LOWER BOUNDS FOR THE COMPLEXITY OF LINEAR FUNCTIONALS IN THE RANDOMIZED SETTING

ERICH NOVAK AND HENRYK WOŹNIAKOWSKI

Abstract. Hinrichs [3] recently studied multivariate integration defined over reproducing kernel Hilbert spaces in the randomized setting and for the normalized error criterion. In particular, he showed that such problems are strongly polynomially tractable if the reproducing kernels are pointwise nonnegative and integrable. More specifically, let $n^{\rm ran}(\varepsilon,{\rm INT}_d)$ be the minimal number of randomized function samples that is needed to compute an ε-approximation for the d-variate case of multivariate integration. Hinrichs proved that

$$n^{\rm ran}(\varepsilon,{\rm INT}_d)\le\left\lceil\frac{\pi}{2}\left(\frac{1}{\varepsilon}\right)^2\right\rceil\quad\text{for all}\quad\varepsilon\in(0,1)\ \text{ and }\ d\in\mathbb{N}.$$

In this paper we prove that the exponent 2 of ε^{-1} is sharp for tensor product Hilbert spaces whose univariate reproducing kernel is *decomposable* and univariate integration is not trivial for the two parts of the decomposition. More specifically we have

$$n^{\rm ran}(\varepsilon,{\rm INT}_d)\ge\left\lceil\frac{1}{8}\left(\frac{1}{\varepsilon}\right)^2\right\rceil\quad\text{for all}\quad\varepsilon\in(0,1)\ \text{ and }\ d\ge\frac{2\,\ln\varepsilon^{-1}\,-\,\ln 2}{\ln\alpha^{-1}},$$

where $\alpha\in[1/2,1)$ depends on the particular space.

We stress that these estimates hold independently of the smoothness of functions in a Hilbert space. Hence, even for spaces of very smooth functions the exponent of strong polynomial tractability must be 2.

Our lower bounds hold not only for multivariate integration but for all linear tensor product functionals defined over a Hilbert space with a decomposable reproducing kernel and with a non-trivial univariate functional for the two spaces corresponding to decomposable parts. We also present lower bounds for reproducing kernels that are not decomposable but have a decomposable part. However, in this case it is not clear if the lower bounds are sharp.

1. Introduction

The motivation of this paper comes from the recent paper of Hinrichs [3] who studied multivariate integration defined over reproducing kernel Hilbert spaces. Multivariate integration is a very popular research subject with numerous applications especially for the d-variate case with large or huge d. Multivariate integration has been studied in many settings including the worst case, average and randomized setting. It is well known that the worst case and average case setting are technically

Date: March 2010; revised: August 2010.

The research reported in this paper has been partially done while the first author visited Columbia University, and has been supported in part by DFG and NSF.

very much related and it is usually easy to translate the results from one setting to the other.

The randomized setting is different and that was the setting studied in [3]. The primary example of an algorithm for multivariate integration in the randomized setting is obviously Monte Carlo (MC). It is well known that the error of MC with n random function samples behaves like $\mathcal{O}(n^{-1/2})$. Here the factor in the big $\mathcal{O}$ notation depends on the variance of the integrand. In general, the variance can be an arbitrary function of d. In particular, the variance can be exponential in d. Then for large d, we must take n exponentially large in d to guarantee a reasonably small error. It is a priori not clear if this bad dependence on d is just a bad property of Monte Carlo or an intrinsic property of multivariate integration in a given space.

The surprising result of Hinrichs is that there is *no* dependence on d if we switch from the standard Monte Carlo to *importance sampling* with a properly chosen density function. This holds under the following assumptions.

- The *normalized* error criterion is chosen. That is, we want to reduce the error that can be achieved without sampling the function by a factor of $\varepsilon \in (0,1)$.
- For all d the reproducing kernel of the Hilbert space for the d-variate case is pointwise nonnegative and integrable.

Hinrichs [3] proved that there exists a density function such that the importance sampling computes an ε approximation for the d-variate case with

$$n = \left\lceil \frac{\pi}{2} \left(\frac{1}{\varepsilon}\right)^2 \right\rceil$$

randomized function samples. So there is no dependence on d, however, the power 2 of ε^{-1} is independent of the Hilbert spaces.

One may hope that at least for some Hilbert spaces, we can get a better result. Ideally, we would like to preserve the independence on d and improve the dependence on ε^{-1} by lowering the exponent 2. This hope can be justified by remembering that smoothness of functions sometimes permits the reduction of the exponent of ε^{-1}. For instance, it is known that for $d = 1$ and r times continuously differentiable functions

$$\Theta(\varepsilon^{-1/(r+1/2)})$$

randomized function samples are enough to compute an ε-approximation. For $d > 1$, if we take the d-fold tensor product of such spaces then we need

$$\mathcal{O}\left(\varepsilon^{-1/(r+1/2)} \left[\ln \varepsilon^{-1}\right]^{p(d,r)}\right)$$

randomized function samples to compute an ε-approximation, where the exponent $p(d,r)$ of $\ln \varepsilon^{-1}$ is linear in d and r. However, it is not known how the factor in the big $\mathcal{O}$ notation depends on d. A priori, we do not know whether there is a tradeoff between the dependence on d and ε^{-1}.

Let $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ denote the minimal number of randomized function samples that is needed to compute an ε-approximation for d-variate integration. We stress that $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ is the intrinsic difficulty of multivariate integration in the randomized setting since we now allow all possible algorithms including Monte Carlo, importance sampling with an arbitrary density function, as well as other linear or nonlinear randomized algorithms. Clearly, the result of Hinrichs can be rewritten as

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \le \left\lceil \frac{\pi}{2} \left(\frac{1}{\varepsilon}\right)^2 \right\rceil \quad \text{for all} \quad \varepsilon \in (0,1) \ \text{ and } \ d \in \mathbb{N}.$$

In this paper we study, in particular, whether the last bound is sharp, or more precisely if we can preserve the independence on d and lower the exponent of ε^{-1}. We study this question for tensor product Hilbert spaces. These spaces are generated by a reproducing kernel Hilbert space of univariate functions. This corresponds to the *unweighted* problem in which all variables and groups of variables play the same role. We prove that the exponent 2 *cannot* be lowered. This holds if we assume that

- the univariate reproducing kernel is *decomposable* in the sense of [5],
- the univariate integration is non-zero if restricted to the space corresponding to the decomposable parts of the kernel.

The first assumption means that the univariate reproducing kernel

$$K_1 : D_1 \times D_1 \to \mathbb{R} \quad \text{with} \quad D_1 \subseteq \mathbb{R},$$

has the property that there exists a point $a \in \mathbb{R}$ such that

$$K_1(x,y) = 0 \quad \text{for all} \quad x, y \in D_1 \ \text{ and } \ x \le a \le y.$$

The second assumption means that univariate integration is not zero when the domain is restricted to one of the domains

$$D_{(0)} := \{x \in D_1 \ : \ x \le a\} \quad \text{and} \quad D_{(1)} := \{x \in D_1 \ : \ x \ge a\}.$$

We stress that these assumptions are not related to the smoothness of functions from the Hilbert space. As we shall see these assumptions hold for certain Sobolev spaces with arbitrary high smoothness of functions.

More specifically, we prove that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \ge \left\lceil \frac{1}{8} \left(\frac{1}{\varepsilon}\right)^2 \right\rceil \quad \text{for all} \quad \varepsilon \in (0,1) \ \text{ and } \ d \ge \frac{2\,\ln \varepsilon^{-1} - \ln 2}{\ln \alpha^{-1}}.$$

Here $\alpha \in [1/2, 1)$ measures the difficulty of the univariate integration problem over $D_{(0)}$ and $D_{(1)}$. If the univariate case is equally difficult over $D_{(0)}$ and $D_{(1)}$ then we have $\alpha = 1/2$.

We now comment on the condition on d which requires that d is large relative to ε^{-1}. First of all, note that the lower bound presented above cannot be true for all $\varepsilon \in (0,1)$, $d \in \mathbb{N}$ and Hilbert spaces satisfying the assumptions mentioned above. The reason is simple since for smooth functions the exponent of ε^{-1} is smaller than 2. That is, for a fixed d and ε tending to zero, the asymptotic behavior of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ may be better, or even much better, than ε^{-2}. That is why the lower bound presented above must relate ε^{-1} and d. On the other hand, note that the condition on d is quite mild since the dependence on ε^{-1} is only logarithmic.

The main point of the lower bound is that smoothness can *not* lower the exponent of ε^{-1} if we insist on the independence on d. This also means that there may be an important difference between the asymptotic behavior of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ when d is fixed and ε goes to zero and the behavior of $n^{\text{ran}}(\varepsilon, \text{INT}_d)$ when ε is fixed and d goes to infinity.

The lower bounds presented in this paper hold not only for multivariate integration but for all linear tensor product functionals defined over Hilbert spaces with decomposable reproducing kernels. We also study non-decomposable kernels which have a decomposable part. In this case, the lower bounds are almost the same as before only if the part of the univariate linear functional corresponding to the non-decomposable part of the reproducing kernel has a small norm. It is not clear what are sharp lower and upper bounds for general linear tensor product functionals.

We now briefly compare the results for linear tensor product functionals for Hilbert spaces with decomposable reproducing kernels in the worst case and randomized settings. In the worst case setting, it is proved in [5] that such problems are *intractable* since they suffer from the *curse of dimensionality*. This means we need to compute exponentially many function values in d to get an ε-approximation for the d-variate case. ¿From this point of view, the positive results on strong polynomial tractability in the randomized setting are even more surprising. We must admit that after we completed the paper [5] on lower bounds in the worst case setting, we started to work on lower bounds in the randomized setting around the year 2002. Fairly soon we realized that we cannot prove the curse of dimensionality for decomposable kernels in the randomized setting since the lower bound had a factor $n^{-1/2}$ independently of the Hilbert space, i.e., independently of the smoothness of functions. At that time we felt sure that our lower bounds were too loose. We regarded the factor $n^{-1/2}$ as a sign that our analysis is not good enough. After a few more trials, we gave up still being (almost) certain that $n^{-1/2}$ is not needed. After a few years, Hinrichs saved, in a way, our previous work by showing that the factor $n^{-1/2}$ is indeed needed and that our intuition was simply wrong.

We finally briefly comment on a number of possible future directions related to the randomized setting.

- The result of Hinrichs is for multivariate integration, and the lower bounds are for linear tensor product functionals. It would be of interest to see if the result of Hinrichs can be extended for linear tensor product functionals. In fact, some linear functionals can be interpreted as multivariate integration, see Section 10.9 of [7] but is it not clear if we can do this for all such functionals.
- We already mentioned that the lower bounds for the case of reproducing kernels with only a decomposable part are not always satisfactory. Of course, it would be of interest to improve them. It is not clear but perhaps the upper bounds can also be improved and strong polynomial tractability with the exponent smaller than 2 can be obtained at least for some linear tensor product functionals with nontrivial decomposable parts.
- We have so far discussed the *unweighted* spaces in which all variables and groups of variables play the same role. Obviously, we should analyze *weighted* spaces in which we moderate the influence of all groups of variables by weights. In the worst case setting, the analysis of weighted spaces has been a major research trend with many positive tractability results under the conditions of

proper decay of weights. Some work has been also done in the randomized setting. However, the consequences of the result of Hinrichs for weighted spaces have not yet been found. In particular, we would like to know what we have to assume about the weights to get a smaller exponent of strong polynomial tractability than 2.

2. The Result of Hinrichs

We briefly define the problem studied by Hinrichs [3]. Let $H(K_d)$ be a reproducing kernel Hilbert space of real functions defined on a Borel measurable set $D_d \subseteq \mathbb{R}^d$. Its reproducing kernel $K_d : D_d \times D_d \to \mathbb{R}$ is assumed to be integrable,

$$C_d^{\rm init} := \left(\int_{D_d}\int_{D_d} K_d(x,y)\,\varrho_d(x)\,\varrho_d(y)\,\mathrm{d}x\,\mathrm{d}y\right)^{1/2} < \infty.$$

Here, ϱ_d is a probability density function on D_d. Without loss of generality we assume that D_d and ϱ_d are chosen such that there is no subset of D_d with positive measure such that all functions from $H(K_d)$ vanish on it.

The inner product and the norm of $H(K_d)$ are denoted by $\langle\cdot,\cdot\rangle_{H(K_d)}$ and $\|\cdot\|_{H(K_d)}$. Consider multivariate integration

$$\mathrm{INT}_d(f) = \int_{D_d} f(x)\,\varrho_d(x)\,\mathrm{d}x \quad \text{for all} \quad f \in H(K_d).$$

We approximate $\mathrm{INT}_d(f)$ in the randomized setting using *importance sampling.* That is, for a probability density function ω_d on D_d we choose n random sample points $x_1, x_2, \dots, x_n$ which are independent and distributed according to ω_d and take the algorithm

$$A_{n,d,\omega_d}(f) = \frac{1}{n}\sum_{j=1}^{n} \frac{f(x_j)\,\varrho_d(x_j)}{\omega_d(x_j)}.$$

The error of A_{n,d,ω_d} is defined as

$$e^{\rm ran}(A_{n,d,\omega_d}) = \sup_{\|f\|_{H(K_d)}\le 1} \left(\mathbb{E}_{\omega_d}\left(\mathrm{INT}_d(f) - A_{n,d,\omega_d}(f)\right)^2\right)^{1/2},$$

where the expectation is with respect to the random choice of the sample points x_j.

For $n = 0$ we formally take $A_{0,d,\omega_d} = 0$ and then

$$e^{\rm ran}(0, I_d) = C_d^{\rm init}.$$

The error $e^{\rm ran}(0)$ is called the *initial* error and can be obtained without sampling the function. This also explains the use of the superscript init.

Hinrichs [3] proved, in particular, the following theorem.

Theorem 1. [3, Theorem 4]

Assume additionally that $K_d(x,y) \ge 0$ for all $x,y \in D_d$. Then there exists a positive density function ω_d such that

$$e^{\rm ran}(A_{n,d,\omega_d}) \le \left(\frac{\pi}{2}\right)^{1/2} \frac{1}{\sqrt{n}}\, e^{\rm ran}(0, I_d).$$

Hence, if we want to achieve $e^{\mathrm{ran}}(A_{n,d,\omega_d}) \le \varepsilon\, e^{\mathrm{ran}}(0, I_d)$ *then it is enough to take*

$$n = \left\lceil \frac{\pi}{2} \left(\frac{1}{\varepsilon}\right)^2 \right\rceil .$$

We briefly comment on the assumption on $K_d(x,y) \ge 0$ for all $x, y \in D_d$. In general, this assumption is needed. Indeed, we will show this for an example which is a modification of the example studied in Section 17.1.6.2 of [7]. More precisely, for $d = 1$ we define the space $H(K_1)$ of real functions defined over $[0,1]$ such that they are constant over $[0,1/2]$ and $(1/2,1]$. That is, $f(x) = f(0)$ for all $x \in [0,1/2]$, and $f(x) = f(1)$ for all $x \in (1/2,1]$. The inner product of f, g from $H(K_1)$ is defined by

$$\langle f, g\rangle_{H(K_1)} = \frac{1}{2}\,[f(0)+f(1)]\,[g(0)+g(1)] + \frac{1}{4}\,[f(0)g(0)+f(1)g(1)]\,.$$

For $i = 1,2$, consider two functions f_i from $H(K_1)$ such that

$$\begin{aligned} f_1(0) &= 2/\sqrt{3}, & f_1(1) &= 0 \\ f_2(0) &= 4/\sqrt{15}, & f_2(1) &= -6/\sqrt{15}. \end{aligned}$$

It is easy to check that $\langle f_i, f_j\rangle_{H(K_1)} = \delta_{i,j}$. Therefore the reproducing kernel is

$$K_1(x,y) = f_1(x)f_1(y) + f_2(x)f_2(y) \quad \text{for all} \quad x, y \in [0,1].$$

We have $K_1(x,t) = 12/5$ if $x, t \in [0,1/2]$ or $x, t \in (1/2,1]$ and $K_1(x,t) = -8/5$ otherwise, i.e., if $x \le 1/2 < t$ or $t \le 1/2 < x$. Univariate integration takes now the form

$$\mathrm{INT}_1(f) = \int_0^1 f(t)\,\mathrm{d}t = \frac{1}{2}\,[f(0)+f(1)] \quad \text{for all} \quad f \in H(K_1).$$

For $d > 1$, we take the space $H(K_d)$ of real functions that are constant on the 2^d products of intervals $[0,1/2]$ and $(1/2,1]$. This is, each function in $H(K_d)$ is uniquely defined by its values at the 2^d points $\{0,1\}^d$. We define the inner product for all $f, g \in H(K_d)$ by

$$\langle f, g\rangle_{H(K_d)} = 2^{-d}\left[\sum_{x\in\{0,1\}^d} f(x)\right]\left[\sum_{x\in\{0,1\}^d} g(x)\right] + 4^{-d}\sum_{x\in\{0,1\}^d} f(x)g(x).$$

It can be checked that the reproducing kernel is

$$K_d(x,t) = 4^d\left(1 - \frac{1}{2^d + 2^{-d}}\right)$$

if all coordinates of x and t lie in the same subinterval $[0,1/2]$ or $(1/2,1]$ while

$$K_d(x,t) = -\frac{2^d}{1+4^{-d}}$$

if at least one of the coordinates x_i and t_i lie in different subintervals. Hence, the kernel K_d does not satisfy the assumption of Theorem 1 of Hinrichs.

Consider multivariate integration

$$\mathrm{INT}_d(f) = \int_{[0,1]^d} f(t)\,\mathrm{d}t = \frac{1}{2^d} \sum_{x\in\{0,1\}^d} f(x) \quad \text{for all} \quad f \in H(K_d).$$

Observe that the norm of multivariate integration is given by $\|\mathrm{INT}_d\| = \sqrt{2^{-d}/(1+4^{-d})}$.

Similarly as in [7], see the proof of Theorem 17.14, we can apply Lemma 1 below (with $N = 2^d$ and f_i being equal to $\sqrt{2^d/(1+2^{-d})}$ on one of the 2^d subregions of $[0,1]^d$ and zero otherwise, which corresponds to $\eta = (1+2^d)^{-1/2}$) to conclude that for $\varepsilon^2 = \frac{1}{2}$

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \geq 2^d \left(1 - \frac{1+2^{-d}}{2+2\cdot 4^{-d}}\right).$$

This means that multivariate integration suffers from the curse of dimensionality and Theorem 1 does not hold for this space $H(K_d)$ since its reproducing kernel takes also negative values.

Formally, the result of Hinrichs seems to be only for multivariate integration. However, it turns out that some linear functionals can be expressed as multivariate integration and this of course extends applicability of Theorem 1, for details see Section 10.9 in [7].

3. Linear Tensor Product Functionals

In this section we define linear tensor product functionals over reproducing kernel Hilbert spaces. These problems are not necessarily given as multivariate integration. The basic information on this subject can be found, e.g, in [1, 9].

For $d = 1$, we assume that $H(K_1)$ is a reproducing kernel Hilbert space of real functions defined over $D_1 \subset \mathbb{R}$ with the kernel $K_1 : D_1 \times D_1 \to \mathbb{R}$. The inner product of $H(K_1)$ is denoted by $\langle\cdot,\cdot\rangle_{H(K_1)}$. Consider the continuous linear functional

$$I_1(f) = \langle f, h_1\rangle_{H(K_1)} \quad \text{for all} \quad f \in H(K_1).$$

Here h_1 is some function from $H(K_1)$.

For $d > 1$, we take

$$H(K_d) = H(K_1) \otimes H(K_1) \otimes \cdots \otimes H(K_1)$$

as the d-fold tensor product of $H(K_1)$. Then $H(K_d)$ is a reproducing kernel Hilbert space of multivariate functions defined over $D_d = D_1 \times D_1 \times \cdots \times D_1$ (d times) with the kernel $K_d : D_d \times D_d \to \mathbb{R}$ given by

$$K_d(x,y) = \prod_{j=1}^{d} K_1(x_j, y_j) \quad \text{for all} \quad x = [x_1, x_2, \ldots, x_d],\ y = [y_1, y_2, \ldots, y_d] \in D_d.$$

The inner product of $H(K_d)$ is denoted by $\langle\cdot,\cdot\rangle_{H(K_d)}$. Finally, the continuous linear functional

$$I_d = I_1 \otimes \cdots \otimes I_1$$

is the d-fold tensor product of I_1. This means that

$$I_d(f) \;=\; \langle f, h_d\rangle_{H(K_d)} \quad \text{for all} \quad f \in H(K_d)$$

and

$$h_d(x) \;=\; h_1(x_1)h_1(x_2)\cdots h_1(x_d) \quad \text{for all} \quad x \in D_d.$$

4. Randomized Setting and Tractability

We approximate linear tensor product functionals in the randomized setting. We now briefly define this setting as well as recall a few notions of tractability. The reader may find more on these subjects, e.g., in [6, 7, 8].

We approximate I_d by algorithms A_{n,d,ω_d} that use n function values on the average and each function value is computed at a random sample point from D_d chosen with respect to a probability distribution on D_d. More precisely, the algorithm A_{n,d,ω_d} is of the following form

$$A_{n,d,\omega_d}(f) = \varphi_{n,d,\omega_d}\left(f(t_{1,\omega_{d,1}}), f(t_{2,\omega_{d,2}}), \dots, f(t_{n(\omega_d),\omega_{d,n(\omega_d)}})\right).$$

Here $\omega_d = [\omega_{d,1}, \omega_{d,2}, \dots]$, and the sample points $t_{j,\omega_{d,j}}$ are random points distributed according to a probability distribution $\omega_{d,j}$ on D_d which may depend on j as well as on the function values already computed. The mapping $\varphi_{n,d,\omega_d} : \mathbb{R}^{n(\omega_d)} \to \mathbb{R}$ is a random mapping, and

$$\mathbb{E}_{\omega_d}\, n(\omega_d) \le n.$$

We also allow adaptive choices of sample points. That is, t_{j,ω_d} may depend on the already selected sample points $t_{1,\omega_d}, t_{2,\omega_2}, \dots, t_{j-1,\omega_d}$.

Without loss of generality, we assume that $A_{n,d,\cdot}(f)$ is measurable, and define the error of A_{n,d,ω_d} as

$$e^{\mathrm{ran}}(A_{n,d,\omega_d}) = \sup_{\|f\|_{H(K_d)}\le 1} \left(\mathbb{E}_{\omega_d}\left(I_d(f) - A_{n,d,\omega_d}(f)\right)^2\right)^{1/2}.$$

For $n = 0$, the algorithm A_{0,d,ω_d} does not depend on any function values and it is easy to check that the error is minimized when we take $A_{0,d,\omega_d} \equiv 0$. Then

$$e^{\mathrm{ran}}(0) = \|I_d\| = \|h_d\|_{H(K_d)} = \|h_1\|^d_{H(K_d)}.$$

Hence, $e^{\mathrm{ran}}(0) = 0$ only for trivial problems with $h_1 \equiv 0$. Therefore, we always assume that $h_1 \ne 0$.

For a given n, we would like to find an algorithm with the nth minimal error. Let

$$e^{\mathrm{ran}}(n, I_d) \;=\; \inf\left\{e^{\mathrm{ran}}(A_{n,d,\omega_d}) \;\middle|\; A_{n,d,\omega_d} \text{ as above}\right\} \tag{1}$$

be the nth minimal error when we use n randomized function values on the average. We stress that we minimize the error with respect to all possible probability distributions ω_d, adaptive sample points x_j, as well as random mappings φ_{n,d,ω_d}. Obviously, $e^{\mathrm{ran}}(0, I_d) = e^{\mathrm{ran}}(0) = \|I_d\|$.

We would like to reduce the initial error by a factor ε, where $\varepsilon \in (0,1)$. We are looking for the smallest $n = n(\varepsilon, I_d)$ for which $e^{\mathrm{ran}}(n,d) \le \varepsilon e^{\mathrm{ran}}(0, I_d)$. That is,

$$n^{\mathrm{ran}}(\varepsilon, I_d) \;=\; \min\left\{\, n \;:\; e^{\mathrm{ran}}(n, I_d) \le \varepsilon\, e^{\mathrm{ran}}(0, I_d) \,\right\}. \tag{2}$$

We now turn to tractability that studies when $n^{\rm ran}(\varepsilon,d)$ is *not* exponential in ε^{-1} or d. Since there are many different ways to define the lack of exponential dependence we have various kinds of tractability.

We say that the problem $I=\{I_d\}$ is *polynomially tractable* iff there exist nonnegative C,q and p such that

$$n^{\rm ran}(\varepsilon,I_d)\,\le\, C\,d^{\,q}\,\varepsilon^{-p}\quad\text{for all}\quad \varepsilon\in(0,1)\ \ \text{and}\ \ d\in\mathbb{N}. \tag{3}$$

Polynomial tractability means that we can reduce the initial error by a factor ε by using a number of function values that is polynomial in ε^{-1} and d.

If $q=0$ in (3), then we say that the problem $I=\{I_d\}$ is *strongly polynomially tractable*. In this case, the number of randomized samples is independent of d and depends polynomially on ε^{-1}. The smallest (or the infimum of) p in (3) is called the *exponent* of strong polynomial tractability.

Finally, we say that $I=\{I_d\}$ is *weakly tractable* iff

$$\lim_{\varepsilon^{-1}+d\to\infty}\frac{\ln\, n^{\rm ran}(\varepsilon,I_d)}{\varepsilon^{-1}+d}=0.$$

Weak tractability means that $n^{\rm ran}(\varepsilon,d)=\exp(o(\varepsilon^{-1}+d))$ is not exponential in $\varepsilon^{-1}+d$ but may increase to infinity faster than any polynomial in $\varepsilon^{-1}+d$.

We illustrate the concepts of this section for multivariate integration. We now need to assume that $H(K_1)$ contains integrable functions with respect to some probability density function $\varrho_1:D_1\to\mathbb{R}$, i.e., $\varrho_1\ge 0$ and $\int_{D_1}\varrho_1(x)\,{\rm d}x=1$. This requires to assume that

$$C_1^{\rm init}:=\left(\int_{D_1}\int_{D_1}K_1(x,y)\,\varrho_1(x)\,\varrho_1(y)\,{\rm d}x\,{\rm d}y\right)^{1/2}<\infty.$$

Without loss of generality we may choose D_1 and ϱ_1 such that there is no subset of D_1 with positive measure for which all functions from $H(K_1)$ vanish on it. Let

$${\rm INT}_1(f)=\int_{D_1}f(x)\,\varrho_1(x)\,{\rm d}x=\langle f,h_1\rangle_{H(K_1)}\quad\text{for all}\ \ f\in H(K_1),$$

where

$$h_1(x)=\int_{D_1}K_1(x,y)\,\varrho_1(y)\,{\rm d}y\quad\text{for all}\quad x\in D_1,$$

and

$$\|{\rm INT}_1\|=\|h_1\|_{H(K_1)}=C_1^{\rm init}.$$

For $d>1$, we have

$${\rm INT}_d(f)=\langle f,h_d\rangle_{H(K_d)}=\int_{D_d}f(x)\,\varrho_d(x)\,{\rm d}x\quad\text{for all}\quad f\in H(K_d),$$

where

$$h_d(x)=\prod_{j=1}^d h_1(x_j)\quad\text{and}\quad \varrho_d(x)=\prod_{j=1}^d\varrho_1(x_j)\quad\text{for all}\quad x\in D_d.$$

For $K_1 \ge 0$, Theorem 1 of Hinrichs states that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \le 1 + \pi/2 \cdot \varepsilon^{-2},$$

so that multivariate integration $I = \{I_d\}$ is strongly polynomially tractable with the exponent at most 2.

5. Decomposable Kernels

We present lower bounds on the minimal errors $e^{\mathrm{ran}}(n, I_d)$ for certain tensor product linear functionals I_d and tensor product spaces $H(K_d)$. From these bounds we will conclude that the exponent 2 of strong polynomial tractability of multivariate integration is sharp.

We proceed similarly as in [5], where the worst case setting was studied. Take first $d = 1$. We say that the kernel K_1 is *decomposable* iff there exists $a^* \in \mathbb{R}$ such that

$$K_1(x, y) \;=\; 0 \quad \text{for all} \quad x \le a^* \le y \;\; \text{and} \;\; x, y \in D_1. \tag{4}$$

For $a^* \in \mathbb{R}$, define

$$D_{(0)} = \{x \in D_1 :\; x \le a^*\} \quad \text{and} \quad D_{(1)} = \{x \in D_1 :\; x \ge a^*\}.$$

Obviously $D_1 = D_{(0)} \cup D_{(1)}$ and $D_{(0)} \cap D_{(1)} = \{a^*\}$ or $D_{(0)} \cap D_{(1)} = \emptyset$ depending on whether a^* belongs or does not belong to D_1. The essence of (4) is that the function K_1 may take nonzero values only if x and t belong to the same quadrant $D_{(0)} \times D_{(0)}$ or $D_{(1)} \times D_{(1)}$.

Observe that if K_1 is decomposable and $a^* \in D_1$ then $K_1(\cdot, a^*) = 0$. This implies that all functions in $H(K_1)$ vanish at a^* since $f(a^*) = \langle f, K_1(\cdot, a^*)\rangle_{H(K_1)} = 0$.

If K_1 is decomposable, then the space $H(K_1)$ can be decomposed as the direct sum of Hilbert spaces $H(K_1)_{(0)}$ and $H(K_1)_{(1)}$ of univariate functions defined by

$$H(K_1)_{(i)} \;=\; \overline{\mathrm{span}}\{K_1(\cdot, t) \;:\; t \in D_{(i)}\,\}$$

and equipped with the inner product of $H(K_1)$.

Indeed, functions of the form $f = \sum_{j=1}^k \beta_j K_1(\cdot, t_j)$ with real β_j and $t_j \in D_1$ are dense in $H(K_1)$. Then for all $t \in D_1$ we have

$$f(t) = \sum_{j=1}^k \beta_j K_1(t, t_j) = \sum_{(t,t_j)\in D_{(0)}^2} \beta_j K_1(t, t_j) + \sum_{(t,t_j)\in D_{(1)}^2} \beta_j K_1(t, t_j) = f_{(0)}(t) + f_{(1)}(t),$$

where $f_{(0)} \in H(K_1)_{(0)}$ and $f_{(1)} \in H(K_1)_{(1)}$. For $f \in H(K_1)_{(i)}$ we have $f(t) = 0$ for $t \in D_{(1-i)}$ and the subspaces $H(K_1)_{(0)}$ and $H(K_1)_{(1)}$ are orthogonal. Hence

$$\|f\|^2_{H(K_1)} = \|f_{(0)}\|^2_{H(K_1)} + \|f_{(1)}\|^2_{H(K_1)} \quad \text{for all} \quad f \in H(K_1).$$

Consider now $I_1(f) = \langle f, h_1\rangle_{H(K_1)}$ for all $f \in H(K_1)$. The function h_1 is from $H(K_1)$ and can be decomposed as

$$h_1 = h_{1,(0)} + h_{1,(1)}$$

where $h_{1,(i)} \in H(K_1)_{(i)}$ for $i = 1, 2$.

Take now arbitrary $d \geq 1$. Then

$$K_d(x,y) = \prod_{j=1}^{d} K_1(x_j, y_j) \quad \text{for all} \quad x, y \in D_d.$$

The continuous linear functional $I_d(f) = \langle f, h_d \rangle_{H(K_d)}$ corresponds to

$$h_d(x) = \prod_{j=1}^{d} h_1(x_j) = \prod_{j=1}^{d} \big(h_{1,(0)}(x_j) + h_{1,(1)}(x_j)\big)$$

and

$$\|I_d\| = \|h_d\|_{H(K_d)} = \left(\|h_{1,(0)}\|^2_{H(K_1)} + \|h_{1,(1)}\|^2_{H(K_1)}\right)^{d/2}.$$

We will apply a modification of Lemma 17.10 of Chapter 17 from [7], which in turn is a slight modification of Lemma 1 from [4] p. 63. For completeness we provide the proof of this lemma here.

Lemma 1. *Let $f_1, f_2, \ldots, f_N$ be such that*

- *$f_i \in H(K_d)$ and $\|f_i\|_{H(K_d)} = 1$ for all $i = 1, 2, \ldots, N$,*
- *the functions f_i have disjoint supports and satisfy $I_d(f_i) \geq \eta > 0$.*

Then for $n < N$ we have

$$e^{\mathrm{ran}}(n, I_d) \geq \left(1 - \frac{n}{N}\right)^{1/2} \eta.$$

Proof. We apply the idea of Bakhvalov [2] which states that the randomized setting is at least as hard as the average case setting for an arbitrary probability measure. For the average case setting, we select the set

$$M = \{\pm f_i \mid i = 1, 2, \ldots, N\}$$

with the uniform distribution so that each $\pm f_i$ occurs with probability $1/(2N)$. That is, the average case error of a deterministic algorithm A is now

$$e^{\mathrm{avg}}(A) = \left[\frac{1}{2N} \sum_{i=1}^{N} \left[(I_d(f_i) - A(f_i))^2 + (I_d(-f_i) - A(-f_i))^2\right]\right]^{1/2}.$$

Suppose first that A uses k function values, $k < N$. Then at least $N - k$ supports of f_i's are missed and for these functions $A(f_i) = A(-f_i)$. Then

$$(I_d(f_i) - A(f_i))^2 + (-I_d(f_i) - A(-f_i))^2 \geq 2I_d^2(f_i) \geq 2\eta^2,$$

and therefore

$$[e^{\mathrm{avg}}(A)]^2 \geq \frac{1}{2N}(N-k)2\eta^2 = \left(1 - \frac{k}{N}\right)\eta^2.$$

Next, let A use k function values with probability p_k such that $\sum_{k=1}^{\infty} p_k = 1$ and $\sum_{k=1}^{\infty} k p_k \leq n$. Then

$$[e^{\mathrm{avg}}(A)]^2 \geq \sum_{k=1}^{\infty} p_k \left(1 - \frac{k}{N}\right)\eta^2 = \left(1 - \frac{\sum_{k=1}^{\infty} k p_k}{N}\right)\eta^2 \geq \left(1 - \frac{n}{N}\right)\eta^2.$$

Since this holds for any deterministic algorithm using n function values on the average, we conclude that

$$\inf_A e^{\rm avg}(A) \ge \left(1-\frac{n}{N}\right)_+^{1/2} \eta. \tag{5}$$

Take now an arbitrary randomized algorithm A_{n,d,ω_d} that uses n function values on the average. The square of its error is

$$\begin{aligned}\left[e^{\rm ran}(A_n)\right]^2 &= \sup_{\|f\|_{H(K_d)}\le 1} \mathbb{E}_{\omega_d}(I_d(f)-A_{n,d,\omega}(f))^2 \\ &\ge \mathbb{E}_{\omega_d}\left(\frac{1}{2N}\sum_{i=1}^N \left[(I_d(f_i)-A_{n,d,\omega}(f_i))^2+(-I_d(f_i)-A_{n,d,\omega}(-f_i))^2\right]\right).\end{aligned}$$

Note that for a fixed ω, the algorithm $A_{n,d,\omega}$ is deterministic. The expression above between the brackets is then the square of the average case error of $A_{n,d,\omega}$ for which we can apply the lower bound (5). Therefore we have

$$\left[e^{\rm ran}(A_n)\right]^2 \ge \mathbb{E}_{\omega_d}\left(1-\frac{n}{N}\right)_+^2 \eta^2 = \left(1-\frac{n}{N}\right)_+^2 \eta^2.$$

This completes the proof. $\square$

We are ready to present a lower bound on the nth minimal error $e^{\rm ran}(n,I_d)$ which is the main result of this paper.

Theorem 2. *Assume that K_1 is decomposable and that $h_{1,(0)}$ and $h_{1,(1)}$ are non-zero. Denote*

$$\alpha = \frac{\max\left(\|h_{1,(0)}\|^2_{H(K_1)}, \|h_{1,(1)}\|^2_{H(K_1)}\right)}{\|h_{1,(0)}\|^2_{H(K_1)}+\|h_{1,(1)}\|^2_{H(K_1)}} \in \left[\tfrac12,1\right).$$

Then

$$e^{\rm ran}(n,I_d) \ge \left(\frac18\right)^{1/2}\frac{1}{\sqrt{n}}\, e^{\rm ran}(0,I_d) \quad \text{for all } n \text{ and } d \text{ such that} \quad 4n\alpha^d \le 1.$$

Hence

$$n^{\rm ran}(\varepsilon,I_d) \ge \left\lceil \frac18\left(\frac{1}{\varepsilon}\right)^2\right\rceil \quad \text{for all} \quad \varepsilon\in(0,1) \quad \text{and} \quad d \ge \frac{2\ln \varepsilon^{-1} - \ln 2}{\ln \alpha^{-1}}.$$

Proof. To apply Lemma 1 we need to construct functions f_i and estimate η. We proceed as follows. Let $[d] := \{1,2,\dots,d\}$. For the function h_d we have

$$h_d(x) = \prod_{j=1}^d (h_{1,(0)}(x_j)+h_{1,(1)}(x_j)) = \sum_{\mathfrak{u}\subseteq[d]} h_{\mathfrak{u}}(x) \tag{6}$$

with

$$h_{\mathfrak{u}}(x) = \prod_{j\in\mathfrak{u}} h_{1,(0)}(x_j) \prod_{j\notin\mathfrak{u}} h_{1,(1)}(x_j).$$

For $\mathfrak{u}=\emptyset$ or $\mathfrak{u}=[d]$, the product over the empty set is taken as 1.

The support of $h_{\mathfrak{u}}$ is

$$D_{\mathfrak{u}} := \{x \in D_d \mid \ x_j \in D_{(1)} \ \text{ for all } \ j \in \mathfrak{u} \quad \text{and} \quad x_j \in D_{(0)} \quad \text{for all} \quad j \notin \mathfrak{u}\}.$$

That is we identify 2^d elements $h_{\mathfrak{u}}$ with disjoint supports and

$$I_d(h_{\mathfrak{u}}) = \|h_{\mathfrak{u}}\|^2_{H(K_d)} = \|h_{1,(0)}\|^{2|\mathfrak{u}|}_{H(K_1)} \ \|h_{1,(1)}\|^{2(d-|\mathfrak{u}|)}_{H(K_1)}.$$

We now order $\{h_{\mathfrak{u}}\}$ according to their decreasing $I_d(h_{\mathfrak{u}})$. That is, let

$$\{g_j\}_{j=1,2,\dots,2^d} = \{h_{\mathfrak{u}}\}_{\mathfrak{u}\subseteq[d]}$$

such that $\|g_1\|_{H(K_d)} \ge \|g_2\|_{H(K_d)} \ge \cdots$. Let

$$p_j = \frac{\|g_j\|^2_{H(K_d)}}{\|h_d\|^2_{H(K_d)}} \quad \text{for} \ \ j = 1,2,\dots,2^d.$$

Clearly, $\sum_{j=1}^{2^d} p_j = 1$ and the largest p_1 is given by

$$p_1 = \alpha^d.$$

Define $k_0 = 0$ and integers $k_1, k_2, \dots, k_s \le 2^d$ such that for $i = 1,2,\dots,s$ we have

$$p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i-1} < \frac{1}{4n} \le p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i}.$$

Since $4n\alpha^d \le 1$ we have

$$p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i-1} + p_{k_i} < \frac{1}{4n} + \alpha^d \le \frac{1}{2n}.$$

This implies that

$$\frac{s}{4n} \le \sum_{j=1}^{k_s} p_j < \frac{s}{2n}.$$

Hence this construction is well defined at least for $s = 2n$.

Finally we apply Lemma 1 with $N = 2n$ and

$$f_j = \frac{\sum_{i=k_{j-1}+1}^{k_j} g_i}{\left\|\sum_{i=k_{j-1}+1}^{k_j} g_i\right\|_{H(K_d)}} = \frac{\sum_{i=k_{j-1}+1}^{k_j} g_i}{\left(\sum_{i=k_{j-1}+1}^{k_j} \|g_i\|^2_{H(K_d)}\right)^{1/2}}$$

for all $j = 1,2,\dots,N$.

Then f_j's have disjoint supports, $\|f_j\|_{H(K_d)} = 1$, and

$$\begin{aligned} I_d(f_j) &= \left(\sum_{i=k_{j-1}+1}^{k_j} \|g_j\|^2_{H(K_d)}\right)^{1/2} = \|h_d\|_{H(K_d)} \left(\sum_{i=k_{j-1}+1}^{k_j} p_j\right)^{1/2} \\ &\ge \eta := \|h_d\|_{H(K_d)} \ \frac{1}{2\sqrt{n}}. \end{aligned}$$

¿From Lemma 1 we conclude that

$$e^{\mathrm{ran}}(n, I_d) \ge \frac{1}{2\sqrt{2n}} \ e^{\mathrm{ran}}(0, I_d),$$

which completes the proof of the first inequality.

To prove the second inequality assume that

$$n < \lceil \varepsilon^{-2}/8 \rceil \quad \text{for} \quad \varepsilon \in (0,1) \quad \text{and} \quad d \geq \frac{2 \ln \varepsilon^{-1} \,-\, \ln 2}{\ln \alpha^{-1}}.$$

Then $n < \varepsilon^{-2}/8$ and $4n\alpha^d \leq 1$. Since $\varepsilon < 1/\sqrt{8n}$, the first inequality yields that

$$e^{\rm ran}(n, I_d) > \varepsilon \, e^{\rm ran}(0, I_d).$$

This means that $n^{\rm ran}(\varepsilon, I_d) > n$, and taking the largest such n we conclude that

$$n^{\rm ran}(\varepsilon, I_d) \geq \left\lceil \frac{1}{8\varepsilon^2} \right\rceil,$$

as claimed. This completes the proof of Theorem 2. $\square$

We stress that the lower estimate of $n^{\rm ran}(\varepsilon, I_d)$ in Theorem 2 holds for all $\varepsilon \in (0,1)$ and sufficiently large d with respect to ε. This has to be so since otherwise if we do not have a condition on the growth of d, then we could fix d and let ε tend to zero. The asymptotic behavior of $n^{\rm ran}(\varepsilon, I_d)$ depends on the smoothness of functions in $H(K_d)$ and may go to infinity much slower than ε^{-2}. In fact, in a moment we will see examples for which this happens. Therefore the lower bound in Theorem 2 cannot be true for all d, in general. On the other hand, it is interesting to note that the condition on d is quite mild since it requires that d grows only logarithmically with ε^{-1}.

Comparing Theorems 1 and 2 for multivariate integration defined over a tensor product Hilbert space we see quite similar lower and upper estimates on $n^{\rm ran}(\varepsilon, {\rm INT}_d)$ of order ε^{-2}. These estimates hold as long as the univariate reproducing kernel K_1 is pointwise nonnegative, integrable and decomposable. We now show two examples for which all these properties of K_1 hold.

Example 1: Multivariate Integration of Smooth Functions

As in Section 11.4.1 of [7], we consider multivariate integration for the Sobolev space of arbitrarily smooth functions. More precisely, let $r \in \mathbb{N}$. We take

$$H(K_1) = W_0^r(\mathbb{R})$$

as the Sobolev space of functions defined over $\mathbb{R}$ whose $(r-1)$st derivatives are absolutely continuous, with the rth derivatives belonging to $L_2(\mathbb{R})$ and their derivatives up to the $(r-1)$st at zero being zero. That is, we now have $D_1 = \mathbb{R}$ and

$$H(K_1) \,=\, \{f : \mathbb{R} \to \mathbb{R} \,:\, f^{(j)}(0) = 0,\, j \in [0, r{-}1],\, f^{(r-1)} \text{ abs. cont. and } f^{(r)} \in L_2(\mathbb{R})\}.$$

The inner product of F_1 is given as

$$\langle f, g \rangle_{F_1} \,=\, \int_{\mathbb{R}} f^{(r)}(t) g^{(r)}(t) \, {\rm d}t.$$

It is known, and not hard to check, that this Hilbert space has the reproducing kernel

$$K_1(x,t) \,=\, 1_M(x,t) \int_0^\infty \frac{(|t| - u)_+^{r-1} \, (|x| - u)_+^{r-1}}{[(r-1)!]^2} \, {\rm d}u,$$

where 1_M is the characteristic (indicator) function of the set $M = \{(x,t) :\ xt \geq 0\}$. For $r = 1$, we have

$$K_1(x,t) \;=\; 1_M(x,t)\, \min(|t|,|x|).$$

For $r \geq 1$, observe that this kernel is *decomposable* at $a^* = 0$ since

$$K_1(x,t) = 0 \quad \text{for all} \quad x \leq 0 \leq t.$$

The kernel K_1 is also symmetric since $K_1(x,t) = K_1(-x,-t)$., and obviously

$$K_1(x,t) \geq 0 \quad \text{for all} \quad x,t \in D_1.$$

For $d > 1$, we take tensor products and

$$H(K_d) \;=\; W_0^{r,r,\dots,r}(\mathbb{R}^d) \;=\; W_0^r(\mathbb{R}) \otimes \cdots \otimes W_0^r(\mathbb{R})$$

is the d-fold tensor product of $W_0^r(\mathbb{R})$. Hence, $H(K_d)$ is the Sobolev space of smooth functions defined over $D_d = \mathbb{R}^d$ such that $D^\alpha f(x) = 0$ if at least one component of x is zero for any multi-index $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_d]$ with integers $\alpha_j \in [0, r-1]$. Here, D^α is the partial differential operator, $D^\alpha f = \partial^{|\alpha|} f / \partial^{\alpha_1} x_1 \cdots \partial^{\alpha_d} x_d$. The inner product of $H(K_d)$ is given by

$$\langle f, g \rangle_{H(K_d)} \;=\; \int_{\mathbb{R}^d} D^{[r,r,\dots,r]} f(x)\, D^{[r,r,\dots,r]} g(x)\, \mathrm{d}x.$$

Obviously,

$$K_d(x,t) = \prod_{j=1}^{d} K_1(x_j,t_j) \geq 0 \quad \text{for all} \quad x,t \in D_d.$$

For $d = 1$, consider univariate integration

$$\mathrm{INT}_1(f) \;=\; \int_{\mathbb{R}} f(t)\, \varrho(t)\, \mathrm{d}t \quad \text{for all} \quad f \in H(K_1)$$

for some measurable non-zero weight function $\varrho : \mathbb{R} \to \mathbb{R}_+$. It is easy to check that INT_1 is a continuous linear functional iff the function

$$h_1(x) \;=\; \int_{\mathbb{R}} K_1(x,t)\, \varrho(t)\, \mathrm{d}t$$

belongs to $H(K_1)$, which holds iff

$$\int_{\mathbb{R}^2} K_1(x,t)\, \varrho(t)\, \varrho(x)\, \mathrm{d}t\, \mathrm{d}x \;<\; \infty. \tag{7}$$

It is also easy to check that $K_1(x,t) = \mathcal{O}(|t\,x|^{r-1/2})$, and (7) holds if

$$\int_{\mathbb{R}} \varrho(t)\, |t|^{r-1/2}\, \mathrm{d}t \;<\; \infty.$$

The last condition imposes a restriction on the behavior of the weight ϱ at infinity. If (7) holds, then

$$\mathrm{INT}_1(f) \;=\; \langle f, h_1 \rangle_{H(K_1)} \quad \text{for all} \quad f \in H(K_1),$$

and

$$\|\mathrm{INT}_1\| \;=\; \|h_1\|_{H(K_1)} \;=\; \left(\int_{\mathbb{R}^2} K_1(x,t)\,\varrho(t)\,\varrho(x)\,\mathrm{d}t\,\mathrm{d}x\right)^{1/2} \;<\; \infty.$$

We also have

$$h_{1,(0)}(x) \;=\; \int_{-\infty}^{0} K_1(x,t)\,\varrho(t)\,\mathrm{d}t \quad \text{and} \quad h_{1,(1)}(x) \;=\; \int_{0}^{\infty} K_1(x,t)\,\varrho(t)\,\mathrm{d}t.$$

Furthermore,

$$\begin{aligned}\|h_{1,(0)}\|^2_{H(K_1)} &= \int_{-\infty}^{0}\int_{-\infty}^{0} K_1(x,t)\,\varrho(t)\,\varrho(x)\,\mathrm{d}t\,\mathrm{d}x,\\ \|h_{1,(1)}\|^2_{H(K_1)} &= \int_{0}^{\infty}\int_{0}^{\infty} K_1(x,t)\,\varrho(t)\,\varrho(x)\,\mathrm{d}t\,\mathrm{d}x.\end{aligned}$$

For $d>1$, we have

$$\mathrm{INT}_d(f) \;=\; \int_{\mathbb{R}^d} f(t)\,\varrho_d(t)\,\mathrm{d}t \quad \text{with} \quad \varrho_d(t) \;=\; \varrho(t_1)\varrho(t_2)\cdots\varrho(t_d).$$

We are ready to apply Theorems 1 and 2 for this multivariate integration problem. All the assumptions of Theorem 1 of Hinrichs are satisfied. To apply Theorem 2, note that if the weight ϱ does not vanish (in the L_2 sense) over $\mathbb{R}_-$ and $\mathbb{R}_+$ then the norms of $h_{1,(0)}$ and $h_{1,(1)}$ are positive and

$$\alpha = \frac{\max\left(\|h_{1,(0)}\|^2_{H(K_1)}, \|h_{1,(1)}\|^2_{H(K_1)}\right)}{\|h_{1,(0)}\|^2_{H(K_1)} + \|h_{1,(0)}\|^2_{H(K_1)}} < 1.$$

Furthermore, if we take a nonzero symmetric ϱ, i.e., $\varrho(t)=\varrho(-t)$, then $\alpha=\frac{1}{2}$. This is the case for *Gaussian integration* for which

$$\varrho(t) = (2\pi\,\sigma)^{-1/2}\exp\left(-\,t^2/(2\,\sigma)\right) \quad \text{for all} \quad t\in\mathbb{R}$$

is symmetric. Here, the variance σ is an arbitrary positive number.

Hence, multivariate integration is strongly polynomially tractable with the exponent 2. We stress that the exponent is independent of the assumed smoothness of functions measured by r. More specifically we have the following bounds

$$\begin{aligned} n^{\mathrm{ran}}(\varepsilon,\mathrm{INT}_d) &\le \left\lceil \frac{\pi}{2}\left(\frac{1}{\varepsilon}\right)^2\right\rceil \quad \text{for all} \quad \varepsilon\in(0,1) \ \text{ and } \ d\in\mathbb{N},\\ n^{\mathrm{ran}}(\varepsilon,\mathrm{INT}_d) &\ge \left\lceil \frac{1}{8}\left(\frac{1}{\varepsilon}\right)^2\right\rceil \quad \text{for all} \quad \varepsilon\in(0,1) \ \text{ and } \ d\ge\frac{2\,\ln\,\varepsilon^{-1}-\ln\,2}{\ln\,\alpha^{-1}}.\end{aligned}$$

We add in passing that this problem was also studied in the worst case setting. If we denote $n^{\mathrm{wor}}(\varepsilon,\mathrm{INT}_d)$ as the minimal number of function values needed to reduce the initial error by a factor ε in the worst case setting then

$$n^{\mathrm{wor}}(\varepsilon,\mathrm{INT}_d) \ge (1-\varepsilon^2)\,\alpha^{-d} \quad \text{for all} \quad \varepsilon\in(0,1) \ \text{ and } \ d\in\mathbb{N},$$

see Theorem 11.8 in [7]. Hence, we have intractability and the curse of dimensionality. This means that the randomized setting allows us to vanquish the curse of dimensionality of this multivariate problem in the worst case setting.

We now briefly discuss the asymptotic behavior of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ for a fixed d and ε tending to zero. For simplicity we take the weight $\varrho(t) = \frac{1}{2}$ for $t \in [-1, 1]$ and $\varrho(t) = 0$ for $|t| > 1$. Then $\alpha = 1/2$ and $\int_{\mathbb{R}} \varrho(t)\,\mathrm{d}t = 1$. For $d = 1$ it is known that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_1) = \Theta\left(\varepsilon^{-1/(r+1/2)}\right) \quad \text{as} \quad \varepsilon \to 0. \tag{8}$$

For $d \geq 2$, we can achieve almost the same dependence modulo some powers of $\ln \varepsilon^{-1}$. More precisely, we first approximate functions from $H(K_d)$ in the worst case setting for the L_2 norm by algorithms using arbitrary linear functionals. Then the minimal worst case error of algorithms that use n such linear functionals is

$$\Theta\left(n^{-r}\ (\ln n)^{(d-1)r}\right).$$

It is also know that in the randomized setting we can approximate functions from $H(K_d)$ by linear algorithms using function values with the error which is modulo a double log the same as the worst case error for arbitrary linear functionals, see [10]. That is, f is approximated by $f_n = \sum_{j=1}^{n} a_{j,\omega_d} f(x_{j,\omega_d})$ with the error for the L_2 norm

$$\mathcal{O}\left(n^{-r}\ (\ln n)^{(d-1)r}\ (\ln \ln n)^{r+1/2}\right).$$

Finally, since

$$\mathrm{INT}_d(f) = \mathrm{INT}_d(f_n) + \mathrm{INT}_d(f - f_n)$$

we approximate the integral of f by adding to $\mathrm{INT}_d(f_n)$ the standard Monte Carlo approximation of the integral of $f - f_n$, and obtain the error

$$\mathcal{O}\left(n^{-(r+1/2)}\ (\ln n)^{(d-1)r}\ (\ln \ln n)^{r+1/2}\right).$$

This implies that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) = \mathcal{O}\left(\varepsilon^{-2/(1+2r)}\ \left(\ln \varepsilon^{-1}\right)^{(d-1)r/(r+1/2)}\ \ln \ln \varepsilon^{-1}\right) \quad \text{as} \quad \varepsilon \to 0.$$

From (8) we conclude that modulo logarithms the last bound is sharp.

We stress that the factor in the big $\mathcal{O}$ notation depends on d and r. In any case, the leading factor $\varepsilon^{-2/(1+2r)}$ is always less 2, and for large r is quite small. Hence, asymptotically in ε and for fixed d, we have a much better behavior than ε^{-2} that is achieved if d varies with ε^{-1}.

Example 2: Centered Discrepancy and Midpoint Conditions

We now consider multivariate integration whose worst case error is closely related to the centered discrepancy, see Section 11.4.3 of [7]. In fact, we have two such multivariate problems defined on specific Sobolev spaces with or without midpoint conditions. Here we discuss the case with midpoint conditions and later we will address the case without midpoint conditions.

Take now $D_1 = [0,1]$ and $H(K_1) = W^1_{1/2}([0,1])$ as the Sobolev space of absolutely continuous functions whose first derivatives are in $L_2([0,1])$ and whose function values are zero at $1/2$. We call $f(\frac{1}{2}) = 0$ the *midpoint condition*. That is,

$$H(K_1) = \{f : [0,1] \to \mathbb{R} \ : f(\tfrac{1}{2}) = 0, \ f \text{ abs. cont. and } f' \in L_2([0,1])\}$$

with the inner product

$$\langle f, g \rangle_{H(K_1)} = \int_0^1 f'(t) g'(t) \, \mathrm{d}t.$$

The reproducing kernel is

$$K_1(x,t) = \tfrac{1}{2}(|x - \tfrac{1}{2}| + |t - \tfrac{1}{2}| - |x - t|),$$

which can be rewritten as

$$K_1(x,t) = 1_M(x,t) \cdot \min(|x - \tfrac{1}{2}|, |t - \tfrac{1}{2}|),$$

where $M = [0, \frac{1}{2}] \times [0, \frac{1}{2}] \cup [\frac{1}{2}, 1] \times [\frac{1}{2}, 1]$, and 1_M denotes the characteristic function of M, i.e., $1_M(y) = 1$ if $y \in M$ and $1_M(y) = 0$ is $y \notin M$. Hence, the kernel K_1 is *decomposable* at $a^* = \frac{1}{2}$, symmetric and clearly $K_1 \ge 0$.

For $d > 1$, we take tensor products and obtain

$$H(K_d) = W^{1,1,\dots,1}_{1/2}([0,1]^d) \ = \ W^1_{1/2}([0,1]) \otimes \cdots \otimes W^1_{1/2}([0,1]), \quad d \text{ times},$$

as the Sobolev space of smooth functions f defined over $D_d = [0,1]^d$ such that $f(x) = 0$ if at least one component of x is $1/2$. They are called the *midpoint conditions*. The inner product of $H(K_d)$ is given by

$$\langle f, g \rangle_{H(K_d)} \ = \ \int_{[0,1]^d} \frac{\partial^d}{\partial x_1 \cdots \partial x_d} f(x) \, \frac{\partial^d}{\partial x_1 \cdots \partial x_d} g(x) \, \mathrm{d}x.$$

Consider the *uniform integration* problem,

$$I_1(f) \ = \ \int_0^1 f(t) \, \mathrm{d}t.$$

It is easy to compute

$$\begin{aligned} h_{1,(0)}(x) \ &= \ \int_0^{1/2} \min(\tfrac{1}{2} - x, \tfrac{1}{2} - t) \, \mathrm{d}t \ &= \ \tfrac{1}{2}\,(\tfrac{1}{2} - x)(\tfrac{1}{2} + x) \quad &\forall\, x \in [0, \tfrac{1}{2}], \\ h_{1,(1)}(x) \ &= \ \int_a^1 \min(x - \tfrac{1}{2}, t - \tfrac{1}{2}) \, \mathrm{d}t \ &= \ \tfrac{1}{2}\,(x - \tfrac{1}{2})(\tfrac{3}{2} - x) \quad &\forall\, x \in [\tfrac{1}{2}, 1]. \end{aligned}$$

Furthermore,

$$\|h_{1,(0)}\|^2_{H(K_1)} = \|h_{1,(1)}\|^2_{H(K_1)} = \tfrac{1}{24} \quad \text{and} \quad \alpha = \tfrac{1}{2}.$$

This means that we can apply Theorems 1 and 2 and obtain strong polynomial tractability with the exponent 2. More specifically we have

$$\begin{aligned} n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) &\le \left\lceil \frac{\pi}{2}\left(\frac{1}{\varepsilon}\right)^2 \right\rceil \quad \text{for all} \quad \varepsilon \in (0,1) \ \text{ and } \ d \in \mathbb{N}, \\ n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) &\ge \left\lceil \frac{1}{8}\left(\frac{1}{\varepsilon}\right)^2 \right\rceil \quad \text{for all} \quad \varepsilon \in (0,1) \ \text{ and } \ d \ge \frac{2\,\ln \varepsilon^{-1} - \ln 2}{\ln 2}. \end{aligned}$$

In the worst case setting we have

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d) \ge (1-\varepsilon^2)\, 2^d \quad \text{for all} \quad \varepsilon \in (0,1) \ \text{ and } \ d \in \mathbb{N},$$

see Theorem 11.8 of [7]. Furthermore, the worst case error of a linear algorithm

$$Q_{n,d}(f) \;=\; \sum_{j=1}^{n} a_j f(z_j)$$

is given by

$$e^{\mathrm{wor}}(Q_{n,d}) = \left(\int_{[0,1]^d} \left| \prod_{j=1}^{d} \min(x_j, 1-x_j) - \sum_{j=1}^{n} a_j \cdot 1_{J(b(x),x)}(z_j) \right|^2 \mathrm{d}x \right)^{1/2},$$

where $J(b(x),x)$ is the rectangular box generated by x and the vertex $b(x)$ of $[0,1]^d$ that is closest to x in the sup-norm. The last integral is the centered discrepancy of the points z_j and the coefficients a_j. This explains in what sense this integration problem is related to the centered discrepancy.

As in the previous example, the curse of dimensionality present in the worst case setting is vanquished in the randomized setting. Also as before we can basically repeat the reasoning on the asymptotic behavior of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ for a fixed d and ε tending to zero, and show that the current case is a variant of the previous case for $r = 1$ and a special $\varrho = 1$ over $[-1,1]$. Therefore we have

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) = \mathcal{O}\left(\varepsilon^{-2/3} \, \left(\ln \varepsilon^{-1}\right)^{2(d-1)/3} \, \ln\ln \varepsilon^{-1} \right) \quad \text{as} \quad \varepsilon \to 0 \tag{9}$$

with the factor in the big $\mathcal{O}$ notation depending on d. Again, modulo logarithms the last bound is sharp. This means that we must have ε^{-2} instead of $\varepsilon^{-2/3}$ if we want to have bounds independent of d.

6. Non-decomposable Kernels

In this final section we briefly discuss tensor product functionals defined over Hilbert spaces with non-decomposable kernels. We indicate how to get a lower bound for such problems. However, the lower and upper bounds are not sharp as before and we think that there is much more work needed to get better bounds.

Our approach is parallel to the approach we took for the worst case setting in [5]. Unfortunately for the randomized setting the situation is much more complicated and it is not clear if some properties of tensor product functionals that were crucial for lower bounds techniques in the worst case setting are also true in the randomized setting. We will be more specific on this point later after we present a lower bound.

As before, we first consider $d = 1$, and assume that

$$K_1 = R_1 + R_2 \tag{10}$$

for some reproducing kernels R_1 and R_2 such that the corresponding Hilbert spaces $H(R_1)$ and $H(R_2)$ satisfy

$$H(R_1) \cap H(R_2) = \{0\} \quad \text{and} \quad \text{the kernel } R_2 \text{ is } \textit{decomposable}. \tag{11}$$

For many standard spaces with non-decomposable kernels K_1 we can take R_1 such that $H(R_1)$ is a finite dimensional space. For example, let $K_1(x,t) = 1 + \min(x,t)$ for $x,t \in [0,1]$. For $a \in (0,1)$ we take $R_2 = K_1 - R_1$ with

$$R_1(x,t) = \frac{(1+\min(x,a))(1+\min(t,a))}{1+a} \quad \text{for all} \quad x,t \in [0,1].$$

Then $H(R_1) = \text{span}(1+\min(\cdot,a)\}$ is one-dimensional, $H(R_2) = \{f \in H(K_1) \mid f(a) = 0\}$ and $H(R_1) \cap H(R_2) = \{0\}$. For $x \le a \le t$ we have

$$R_2(x,t) = 1 + x - \frac{(1+x)(1+a)}{1+a} = 0,$$

so that R_2 is decomposable at a.

Due to (10) we have a unique decomposition for $f \in H(K_1)$,

$$f = f_1 + f_2 \quad \text{with} \quad f_i \in H(R_i),\ i = 1,2.$$

Furthermore, for $f,g \in H(K_1)$ we have

$$\langle f,g\rangle_{H(K_1)} = \langle f_1,g_1\rangle_{H(R_1)} + \langle f_2,g_2\rangle_{H(R_2)}\,.$$

This implies that all $f \in H(K_1)$ can be uniquely represented as

$$f = f_1 + f_{2,(0)} + f_{2,(1)} \quad \text{with} \quad f_1 \in H(R_1),\ f_{2,(0)} \in H(R_2)_{(0)},\ f_{2,(1)} \in H(R_2)_{(1)},$$

and

$$\|f\|^2_{H(K_1)} = \|f_1\|^2_{H(R_1)} + \|f_{2,(0)}\|^2_{H(R_2)} + \|f_{2(1)}\|^2_{H(R_2)}.$$

For $I_1(f) = \langle f,h_1\rangle_{H(K_1)}$ for all $f \in H(K_1)$ and some $h_1 \in H(K_1)$, we have

$$h_1 = h_{1,1} + h_{1,2,(0)} + h_{1,2,(1)}$$

and

$$\|h_1\|^2_{H(K_1)} = \|h_{1,1}\|^2_{H(R_1)} + \|h_{1,2,(0)}\|^2_{H(R_2)} + \|h_{1,2,(1)}\|^2_{H(R_2)}.$$

For $d > 1$, we take tensor products and obtain $H(K_d)$ with

$$K_d(x,t) = \prod_{j=1}^{d} \left(R_1(x_j,t_j) + R_2(x_j,t_j)\right) = \sum_{\mathfrak{u}\subseteq[d]} R_{\overline{\mathfrak{u}},1}(x_{\overline{\mathfrak{u}}},t_{\overline{\mathfrak{u}}})\, R_{\mathfrak{u},2}(x_{\mathfrak{u}},t_{\mathfrak{u}}),$$

where

$$R_{\overline{\mathfrak{u}},1}(x_{\overline{\mathfrak{u}}},t_{\overline{\mathfrak{u}}}) \;=\; \prod_{j\notin\mathfrak{u}} R_1(x_j,t_j) \quad \text{and} \quad R_{\mathfrak{u},2}(x_{\mathfrak{u}},t_{\mathfrak{u}}) = \prod_{j\in\mathfrak{u}} R_2(x_j,t_j)$$

are the reproducing kernels of the Hilbert spaces $H(R_{\overline{\mathfrak{u}},1})$ and $H(R_{\mathfrak{u},2})$.

For $I_d(f) = \langle f, h_d \rangle_{H(K_d)}$ for all $f \in H(K_d)$, we have

$$h_d(x) = \prod_{j=1}^{d} h_1(x_j) = \prod_{j=1}^{d} (h_{1,1}(x_j) + h_{1,2}(x_j)) = \sum_{\mathfrak{u} \subseteq [d]} h_{\overline{\mathfrak{u}},1}(x_{\overline{\mathfrak{u}}})\, h_{\mathfrak{u},2}(x_{\mathfrak{u}}),$$

where

$$h_{\overline{\mathfrak{u}},1}(x_{\overline{\mathfrak{u}}}) = \prod_{j \notin \mathfrak{u}} h_{1,1}(x_j), \quad \text{and} \quad h_{\mathfrak{u},2}(x_{\mathfrak{u}}) = \prod_{j \in \mathfrak{u}} h_{1,2}(x_j).$$

Then $h_{\overline{\mathfrak{u}},1} \in H(R_{\overline{\mathfrak{u}},1})$ and $h_{\mathfrak{u},2} \in H(R_{\mathfrak{u},2})$. For $\mathfrak{u} = \emptyset$ or $\mathfrak{u} = [d]$, we take $h_{\emptyset,2} = 1$ and $h_{[d],1} = 1$. We also have

$$\|h_{\overline{\mathfrak{u}},1}\|_{H(R_{\overline{\mathfrak{u}},1})} = \|h_{1,1}\|_{H(R_1)}^{d-|\mathfrak{u}|} \quad \text{and} \quad \|h_{\mathfrak{u},2}\|_{H(R_{\mathfrak{u},2})} = \|h_{1,2}\|_{H(R_2)}^{|\mathfrak{u}|}.$$

Obviously, $h_{1,2} = h_{1,2,(0)} + h_{1,2,(1)}$ and

$$\|h_{1,2}\|_{H(R_2)}^2 = \|h_{1,2,(0)}\|_{H(R_2)}^2 + \|h_{1,2,(1)}\|_{H(R_2)}^2.$$

We are ready to present our lower bound.

Theorem 3. *Assume that* (11) *holds. Let*

$$\|h_{1,2,(0)}\|_{H(R_2)}^2 > 0 \quad \textit{and} \quad \|h_{1,2,(1)}\|_{H(R_2)}^2 > 0$$

so that

$$\alpha = \frac{\max(\|h_{1,2,(0)}\|_{H(R_2)}^2, \|h_{1,2,(1)}\|_{H(R_2)}^2)}{\|h_{1,2,(0)}\|_{H(R_2)}^2 + \|h_{1,2,(1)}\|_{H(R_2)}^2} \in \left[\tfrac{1}{2}, 1\right).$$

Let

$$\beta = \frac{\|h_{1,1}\|_{H(R_1)}^2}{\|h_{1,1}\|_{H(R_1)}^2 + \|h_{1,2,(0)}\|_{H(R_2)}^2 + \|h_{1,2,(1)}\|_{H(R_2)}^2} \in [0,1).$$

Then

$$e^{\rm ran}(n, I_d) \ge \frac{(1-\beta)^{1/2}}{4^{\gamma_{\alpha,\beta}+3/4}} \, \frac{1}{n^{\gamma_{\alpha,\beta}+1/2}} \, e^{\rm ran}(0, I_d) \quad \textit{with} \quad \gamma_{\alpha,\beta} = \frac{\ln 1/(1-\beta)}{2 \ln \alpha^{-1}}$$

for all n and d such that $4n\alpha^d \le 1$. Hence

$$n^{\rm ran}(\varepsilon, I_d) \ge \left\lceil \frac{1}{4} \left(\frac{1-\beta}{2}\right)^{1/(1+2\gamma_{\alpha,\beta})} \left(\frac{1}{\varepsilon}\right)^{2/(1+2\gamma_{\alpha,\beta})} \right\rceil$$

for all $\varepsilon \in (0,1)$ and

$$d \ge \frac{\frac{2}{1+2\gamma_{\alpha,\beta}} \ln \varepsilon^{-1} + \frac{1}{1+2\gamma_{\alpha,\beta}} \ln \frac{1-\beta}{2}}{\ln \alpha^{-1}}.$$

Proof. Let $k = \lceil \ln(4n)/\ln(\alpha^{-1}) \rceil$ so that $\alpha^{-k} \ge 4n$. It is easy to check that $4n\alpha^d \le 1$ implies that $k \le d$. Indeed, since $\alpha^{-k+1} < 4n$ we have $4n = \alpha^{-k}(1-x)$ for $x \in [0, 1-\alpha)$. Then

$$\alpha^{d-k}(1-x) = 4n\alpha^d \le 1$$

and this implies that $k \le d$, as claimed.

Consider 2^k pairs $(\mathfrak{v}_j, \mathfrak{w}_j)$ such that

$$\mathfrak{v}_j \cap \mathfrak{w}_j = \emptyset \quad \text{and} \quad \mathfrak{v}_j \cup \mathfrak{w}_j = [k] = \{1, 2, \ldots, k\}.$$

Consider the function

$$g_j(x) = \prod_{j\in\mathfrak{v}_j} h_{1,2,(0)}(x_j) \prod_{j\in\mathfrak{w}_j} h_{1,2,(1)}(x_j) \prod_{j\in[d]\setminus[k]} h_1(x_j).$$

Clearly,

$$\begin{aligned} I_d(g_j) &= \langle g_j, h_d\rangle_{H(K_d)} = \|g_j\|^2_{H(K_d)} \\ &= \|h_{1,2,(0)}\|^{2|\mathfrak{v}_j|}_{H(R_2)} \|h_{1,2,(1)}\|^{2|\mathfrak{w}_j|}_{H(R_2)} \|h_1\|^{2(d-k)}_{H(K_1)}. \end{aligned}$$

The support of g_j is included in the set

$$D_j = \{x \in D_d \mid \ x_j \in D_{(0)} \ \text{ for all } j \in \mathfrak{v}_j \ \text{ and } \ x_j \in D_{(1)} \ \text{ for all } j \in \mathfrak{w}_j\}.$$

Therefore the functions g_j for $j = 1, 2, \dots, 2^k$ have disjoint support.

We now basically repeat a part of the proof of Theorem 2. More precisely, we define

$$p_j = \frac{\|g_j\|^2_{H(K_d)}}{\|h_d\|^2_{H(K_d)}} \quad \text{for all} \quad j = 1, 2, \dots, 2^k.$$

We now have

$$\sum_{j=1}^{2^k} p_j = \left(\|h_{1,2,(0)}\|^2_{H(R_2)} + \|h_{1,2,(1)}\|^2_{H(R_2)}\right)^k \ \|h_1\|^{-2k}_{H(K_1)} = (1-\beta)^k.$$

Furthermore, it is easy to check that

$$\max_{j=1,2,\dots,2^k} p_j = \alpha^k (1-\beta)^k.$$

As in the proof of Theorem 2, we define $k_0 = 0$ and $k_i \le 2^k$ for $i = 1, 2, \dots, s$ such that

$$p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i-1} < \frac{(1-\beta)^k}{4n} \le p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i}.$$

Then we check as before that $4n\alpha^k \le 1$ implies

$$p_{k_{i-1}+1} + p_{k_{i-1}+2} + \cdots + p_{k_i} \le \frac{(1-\beta)^k}{4n} + p_{k_i} \le \frac{(1-\beta)^k}{4n} + \alpha^k (1-\beta)^k \le \frac{(1-\beta)^k}{2n},$$

so that the construction is well defined for $s = 2n$.

We are ready to apply Lemma 1 with $N = 2n$ and

$$f_j = \frac{\sum_{i=k_{j-1}+1}^{k_j} g_i}{\left\|\sum_{i=k_{j-1}+1}^{k_j} g_i\right\|_{H(K_d)}} = \frac{\sum_{i=k_{j-1}+1}^{k_j} g_i}{\left(\sum_{i=k_{j-1}+1}^{k_j} \|g_i\|^2_{H(K_d)}\right)^{1/2}}$$

for all $j = 1, 2, \dots, N$.

Then the f_j have disjoint supports, $\|f_j\|_{H(K_d)} = 1$, and

$$\begin{aligned} I_d(f_j) &= \left(\sum_{i=k_{j-1}+1}^{k_j} \|g_j\|^2_{H(K_d)} \right)^{1/2} = \|h_d\|_{H(K_d)} \left(\sum_{i=k_{j-1}+1}^{k_j} p_j \right)^{1/2} \\ &\geq \eta := \|h_d\|_{H(K_d)} \, \frac{(1-\beta)^{k/2}}{2\sqrt{n}}. \end{aligned}$$

¿From Lemma 1 we conclude that

$$e^{\mathrm{ran}}(n, I_d) \geq \frac{(1-\beta)^{k/2}}{2\sqrt{2n}} \, e^{\mathrm{ran}}(0, I_d).$$

We estimate

$$\begin{aligned} (1-\beta)^{k/2} &\geq (1-\beta)^{1/2} \exp\left((\ln 4n) \, \frac{\ln(1-\beta)}{2 \ln \alpha^{-1}} \right) \\ &= (1-\beta)^{1/2} \, (4n)^{-\gamma_{\alpha,\beta}} \end{aligned}$$

which completes the proof of the first inequality.

To prove the second inequality assume that

$$n < \left\lceil \frac{1}{4} \left(\frac{1-\beta}{2} \right)^{1/(1+2\gamma_{\alpha,\beta})} \left(\frac{1}{\varepsilon} \right)^{2/(1+2\gamma_{\alpha,\beta})} \right\rceil .$$

Then

$$4n < \left(\frac{1-\beta}{2} \right)^{1/(1+2\gamma_{\alpha,\beta})} \left(\frac{1}{\varepsilon} \right)^{2/(1+2\gamma_{\alpha,\beta})}$$

and $4n\alpha^d \leq 1$. Since

$$\varepsilon < \left(\frac{1-\beta}{2} \right)^{1/2} \frac{1}{(4n)^{\gamma_{\alpha,\beta}+1/2}} = \frac{(1-\beta)^{1/2}}{4^{\gamma_{\alpha,\beta}+3/4}} \, \frac{1}{n^{\gamma_{\alpha,\beta}+1/2}},$$

the first inequality yields that

$$e^{\mathrm{ran}}(n, I_d) > \varepsilon \, e^{\mathrm{ran}}(0, I_d).$$

This means that $n^{\mathrm{ran}}(\varepsilon, I_d) > n$, and taking the largest such n we conclude the second inequality, as claimed. This completes the proof of Theorem 2. □

We comment on Theorem 3. First of all note that for $h_{1,1} = 0$ we have $\beta = 0$ and the estimates of Theorem 3 are exactly the same as the estimates of Theorem 2. This means that Theorem 3 generalizes Theorem 2 for non-decomposable kernels and linear tensor product functions with the zero function $h_{1,1}$.

Assume that $\beta > 0$. Then the lower bound on $e^{\mathrm{ran}}(n, I_d)/e^{\mathrm{ran}}(0, I_d)$ is roughly $n^{-(\gamma_{\alpha,\beta}+1/2)}$ which is smaller than the bound $n^{-1/2}$ obtained before since $\gamma_{\alpha,\beta} > 0$. Of course, this results in the lower bound on $n^{\mathrm{ran}}(\varepsilon, I_d)$ roughly $\varepsilon^{-2/(1+2\gamma_{\alpha,\beta})}$, again smaller than the bound ε^{-2} before. If we assume additionally that the reproducing kernel is nonnegative, then Theorem 1 of Hinrichs for multivariate integration says that $e^{\mathrm{ran}}(n, I_d)/e^{\mathrm{ran}}(0, I_d) = \mathcal{O}(n^{-1/2})$ and $n^{\mathrm{ran}}(\varepsilon, I_d) = \mathcal{O}(\varepsilon^{-2})$. This means that there is a gap between the lower and upper bounds. We do not know whether the lower or upper bounds can be improved. Of course, for small β relative to α the bounds are pretty tight. However if β is close to 1, the exponent of n^{-1} is large, and the exponent of ε^{-1} small. The same also holds if α is close to 1 and β is not too close to zero. In this case the lower bound cannot be sharp since even the asymptotic bounds yield better estimates of the exponents since asymptotically the exponents of n^{-1} and ε^{-1} do not depend on β. This will be illustrated by the following example.

Example 3: Centered Discrepancy and no Midpoint Conditions

We now remove the midpoint conditions by taking the reproducing kernel

$$K_d(x,t)=\prod_{j=1}^d\left[b+1_M(x_j,t_j)\cdot\min(|x_j-\tfrac12|,|t_j-\tfrac12|)\right]\quad\text{for all}\quad x,t\in[0,1],$$

where, as before, 1_M is the characteristic function of $M=[0,\frac12]\times[0,\frac12]\cup[\frac12,1]\times[\frac12,1]$. Here b is a non-negative number. For $b=0$ we have the case studied before where we assume the midpoint condition, i.e., $f(x)=0$ if at least one component of x is $\frac12$. For $b>0$, there are no midpoint conditions and the inner product of $H(K_d)$ is given by

$$\langle f,g\rangle_{H(K_d)}=\frac{1}{b^d}\,f\left(\vec{\tfrac12}\right)\,g\left(\vec{\tfrac12}\right)+\sum_{\emptyset\neq\mathfrak{u}\subseteq[d]}\frac{1}{b^{d-|\mathfrak{u}|}}\int_{[0,1]^{|\mathfrak{u}|}}\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}}f\left(x_{\mathfrak{u}},\vec{\tfrac12}\right)\,\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}}g\left(x_{\mathfrak{u}},\vec{\tfrac12}\right)\,\mathrm{d}\,x_{\mathfrak{u}},$$

where $y=(x_{\mathfrak{u}},\vec{\frac12})$ is the vector for which $y_j=x_j$ for $j\in\mathfrak{u}$ and $y_j=\frac12$ for $j\notin\mathfrak{u}$, whereas $\mathrm{d}\,x_{\mathfrak{u}}=\prod_{j\in\mathfrak{u}}\mathrm{d}\,x_j$ and $\partial\,x_{\mathfrak{u}}=\prod_{j\in\mathfrak{u}}\partial\,x_j$.

For multivariate integration

$$\mathrm{INT}_d(f)=\int_{[0,1]^d}f(t)\,\mathrm{d}t=\langle f,h_d\rangle_{H(K_d)}\quad\text{for all}\quad f\in H(K_d)$$

we now have, similarly as before,

$$\begin{aligned}h_d(x)&=\int_{[0,1]^d}K_d(x,t)\,\mathrm{d}t\\&=\prod_{j=1}^d\left[b+\tfrac12(\tfrac12-x_j)\left(\delta_{x_j\le1/2}(\tfrac12+x_j)+(1-\delta_{x_j\le1/2})(x+j-\tfrac32)\right)\right]\end{aligned}$$

for all $x\in[0,1]^d$. Here $\delta_{x\le1/2}=1$ for $x\le\frac12$ and $\delta_{x\le1/2}=0$ for $x>\frac12$.

This implies that we can take

$$R_1(x,t)=b\quad\text{and}\quad R_2(x,t)=1_M(x,t)\min(|x-\tfrac12|,|t-\tfrac12|)\quad\text{for all}\quad x,t\in[0,1].$$

Then $h_{1,1}=b$ and $\|h_{1,2,(0)}\|^2_{H(R_2)}=\|h_{1,2,(1)}\|^2_{H(R_2)}=1/(24)$. Hence we have

$$\alpha=\tfrac12\quad\text{and}\quad\beta=\frac{12b}{1+12b}.$$

The lower bound on $n^{\mathrm{ran}}(\varepsilon,\mathrm{INT}_d)$ is now of the form $\Omega(\varepsilon^{-p_b})$ with

$$p_b=\frac{2}{1+\ln(1+12b)/\ln(2)}.$$

For small b we have $p_b \approx 2$. Furthermore $p_{1/12} = 1$, $p_{1/4} = 2/3$ and $p_b < 2/3$ for $b > 1/4$.

Observe that for a fixed d, the value of b does not change the asymptotic behavior of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$. Therefore (9) holds and the exponent of ε^{-1} must be at least $\frac{2}{3}$. This means that the lower bound does not tell us anything useful for $b \geq 1/4$.

Acknowledgments

We are grateful for valuable comments of A. Hinrichs, A. Papageorgiou, J. F. Traub and two anonymous referees.

References

[1] N. Aronszajn, Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68, 337–404, 1950.

[2] N. S. Bakhvalov, On approximate computation of integrals, *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* 4, 3–18, 1959, in Russian.

[3] A. Hinrichs, Optimal importance sampling for the approximation of integrals, J. Complexity **26**, 125–134, 2009.

[4] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, LNiM, 1349, Springer-Verlag, Berlin, 1988.

[5] E. Novak and H. Woźniakowski, Intractability results for integration and discrepancy, J. Complexity **17**, 388–441, 2001.

[6] E. Novak and H. Woźniakowski, Tractability of Multivariate Problems, Volume I: Linear Information, EMS, Zürich, 2008.

[7] E. Novak and H. Woźniakowski, Tractability of Multivariate Problems, Volume II: Standard Information for Functionals, EMS, Zürich, 2010, to appear.

[8] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, Information-Based Complexity, Academic Press, 1988.

[9] G. Wahba, Spline Models for Observational Data. SIAM-NSF Regional Conference Series in Appl. Math, 59, *SIAM*, Philadelphia, 1990.

[10] G. W. Wasilkowski and H. Woźniakowski, The power of standard information for multivariate approximation in the randomized setting, Math. Comp. **76**, 965-988, 2007.

E. Novak, Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

E-mail address: erich.novak@uni-jena.de

H. Woźniakowski, Department of Computer Science, Columbia University, New York 10027, NY, USA, and, Institute of Applied Mathematics, University of Warsaw, Banacha2, 02-097 Warsaw, Poland

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