greater than 1, and $\gamma = \{\gamma_{d,u}\}_{d,u}$ a given family of non-negative numbers, called weights. This family is indexed by $d \in \mathbb{N}_+$ and $u \subseteq \{1, 2, \dots, d\}$. The reproducing kernel K_d of the weighted Korobov space H_d is of the form

$$K_d(\mathbf{x}, \mathbf{y}) = \gamma_{d,\emptyset} + \sum_{\emptyset \neq u \subseteq \{1, 2, \dots, d\}} \gamma_{d,u} K_{d,u}(\mathbf{x}, \mathbf{y}),$$

where

$$K_{d,u}(\mathbf{x}, \mathbf{y}) = \prod_{j \in u} \sum_{k \neq 0} \frac{\exp(2\pi i k(x_j - y_j))}{|k|^\alpha} \quad \text{with } i = \sqrt{-1}.$$

It is convenient to write $K_d(\mathbf{x}, \mathbf{y})$ (see [3]) as

$$K_d(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{h} \in \mathbb{Z}^d} \beta(\mathbf{h}, \gamma) e^{2\pi i \mathbf{h} \cdot (\mathbf{x} - \mathbf{y})},$$

where $\mathbb{Z}^d = \{\dots, -1, 0, 1, \dots\}^d$, $\mathbf{h} \cdot \mathbf{x} = \sum_{j=1}^d h_j x_j$ denotes the inner product of the vectors \mathbf{h} and \mathbf{x} , and

$$\beta(\mathbf{h}, \gamma) = \begin{cases} \gamma_{d,\emptyset} & \text{if } \mathbf{h} = (0, 0, \dots, 0), \\ \gamma_{d,u_h} \prod_{j \in u_h} |h_j|^{-\alpha} & \text{if } \mathbf{h} \neq 0 \text{ with } u_h = \{j : h_j \neq 0\}. \end{cases}$$

The weighted Korobov space H_d is equipped with the inner product

$$\langle f, g \rangle_{H_d} = \sum_{\mathbf{h} \in \mathbb{Z}^d} \beta(\mathbf{h}, \gamma)^{-1} \hat{f}(\mathbf{h}) \overline{\hat{g}(\mathbf{h})} \quad \forall f, g \in H_d,$$

where

$$\hat{f}(\mathbf{h}) = \int_{[0,1]^d} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

is a Fourier coefficient of f . Here, if $\beta(\mathbf{h}, \gamma) = 0$, then we assume that $\hat{f}(\mathbf{h}) = 0$ for all $f \in H_d$, and we adopt the convention that $0/0 = 0$.

If α is an even integer, the reproducing kernel K_d is related to the Bernoulli polynomials B_α . For α even, we have for $x \in [0, 1]$,

$$\sum_{h \neq 0} \frac{e^{2\pi i hx}}{|h|^\alpha} = \frac{(-1)^{\alpha/2+1} (2\pi)^\alpha}{\alpha!} B_\alpha(x).$$

In this case the kernel K_d can be written as

$$K_d(\mathbf{x}, \mathbf{y}) = \gamma_{d,\emptyset} + \sum_{\emptyset \neq u \subseteq \{1, 2, \dots, d\}} \gamma_{d,u} \left(\frac{(2\pi)^\alpha (-1)^{\alpha/2+1}}{\alpha!} \right)^{|u|} \prod_{j \in u} B_\alpha(\{x_j - y_j\}),$$

where the notation $\{x\}$ means the fractional part of x , $\{x\} = x - [x]$.

The role of α is to specify how fast Fourier coefficients decay, which is also related to the regularity of the functions from the space H_d . Roughly speaking, functions from H_d have square integrable mixed partial derivatives of order $\leq \alpha/2$. The role of the weights $\gamma_{d,u}$ is to quantify the importance of the interactions of variables from the subset u ; the smaller the weight $\gamma_{d,u}$ the less significant the interaction.

Often in the literature, the so-called product weights are considered, i.e.,

$$\gamma_{d,\emptyset} = 1 \quad \text{and} \quad \gamma_{d,u} = \prod_{j \in u} \gamma_{d,j}$$

for given weights $0 \leq \gamma_{d,d} \leq \gamma_{d,d-1} \leq \dots \leq \gamma_{d,1} \leq 1$. For product weights, the kernel takes the form

$$K_d(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^d \left(1 + \gamma_{d,j} \sum_{k \neq 0} \frac{\exp(2\pi i k(x_j - y_j))}{|k|^{-\alpha}} \right).$$

For general weights $\gamma_{d,u}$, it follows from [8] that the eigenpairs of the operator $W = W_d$ are:

- for $u = \emptyset$,

$$\lambda_{d,\emptyset} = \gamma_{d,\emptyset} \quad \text{and} \quad \bar{\eta}_{d,\emptyset} \equiv 1,$$

- for $u \neq \emptyset$,

$$(\lambda_{d,\mathbf{k}_u}, \bar{\eta}_{d,\mathbf{k}_u,\ell})$$

for $\ell = 1, 2$. Here, $\mathbf{k}_u = [k_1, k_2, \dots, k_{|u|}]$ with all $k_j \neq 0$ for $j \in u$, and

$$\lambda_{d,\mathbf{k}_u} = \gamma_{d,u} \prod_{j \in u} k_j^{-\alpha}$$

is an eigenvalue of multiplicity two, and the corresponding two eigenvectors, orthonormal in G_d , are

$$\bar{\eta}_{d,\mathbf{k}_u,1}(\mathbf{x}) = \sqrt{2} \cos \left(2\pi \sum_{j \in u} k_j x_j \right) \quad \text{and} \quad \bar{\eta}_{d,\mathbf{k}_u,2}(\mathbf{x}) = \sqrt{2} \sin \left(2\pi \sum_{j \in u} k_j x_j \right).$$

For the weighted Korobov space we may apply Theorem 2 with

$$C_2 = \sqrt{2} \quad \forall d.$$

This means that we can now use uniform distribution for all sample points in the algorithm $A'_{n,k}$ whose error bounds are worse only by a factor $\sqrt{2}$ than the error bounds of the algorithm $A_{n,k}$ which uses a more complicated distribution of its sample points.

As in [8], for $r \in [1, \alpha)$, consider

$$(26) \quad M_{d,r} := \sum_{u, \mathbf{k}_u} \lambda_{d,\mathbf{k}_u}^{1/r} = \gamma_{d,\emptyset}^{1/r} + \sum_{\emptyset \neq u} \gamma_{d,u}^{1/r} [2\zeta(\alpha/r)]^{|u|},$$

where ζ is the Riemann zeta function, $\zeta(x) = \sum_{j=1}^{\infty} j^{-x}$ for $x > 1$. For product weights, $M_{d,r}$ reduces to

$$M_{d,r} = \prod_{j=1}^d \left(1 + 2\gamma_{d,j}^{1/r} \zeta(\alpha/r) \right).$$

If $\lambda_{d,1} \geq \lambda_{d,2} \geq \dots$ are the ordered eigenvalues λ_{d,\mathbf{k}_u} , then the n th largest eigenvalue satisfies $\lambda_{d,n}^{1/r} n \leq M_{d,r}$, i.e.,

$$(27) \quad \lambda_{d,n} \leq M_{d,r}^r n^{-r}.$$

Therefore, we can apply Theorem 1 with $p = r/2$, $C_0 = M_{d,r}^{r/2}$, and $C(n) \equiv 1$ to conclude that

$$e^{\text{ran}}(A_{d,n,k}; H_d) \leq \frac{\sqrt{(k+1)M_{d,r}^r}}{n^{p_k}} \quad \text{with} \quad p_k = \frac{r}{2} \left[1 - \left(\frac{r}{r+1} \right)^k \right] \rightarrow \frac{r}{2}.$$

For the algorithms $A'_{n,k}$, Theorem 2 yields

$$e^{\text{ran}}(A'_{n,k}; H_d) \leq \frac{\sqrt{2(k+1)M_{d,r}^r}}{n^{p_k}}.$$

Since r can be arbitrarily close to α , we have

$$e^{\text{ran}}(n; H_d, \Lambda^{\text{std}}) = O(n^{-\alpha/2+\delta}) \quad \text{as } n \rightarrow \infty.$$

The exponent $\alpha/2$ is optimal since even for $d = 1$ we have $e(n; H_1, \Lambda^{\text{all}}) = \Theta(n^{-\alpha/2})$. We stress that the implied factor in the O -notation may depend on δ and d , and may go to infinity as d and/or δ^{-1} goes to infinity; see the next section where this problem is discussed in detail.

6. TRACTABILITY

As in the previous section, we consider multivariate approximation for a sequence of spaces H_d and G_d for $d = 1, 2, \dots$. The weight ρ in the space G_d may also depend on d , and we write $\rho = \rho_d$.

We are now mostly interested in large d and want to verify when the minimal error bounds are polynomially dependent on d or even independent of d . This leads us to *tractability* which has been extensively studied in information-based complexity.

To stress the dependence on d , we denote the minimal errors as

$$\begin{aligned} e^{\text{wor}}(n, d) &= e^{\text{wor}}(n; H_d, \rho_d, \Lambda^{\text{all}}), \\ e^{\text{ran}}(n, d) &= e^{\text{ran}}(n; H_d, \rho_d, \Lambda^{\text{std}}). \end{aligned}$$

For $n = 0$, we do not sample functions, and we have the initial errors

$$e^{\text{wor}}(0, d) = e^{\text{ran}}(0, d) = \|W_d^{1/2}\|,$$

where $W_d = W$ is given by (3).

For $\varepsilon \in (0, 1)$, we denote

$$\begin{aligned} n^{\text{wor}}(\varepsilon, d) &= \min\{n : e^{\text{wor}}(n, d) \leq \varepsilon e^{\text{wor}}(0; d)\}, \\ n^{\text{ran}}(\varepsilon, d) &= \min\{n : e^{\text{ran}}(n, d) \leq \varepsilon e^{\text{ran}}(0; d)\} \end{aligned}$$

as the minimal number of information evaluations needed to reduce the initial error by a factor ε . We stress that one information evaluation in the worst case setting means a functional from the class Λ^{all} , and in the randomized setting a functional from the class Λ^{std} . We are ready to recall the notion of tractability and strong tractability; see [31].

Multivariate approximation is *tractable* in the worst case setting for the class Λ^{all} (and for the sequence of spaces H_d and G_d) iff there are non-negative numbers c, p_{err} and q_{dim} such that

$$n^{\text{wor}}(\varepsilon, d) \leq c \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \quad \forall \varepsilon \in (0, 1), \forall d = 1, 2, \dots$$

Multivariate approximation is *tractable* in the randomized setting for the class Λ^{std} (and for the sequence of spaces H_d and G_d) iff there are non-negative numbers c, p_{err} and q_{dim} such that

$$n^{\text{ran}}(\varepsilon, d) \leq c \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \quad \forall \varepsilon \in (0, 1), \forall d = 1, 2, \dots$$

The numbers p_{err} and q_{dim} are called the exponents of tractability. We stress that they need not to be uniquely defined.

If $q_{\text{dim}} = 0$ in the formulas above, then multivariate approximation is *strongly* tractable and the infimum of p_{err} satisfying the formulas above with $q_{\text{dim}} = 0$ is called the exponent of strong tractability.

Based on Theorem 1, we now prove the following theorem.

Theorem 4. *Consider multivariate approximation for the spaces H_d and G_d . Then*

- (i) *Strong tractability in the worst case setting for the class Λ^{all} is equivalent to strong tractability in the randomized setting for the class Λ^{std} . Furthermore, the exponents of strong tractability are the same.*
- (ii) *Tractability in the worst case setting for the class Λ^{all} is equivalent to tractability in the randomized setting for the class Λ^{std} . Furthermore, the exponents of tractability are roughly the same, i.e., if*

$$n^{\text{wor}}(\varepsilon, d) \leq c \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d = 1, 2, \dots,$$

then

$$\begin{aligned} n^{\text{ran}}(\varepsilon, d) &\leq \min_k k \cdot \left[\left[3 c k^{p_{\text{err}}/2} \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \right] \left[1 - \left(\frac{2}{2+p_{\text{err}}} \right)^k \right]^{-1} \right] \\ &= O \left(\left[\ln(\ln(\varepsilon^{-1} + d + 1)) \right]^{(1+p_{\text{err}})/2} \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \right). \end{aligned}$$

Proof. Clearly, $e^{\text{ran}}(n; H_d, \rho_d, \Lambda^{\text{all}}) \leq e^{\text{ran}}(n; H_d, \rho_d, \Lambda^{\text{std}})$. Remark 1 states that

$$e^{\text{wor}}(n; H_d, \rho_d, \Lambda^{\text{all}}) \leq 2\sqrt{2} e^{\text{ran}}(\lfloor (n+1)/4 \rfloor; H_d, \rho_d, \Lambda^{\text{std}}).$$

This means that (strong) tractability in the randomized setting for the class Λ^{std} implies (strong) tractability in the worst case setting for the class Λ^{all} . Thus, it is enough to prove that (strong) tractability in the worst case setting for Λ^{all} implies (strong) tractability in the randomized setting for Λ^{std} . Assume then that

$$n^{\text{wor}}(\varepsilon, d) \leq c \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \quad \forall \varepsilon \in (0, 1), \forall d = 1, 2, \dots$$

Here, $q_{\text{dim}} = 0$ for strong tractability, and $q_{\text{dim}} > 0$ for tractability. Note that $n^{\text{wor}}(\varepsilon, d) \geq 1$ which for ε tending to one and $d = 1$ implies that $c \geq 1$.

Let $n = \lfloor c \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \rfloor$. If we vary $\varepsilon \in (0, 1)$, then n varies in $(\lfloor c d^{q_{\text{dim}}} \rfloor, \infty)$. It is easy to check that

$$\varepsilon = \left(\frac{c d^{q_{\text{dim}}}}{n + \eta} \right)^{1/p_{\text{err}}} \quad \text{for some } \eta \in [0, 1).$$

From Section 2.1 we know that $e^{\text{wor}}(0, d) = \sqrt{\lambda_{d,1}}$ and

$$n^{\text{wor}}(\varepsilon, d) = \min \left\{ k : \sqrt{\lambda_{d,k+1}} \leq \varepsilon \sqrt{\lambda_{d,1}} \right\},$$

where $\lambda_{d,k}$'s are ordered eigenvalues of the operator $W_d = W$. This yields that

$$(28) \quad \sqrt{\lambda_{d,n+1}} \leq \left(\frac{c d^{q_{\text{dim}}}}{n} \right)^{1/p_{\text{err}}} \sqrt{\lambda_{d,1}} \quad \text{for } n > \lceil c d^{q_{\text{dim}}} \rceil.$$

This implies that (13) holds with $p = 1/p_{\text{err}}$, $C_0 = \sqrt{\lambda_1}$ and

$$C(n) = \begin{cases} \lceil c d^{q_{\text{dim}}} \rceil & \text{for } n = 1, 2, \dots, \lceil c d^{q_{\text{dim}}} \rceil, \\ \left(1 + \frac{1}{n}\right)^{1/p_{\text{err}}} (c d^{q_{\text{dim}}})^{1/p_{\text{err}}} & \text{for } n > \lceil c d^{q_{\text{dim}}} \rceil. \end{cases}$$

From (15) we have

$$p_k = p \left[1 - \left(\frac{2p}{2p+1} \right)^k \right] = \frac{1}{p_{\text{err}}} \left[1 - \left(\frac{2}{2+p_{\text{err}}} \right)^k \right].$$

From Theorem 1 we conclude that

$$e^{\text{ran}}(A_{n,k}; H_d, \rho_d) \leq \frac{\sqrt{k [C(n+1)]^2 + 1}}{n^{p_k}} \sqrt{\lambda_{d,1}} \leq \varepsilon \sqrt{\lambda_{d,1}}$$

holds if

$$\begin{aligned} n &= \left\lceil \varepsilon^{-1/p_k} \left[k \left(1 + \frac{1}{n+1} \right)^{2/p_{\text{err}}} (c d^{q_{\text{dim}}})^{2/p_{\text{err}}} + 1 \right]^{1/(2p_k)} \right\rceil \\ &\leq \left\lceil \varepsilon^{-1/p_k} \left[k 2^{2/p_{\text{err}}} \left(1 + \frac{1}{n+1} \right)^{2/p_{\text{err}}} (c d^{q_{\text{dim}}})^{2/p_{\text{err}}} \right]^{1/(2p_k)} \right\rceil \\ &\leq \left\lceil \left[3 c k^{p_{\text{err}}/2} \varepsilon^{-p_{\text{err}}} d^{q_{\text{dim}}} \right]^{1 - \left(\frac{2}{2+p_{\text{err}}} \right)^k} \right\rceil. \end{aligned}$$

The cardinality of $A_{n,k}$ is at most nk , and therefore $n^{\text{wor}}(\varepsilon, d) \leq nk$, which proves (strong) tractability in the randomized setting for the class Λ^{std} . The exponents of tractability can be arbitrarily close to p_{err} and q_{dim} if k is large enough. This completes the proof of (i) and the first part of (ii). We now take

$$k = \left\lceil \frac{\ln(\ln(\varepsilon^{-1} + d + 1))}{\ln((2 + p_{\text{err}})/2)} \right\rceil.$$

Then $(2/(2 + p_{\text{err}}))^k \leq 1/\ln(\varepsilon^{-1} + d + 1)$ and the second formula in (ii) easily follows. This completes the proof. \square

Theorem 4 states the equivalence between (strong) tractability in the randomized setting for the class Λ^{std} with (strong) tractability in the worst case setting for the class Λ^{all} . We stress that it is relatively easy to verify when (strong) tractability holds in the worst case setting for the class Λ^{all} since it requires to analyze the behavior of the eigenvalues $\lambda_i = \lambda_{d,i}$. The latter problem has already been analyzed in many papers.

We illustrate the last point for the weighted Korobov spaces from Section 5 for general weights $\gamma_{d,\mathbf{u}}$ with $d = 1, 2, \dots$ and $\mathbf{u} \subseteq \{1, 2, \dots, d\}$. For $\boldsymbol{\gamma} = \{\gamma_{d,\mathbf{u}}\}$ and $q \geq 0$, define

$$r_\alpha(\boldsymbol{\gamma}, q) := \sup \left\{ r \in (0, \alpha) : \sup_{d=1,2,\dots} \frac{1}{d^q} \sum_{\mathbf{u} \subseteq \{1,2,\dots,d\}} \left(\frac{\gamma_{d,\mathbf{u}}}{\gamma_{d,\max}} \right)^{\frac{1}{r}} [2\zeta(\alpha/r)]^{|\mathbf{u}|} < \infty \right\}.$$

Here,

$$\gamma_{d,\max} = \max_{\mathbf{u} \subseteq \{1,2,\dots,d\}} \gamma_{d,\mathbf{u}},$$

ζ is, as in the previous section, the zeta Riemann function, and $|\mathbf{u}| = 0$ for $\mathbf{u} = \emptyset$. Note that $\zeta(\alpha/r)$ is well defined since $\alpha/r > 1$ and $2\zeta(\alpha/r) > 2$. We also adopt the convention that $\sup \emptyset = 0$.

Theorem 5. *Consider multivariate approximation for the weighted Korobov spaces defined as in Section 5. Then*

- (i) *Strong tractability in the worst case setting for the class Λ^{all} and in the randomized setting for the class Λ^{std} holds iff $r_\alpha(\boldsymbol{\gamma}, 0) > 0$. When this holds, then the exponent of strong tractability is $2/r_\alpha(\boldsymbol{\gamma}, 0)$.*
- (ii) *Tractability in the worst case setting for the class Λ^{all} and in the randomized setting for the class Λ^{std} holds iff $r_\alpha(\boldsymbol{\gamma}, q) > 0$ for some positive q .*
- (iii) *If $r_\alpha(\boldsymbol{\gamma}, q) > 0$ for some $q \geq 0$, then*

$$n^{\text{wor}}(\varepsilon, d) \leq C_r d^q \varepsilon^{-2/r} \quad \forall r \in (0, r_\alpha(\boldsymbol{\gamma}, q))$$

with

$$C_r = \sup_{d=1,2,\dots} \frac{1}{d^q} \sum_{\mathbf{u} \subseteq \{1,2,\dots,d\}} \left(\frac{\gamma_{d,\mathbf{u}}}{\gamma_{d,\max}} \right)^{\frac{1}{r}} [2\zeta(\alpha/r)]^{|\mathbf{u}|},$$

which is finite.

Proof. Assume first that (strong) tractability holds. Then for some non-negative c, p_{err} and q_{dim} we have

$$n^{\text{wor}}(\varepsilon, d) \leq c d^{q_{\text{dim}}} \varepsilon^{-p_{\text{err}}} \quad \forall \varepsilon \in (0, 1), \forall d = 1, 2, \dots$$

We know that $p_{\text{err}} \geq 2/\alpha$ since $n^{\text{wor}}(\varepsilon, 1) = \Theta(\varepsilon^{2/\alpha})$, and in the case of strong tractability $q_{\text{dim}} = 0$. From (28) we get

$$\lambda_{d,n+1} \leq \left(\frac{c d^{q_{\text{dim}}}}{n} \right)^{2/p_{\text{err}}} \lambda_{d,1} \quad \text{for } n > \lceil c d^{q_{\text{dim}}} \rceil$$

with $\lceil c d^{q_{\text{dim}}} \rceil \geq 1$, and therefore $\lceil c d^{q_{\text{dim}}} \rceil \leq 2c d^{q_{\text{dim}}}$. Note that $\lambda_{d,1} = \gamma_{d,\max}$.

Using (26) we then obtain for $r \in (0, 2/p_{\text{err}})$ and $q \geq 2q_{\text{dim}}/(rp_{\text{err}}) > q_{\text{dim}}$,

$$\begin{aligned} & \frac{1}{d^q} \sum_{\mathbf{u} \subseteq \{1,2,\dots,d\}} \left(\frac{\gamma_{d,\mathbf{u}}}{\gamma_{d,\max}} \right)^{\frac{1}{r}} [2\zeta(\alpha/r)]^{|\mathbf{u}|} = \frac{1}{d^q \lambda_{d,1}^{1/r}} \sum_{j=1}^{\infty} \lambda_{d,j}^{1/r} \\ & \leq \frac{\lceil c d^{q_{\text{dim}}} \rceil}{d^q} + \frac{(c d^{q_{\text{dim}}})^{2/(rp_{\text{err}})}}{d^q} \sum_{j=\lceil c d^{q_{\text{dim}}} \rceil+1}^{\infty} j^{-2/(rp_{\text{err}})} \\ & \leq \frac{2c}{d^{q-q_{\text{dim}}}} + \frac{c^{2/(rp_{\text{err}})}}{d^{q-2q_{\text{dim}}/(2p_{\text{err}})}} \zeta \left(\frac{2}{rp_{\text{err}}} \right). \end{aligned}$$

Hence $r_\alpha(\boldsymbol{\gamma}, q) \geq r > 0$. Since r can be arbitrarily close to $2/p_{\text{err}}$ this yields that $r_\alpha(\boldsymbol{\gamma}, q) \geq 2/p_{\text{err}} > 0$. For strong tractability we have $q_{\text{dim}} = 0$ and we can take $q = 0$ which implies that $r_\alpha(\boldsymbol{\gamma}, 0) \geq 2/p_{\text{err}} > 0$.

Assume now that $r_\alpha(\boldsymbol{\gamma}, q) > 0$ with $q = 0$ for strong tractability. Hence, for $r \in (0, r_\alpha(\boldsymbol{\gamma}, q))$ we have $C_r < \infty$ with C_r given in (iii). From (27) we conclude that

$$\begin{aligned} \lambda_{d,n} &\leq \left(\sum_{j=1}^{\infty} \lambda_{d,j}^{1/r} \right)^r n^{-r} \\ &= \left(\sum_{u \subseteq \{1,2,\dots,d\}} \left(\frac{\gamma_{d,u}}{\gamma_{d,\max}} \right)^{\frac{1}{r}} [2\zeta(\alpha/r)]^{|u|} \right)^r n^{-r} \lambda_{d,1} \\ &\leq C_r^r d^{qr} n^{-r} \lambda_{d,1}. \end{aligned}$$

Then $\lambda_{d,n+1} \leq \varepsilon^2 \lambda_{d,1}$ if $C_r^r d^{qr} (n+1)^{-r} \leq \varepsilon^2$, which holds for

$$n^{\text{wor}}(\varepsilon, d) \leq n \leq C_r d^q \varepsilon^{-2/r}.$$

Hence, we have tractability for $q > 0$ and strong tractability for $q = 0$. Furthermore, r can be arbitrarily close to $r_\alpha(\boldsymbol{\gamma}, q)$ and therefore

$$n^{\text{wor}}(\varepsilon, d) \leq C_r d^q \varepsilon^{-p_{\text{err}}}$$

with p_{err} arbitrarily close to $2/r_\alpha(\boldsymbol{\gamma}, q)$. For $q = 0$ this means that the exponent of strong tractability is $2/r_\alpha(\boldsymbol{\gamma}, q)$. This completes the proof. \square

The definition of $r(\boldsymbol{\gamma}, q)$ simplifies for product weights, i.e., for

$$\gamma_{d,\emptyset} = 1 \quad \text{and} \quad \gamma_{d,u} = \prod_{j \in u} \gamma_{d,j} \quad \text{with} \quad 0 \leq \gamma_{d,d} \leq \dots \leq \gamma_{d,1} \leq 1.$$

Indeed, we then have

$$r_\alpha(\boldsymbol{\gamma}, q) = \sup \left\{ r \in (0, \alpha) : \sup_d d^{-q} \prod_{j=1}^d \left(1 + 2\gamma_{d,j}^{1/r} \zeta(\alpha/r) \right) < \infty \right\}.$$

The necessary and sufficient conditions for (strong) tractability can be simplified due to the following equivalences:

$$(29) \quad r_\alpha(\boldsymbol{\gamma}, 0) > 0 \quad \text{iff} \quad \sup_d \sum_{j=1}^d \gamma_{d,j}^{1/r} < \infty \quad \text{for some } r \in (0, \alpha),$$

and
(30)

$$r_\alpha(\boldsymbol{\gamma}, q) > 0 \quad \text{for some } q > 0 \quad \text{iff} \quad \sup_d \frac{\sum_{j=1}^d \gamma_{d,j}^{1/r}}{\ln(d+1)} < \infty \quad \text{for some } r \in (0, \alpha).$$

For example, take $\gamma_{d,j} = 1$ for $j = 1, 2, \dots, \min(d, \lceil \beta \ln(d+1) \rceil)$ for some positive β . Then strong tractability does *not* hold since $r_\alpha(\boldsymbol{\gamma}, 0) = 0$. However, tractability *holds* since $r_\alpha(\boldsymbol{\gamma}, q) > 0$ for $q \geq \beta \ln(1 + 2\zeta(\alpha/r))$ with $r \in (0, \alpha)$.

The situation is even simpler when the weights are independent of d as indicated in the following proposition.

Proposition 4. Consider multivariate approximation for the weighted Korobov spaces as in Section 5. If the weights γ are of product form and do not depend on d , i.e.,

$$\gamma_{d,j} = \gamma_j \quad \forall d, j \quad \text{and} \quad 1 \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq 0,$$

then tractability is equivalent to strong tractability which, in turn, is equivalent to

$$(31) \quad \sum_{j=1}^{\infty} \gamma_j^{1/r} < \infty$$

for some $r \in (0, \alpha)$. When this holds, then the exponent of strong tractability equals $2/\min(\alpha, r^*)$, where r^* is the supremum of r satisfying (31).

Proof. It is enough to prove that tractability implies strong tractability. Assume that we have tractability. Then $r(\gamma, q) > 0$ for some positive q , and (30) yields

$$c_{\infty} := \sup_d \frac{\sum_{j=1}^d \gamma_j^{1/r}}{\ln(d+1)} < \infty$$

for a positive number $r < \alpha$. Then for any $d \geq 1$, we have $d\gamma_d^{1/r} \leq c_{\infty} \ln(d+1)$. For any positive δ , and $a = r/(1+\delta)$ we obtain

$$\sum_{j=1}^{\infty} \gamma_j^{1/a} \leq c_{\infty}^{1+\delta} \sum_{j=1}^{\infty} \frac{\ln^{1+\delta}(d+1)}{d^{1+\delta}} < \infty.$$

This and the equivalence (29) imply strong tractability and that $r_{\alpha}(\gamma, 0) \geq a$. Since a can be arbitrarily close to r and r can be arbitrarily close to $\min(\alpha, r^*)$, we conclude that $r_{\alpha}(\gamma, 0) \geq \min(\alpha, r^*)$. For $r \geq \min(\alpha, r^*)$, the product $\prod_{j=1}^{\infty} (1 + 2\gamma_j^{1/r} \zeta(\alpha/r))$ is divergent or not well defined, and therefore $r_{\alpha}(\gamma, 0) \leq r^*$. Hence, $r_{\alpha}(\gamma, 0) = \min(\alpha, r^*)$. Theorem 5 states that the exponent of strong tractability is $2/r_{\alpha}(\gamma, 0)$ which completes the proof. \square

ACKNOWLEDGMENTS

We would like to thank Frances Kuo and Joseph F. Traub for valuable comments. We also thank Przemyslaw Wojtaszczyk for a question that led us to Proposition 2. The authors were partially supported by the National Science Foundation under Grants CCR-0511994 and DMS-0308713, respectively.

REFERENCES

- [1] N. Aronszajn, Theory of reproducing kernels, *Trans. AMS* **68** (1950), 337-404. MR0051437 (14:479c)
- [2] N. S. Bakhvalov, On approximate calculation of integrals, *Vestnik MGU, Ser. Mat. Mekh. Astron. Fiz. Him.* **4** (1959), 3-18. MR0115275 (22:6077)
- [3] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights, to appear in *Numer. Math.*
- [4] B. L. Granovskii and S. M. Ermakov, The Monte Carlo method, *J. Soviet Math.* **7** (1977), 161-192.
- [5] M. Griebel and H. Woźniakowski, On the optimal convergence rate of universal and non-universal algorithms for multivariate integration and approximation, to appear in *Math. Computation.*
- [6] S. Heinrich, Lower bounds for the complexity of Monte Carlo function approximation, *J. Complexity* **8** (1992), 277-300. MR1187419 (93h:65007)
- [7] S. Heinrich, Monte Carlo complexity of global solution of integral equations, *J. Complexity* **14** (1998), 151-175. MR1629093 (99g:65135)

- [8] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimensions, *Advances in Comput. Math.* **12** (2000), 25-58. MR1758946 (2001d:65017)
- [9] F. Kuo, I. H. Sloan and H. Woźniakowski, Lattice rules for multivariate approximation in the worst case setting, in *Monte Carlo and Quasi-Monte Carlo Methods 2004*, eds. H. Niederreiter and D. Talay, Springer-Verlag, Berlin, pp. 289-330, 2006. MR2208715 (2006j:41044)
- [10] F. Kuo, I. H. Sloan and H. Woźniakowski, Lattice rules for multivariate approximation in the average case setting, in progress.
- [11] P. Mathé, Random approximation of Sobolev embeddings, *J. Complexity* **7** (1991), 261-281. MR1128021 (93d:65008)
- [12] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley, New York, 1983. MR0702836 (84g:90079)
- [13] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lect. Notes in Math. **1349**, Springer-Verlag, Berlin, 1988. MR0971255 (90a:65004)
- [14] E. Novak, Stochastic properties of quadrature formulas, *Numer. Math.* **53** (1988), 609-620. MR0954773 (89i:65026)
- [15] E. Novak, Optimal linear randomized methods for linear operators in Hilbert spaces, *J. Complexity* **8** (1992), 22-36. MR1153612 (93f:65048)
- [16] E. Novak, I. H. Sloan and H. Woźniakowski, Tractability of approximation for weighted Korobov spaces on classical and quantum computers, *Found. Comput. Math.* **4** (2004), 121-156. MR2049868 (2005h:81097)
- [17] A. Papageorgiou and G. W. Wasilkowski, On the average complexity of multivariate problems, *J. Complexity* **6** (1990), 1-23. MR1048027 (91b:94020)
- [18] H. Pfeiffer, Monte Carlo methods with few random bits, *Ph.D. Thesis*, University of Jena, 2005.
- [19] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge Univ. Press, Cambridge, 1996. MR1446005 (99b:65189)
- [20] K. Ritter, *Average-Case Analysis of Numerical Problems*, Lect. Notes in Math. **1733**, Springer-Verlag, Berlin, 2000. MR1763973 (2001i:65001)
- [21] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988. MR0958691 (90f:68085)
- [22] J. F. Traub and A. G. Werschulz, *Complexity and Information*, Cambridge Univ. Press, Cambridge, 1998. MR1692462 (2000m:65170)
- [23] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980. MR0584446 (84m:68041)
- [24] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF **59**, Philadelphia, 1990. MR1045442 (91g:62028)
- [25] G. W. Wasilkowski, Some nonlinear problems are as easy as the approximation problem, *Comput. Math. Appl.* **10** (1984), 351-363. MR0777390 (86j:41023)
- [26] G. W. Wasilkowski, Randomization for continuous problems, *J. Complexity* **5** (1989), 195-218. MR1006106 (91a:65006)
- [27] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor-product algorithms for linear multivariate problems, *J. of Complexity* **15**, (1999), 402-447. MR1716741 (2000h:65200)
- [28] G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for weighted approximation, *Found. Comput. Math.* **1** (2001), 417-434. MR1857723 (2002g:41052)
- [29] A. G. Werschulz, *The Computational Complexity of Differential and Integral Equations*, Oxford Univ. Press, Oxford, 1991. MR1144521 (93a:68061)
- [30] A. G. Werschulz and H. Woźniakowski, Tractability of quasilinear problems I: general results, II: second-order elliptic problems, submitted for publication, 2005.
- [31] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity*, **10** (1994), 96-128. MR1263379 (95d:65117)

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF KENTUCKY, 773 ANDERSON HALL, LEXINGTON, KENTUCKY 40506-0046

E-mail address: `greg@cs.uky.edu`

DEPARTMENT OF COMPUTER SCIENCE, COLUMBIA UNIVERSITY, NEW YORK, NY 10027, USA,
AND INSTITUTE OF APPLIED MATHEMATICS, UNIVERSITY OF WARSAW, WARSAW, POLAND

E-mail address: `henryk@cs.columbia.edu`