MATHEMATICS OF COMPUTATION
Volume 75, Number 255, July 2006, Pages 1259–1286
S 0025-5718(06)01865-5
Article electronically published on May 1, 2006

ON THE OPTIMAL CONVERGENCE RATE OF UNIVERSAL AND NONUNIVERSAL ALGORITHMS FOR MULTIVARIATE INTEGRATION AND APPROXIMATION

MICHAEL GRIEBEL AND HENRYK WOŹNIAKOWSKI

Abstract. We study the maximal rate of convergence (mrc) of algorithms for (multivariate) integration and approximation of d-variate functions from reproducing kernel Hilbert spaces $H(K_d)$. Here K_d is an arbitrary kernel all of whose partial derivatives up to order r satisfy a Hölder-type condition with exponent 2β. Algorithms use n function values and we analyze their rate of convergence as n tends to infinity. We focus on *universal* algorithms which depend on d, r, and β but not on the specific kernel K_d, and *nonuniversal* algorithms which may depend additionally on K_d.

For universal algorithms the mrc is $(r+\beta)/d$ for both integration and approximation, and for nonuniversal algorithms it is $1/2+(r+\beta)/d$ for integration and $a+(r+\beta)/d$ with $a \in [1/(4+4(r+\beta)/d), 1/2]$ for approximation. Hence, the mrc for universal algorithms suffers from the curse of dimensionality if d is large relative to $r+\beta$, whereas the mrc for nonuniversal algorithms does *not* since it is always at least $1/2$ for integration, and $1/4$ for approximation. This is the price we have to pay for using universal algorithms. On the other hand, if $r+\beta$ is large relative to d, then the mrc for universal and nonuniversal algorithms is approximately the same.

We also consider the case when we have the additional knowledge that the kernel K_d has product structure, $K_{d,r,\beta} = \bigotimes_{j=1}^{d} K_{r_j,\beta_j}$. Here K_{r_j,β_j} are some univariate kernels whose all derivatives up to order r_j satisfy a Hölder-type condition with exponent $2\beta_j$. Then the mrc for universal algorithms is $q := \min_{j=1,2,\dots,d}(r_j+\beta_j)$ for both integration and approximation, and for nonuniversal algorithms it is $1/2+q$ for integration and $a+q$ with $a \in [1/(4+4q), 1/2]$ for approximation. If $r_j \geq 1$ or $\beta_j \geq \beta > 0$ for all j, then the mrc is at least $\min(1,\beta)$, and the curse of dimensionality is *not* present. Hence, the product form of reproducing kernels breaks the curse of dimensionality even for universal algorithms.

1. Introduction

High-dimensional integration and approximation are important in many practical applications ranging from statistical mechanics to financial engineering. Depending on the particular problem, functions belong to a specific function class that captures their a priori known properties. Then the question arises: which algorithm is

Received by the editor July 25, 2005 and, in revised form, August 15, 2005.
2000 *Mathematics Subject Classification.* Primary 65D30, 65D15, 41A45.
The research of the first author was supported in part by the Sonderforschungsbereich 611 *Singuläre Phänomene und Skalierung in Mathematischen Modellen* sponsored by the *Deutsche Forschungsgemeinschaft.*
The research of the second author was supported in part by the National Science Foundation.

optimal? We can often construct optimal algorithms for a known function class. But a slight change of the underlying class usually results in a different optimal algorithm. Furthermore, the class of functions from which a specific given function stems is usually not known, and thus it is not clear which optimal algorithm should be used. Therefore, from a practical point of view, algorithms that are optimal for only one class of functions are of limited interest. Instead one is interested in a *universal* algorithm, which works well or (better yet) is almost optimal for a wide range of function classes.

The problem of universality for integration has been studied for many years. The earliest papers we could find where universality for integration is considered were [3, 9]. In the univariate case, this problem was also addressed in [5, 14]. In the multivariate case, the periodic setting was studied in [20, 21]. The nonperiodic case was considered in [11, 12]. Here, universality of quadrature was (up to logarithmic terms) achieved for the classes C_d^r and F_d^r, i.e., for Hölder spaces and spaces of functions with bounded *mixed* derivatives. In more general cases, however, no results on the convergence rate of universal quadratures are known. The reader may also consult [2], where the problem of universality is discussed from a more general point of view.

In this article we study the optimal rate of convergence of universal and nonuniversal algorithms for multivariate integration and approximation for d-variate functions from Hilbert spaces. Since we consider algorithms that use finitely many function values, we assume that the computation of a function value is a continuous linear functional. This is equivalent to the assumption that the Hilbert spaces have reproducing kernels. The theory of reproducing kernel Hilbert spaces [1] allows us to describe function spaces in a concise and elegant way by means of a kernel function. In this article, we consider universal algorithms that work for *all* reproducing kernel Hilbert spaces $H(K_d)$ whose kernels K_d have all partial derivatives up to order r satisfying a Hölder-type condition with exponent 2β. Here, r is a nonnegative integer and $\beta \in [0,1]$. We denote the class of such kernels by $\mathcal{K}_{d,r,\beta}$. This type of smoothness for reproducing kernels has been studied in many papers. An extensive list of references up to the year 2000 can be found in [15]. We stress that universal algorithms depend only on the smoothness properties of the kernels; i.e., they depend only on the smoothness parameters r and β and are independent of the specific form of the kernel, whereas nonuniversal algorithms may additionally depend on the form of the kernel.

We show that without loss of generality we may use *nested* linear algorithms, i.e., linear algorithms that reuse the previously computed function values. To be precise, the $(n+1)$st step of such an algorithm uses n function values computed in previous steps and at most one new function value that needs to be computed at this step. We present a nested linear algorithm that is universal, since its weights and sample points are dependent only on d, r, and β but not on the specific kernel. We show that the error of this algorithm is of the order $O(n^{-(r+\beta)/d})$ for *all* spaces $H(K_d)$ with K_d from the class $\mathcal{K}_{d,r,\beta}$. Hence, its rate of convergence is $(r+\beta)/d$. We prove that the rate of convergence $(r+\beta)/d$ is optimal; i.e., there is no universal algorithm with a better rate of convergence. This holds for both multivariate integration and approximation. Lower bounds on the optimal order follow from known results, mostly from [10, 17]. Upper bounds are proved here by constructing a kernel (or equivalently, a Hilbert space) that is difficult for a given universal algorithm.

We also study the optimal rate of convergence for nonuniversal algorithms. That is, we want to find the largest possible p such that for every K_d from the class $\mathcal{K}_{d,r,\beta}$, there exists an algorithm depending on K_d whose error is of the order $O(n^{-p})$ for all functions from $H(K_d)$. It turns out that for multivariate integration, the optimal rate of convergence for nonuniversal algorithms is $1/2 + (r + \beta)/d$. For multivariate approximation, we are able to present a bound $a + (r + \beta)/d$, with $a \in [1/(4 + 4(r + \beta)/d), 1/2]$, on the optimal rate of convergence.

Let us now compare the optimal rates of convergence of universal and nonuniversal algorithms. We start with multivariate integration. Observe that for d large relative to $r + \beta$, the optimal rate of convergence of universal algorithms is small and goes linearly to zero with d^{-1}. This bad property can be seen more clearly if we want to guarantee that the error achieved by an algorithm is at most ε. Then we need to perform $n = \Theta(\varepsilon^{-d/(r+\beta)})$ steps of the corresponding universal algorithm, and we encounter the curse of dimensionality, since n depends exponentially on the number d of variables. For nonuniversal algorithms, however, the curse of dimensionality is *broken*, since their optimal rate of convergence is always at least $1/2$; hence, it is enough to perform $n = \Theta(\varepsilon^{-2/(1+2(r+\beta)/d)})$ steps of the corresponding nonuniversal algorithm to guarantee an error of size ε. Note that the exponent of n is now at most 2. However, the factor in the Θ-notation has an (as yet) unknown dependence on d; this dependence may be exponential, if not worse.

For multivariate approximation, the situation is similar. We have the curse of dimensionality for universal algorithms, but no curse for nonuniversal algorithms, since their optimal rate of convergence is always at least $1/4$ and the number of steps of the corresponding nonuniversal algorithm is at most $\Theta(\varepsilon^{-4})$. As before, the factor in the Θ-notation is an unknown function of d, which may grow at an alarming rate.

In short, the improvements in the optimal rate of convergence for nonuniversal algorithms are significant for both multivariate integration and approximation. This shows the price we have to pay if we want to use universal algorithms when d is large relative to $r+\beta$. On the other hand, if $r+\beta$ is large relative to d, the optimal rates of convergence of universal and nonuniversal algorithms are approximately the same, so that there is no serious loss in the order of convergence if we use universal algorithms in this case.

Next, we consider kernels with product structure. In this case, the Hilbert space $H(K_d)$ is a tensor product of Hilbert spaces $H(K_{r_j,\beta_j})$ of univariate functions with the reproducing kernel K_{r_j,β_j} that is r_j times continuously differentiable and satisfies a Hölder-type condition with exponent $2\beta_j$ for $j = 1, 2, \dots, d$. The corresponding class of such kernels is denoted by $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$. This class corresponds to Hilbert spaces $H(K_d)$ of functions with bounded mixed derivatives. We prove that for the class $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$, the optimal order of convergence for universal algorithms is $q := \min_{j=1,2,\dots,d}(r_j + \beta_j)$ for both multivariate integration and approximation. For nonuniversal algorithms, the optimal rate of convergence is $1/2 + q$ for multivariate integration, and $a + q$ with $a \in [1/(4 + 4q), 1/2]$ for multivariate approximation. The number d of variables now plays a different role than before. Observe that the optimal rate of convergence of universal algorithms depends on d only through the minimum of the local regularities $r_j + \beta_j$. If we consider the class $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$ for which $\min_{j=1,2,\dots,d}\{r_j + \beta_j\} \geq a > 0$ with a independent of d, then the optimal rate is at least a, and the curse of dimensionality is *not* present for product kernels. Indeed, it is enough to use the corresponding universal algorithm with $\Theta(\varepsilon^{-1/a})$ function values to compute an ε-approximation. As before, the factor in the Θ-notation may be exponential in d. For nonuniversal algorithms we have even better bounds on the optimal orders of at least $1/2$ for multivariate integration, and of at least the number a for multivariate approximation. Hence, we break the curse of dimensionality of universal algorithms for the class $\mathcal{K}_{d,r,\beta}$ if we switch to the class $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$ of product kernels with $\min_{j=1,2,\dots,d}\{r_j+\beta_j\} \ge a > 0$ and a independent of d. This means that the product structure of reproducing kernels is a powerful property.

The remainder of this paper is organized as follows. In §2 we give some basics of the theory of reproducing kernel Hilbert spaces. In §3 we state the problem of multivariate integration and multivariate approximation. Then in §4, we introduce the class of kernels $\mathcal{K}_{d,r,\beta}$ with global smoothness properties measured by the two parameters r and β. We also recall that the Hilbert spaces $H(K_d)$ with K_d from the class $\mathcal{K}_{d,r,\beta}$ are subsets of the space of smooth continuous functions $C^{r,\beta}(D)$. In §5, we consider linear algorithms for the spaces $C^{r,\beta}(D)$ and $H(K_d)$ for multivariate integration and approximation. We present nested linear universal algorithms with optimal rates of convergence in §6. In §7 we study nonuniversal algorithms and their optimal rates of convergence. In §8, we consider kernels with product structure and derive their optimal rates of convergence. We finish with some concluding remarks and open problems in §9.

2. Reproducing kernel Hilbert spaces

In this article we study multivariate integration and approximation for real functions defined on $D=[0,1]^d$. We assume that these functions belong to a Hilbert space H with associated inner product $\langle\cdot,\cdot\rangle_H$ and norm $\|f\|_H = \langle f,f\rangle_H^{1/2}$. We assume that H is continuously embedded into $L_2(D)$. Thus, we consider integrable functions f with respect to the Lebesgue measure for which $\|f\|_{L_2(D)} := \left(\int_D f^2(t)dt\right)^{1/2} < \infty$. Furthermore, there is a nonnegative number $c(H)$ depending on the space H such that

$$\|f\|_{L_2(D)} \le c(H)\,\|f\|_H \qquad \text{for all } f\in H. \tag{2.1}$$

We will study algorithms that use finitely many function values. Then we need to assume that for any $x\in D$, the linear functional $f\in H \mapsto f(x)$ is continuous. This is equivalent to the requirement that H be a reproducing kernel Hilbert space; see [1]. Hence, H has an associated kernel $K_d : D\times D \to \mathbb{R}$ that is uniquely defined by the following three conditions:

- $K_d(\cdot,t)\in H_d$ for all $t\in D$,
- $(K_d(x_i,x_j))_{i,j=1}^n$ is a symmetric and nonnegative definite matrix for all n and points x_i from D,
- $f(t) = \langle f, K_d(\cdot,t)\rangle_H$ for all $f\in H_d$ and all $t\in D$ (reproducing kernel property).

The theory of reproducing kernel Hilbert spaces can be found in detail in [1]; further aspects are discussed in, e.g., [15, 24]. This theory allows us to describe function spaces in a concise and elegant way by means of a reproducing kernel. Therefore, we denote in the following the Hilbert space H by $H(K_d)$ and the associated inner product and norm by $\langle\cdot,\cdot\rangle_{H(K_d)}$ and $\|\cdot\|_{H(K_d)}$, respectively. We now review some known properties of reproducing kernel Hilbert spaces, which are especially relevant to the proof technique presented in this paper.

From the three properties of reproducing kernels it easily follows that

$$\begin{aligned} K_d(t,x) &= \langle K_d(\cdot,x), K_d(\cdot,t)\rangle_H && \text{for all } t,x \in D,\\ \sqrt{K_d(t,t)} &= \|K_d(\cdot,t)\|_H && \text{for all } t \in D,\\ |f(t)| &\le \|f\|_H \sqrt{K_d(t,t)} && \text{for all } f \in H,\ t \in D. \end{aligned}$$

If $H = H(K_d)$ is separable, then for an arbitrary orthonormal basis $\{\eta_i\}$, we have $K_d(\cdot,x) = \sum_{i=1}^{\dim(H)} c_i\eta_i$ with $c_i = \langle \eta_i, K_d(\cdot,x)\rangle_H = \eta_i(x)$. Therefore

$$K_d(x,t) = \sum_{i=1}^{\dim(H)} \eta_i(x)\eta_i(t) \qquad \text{for all } x,t \in D. \tag{2.2}$$

In a way, the reverse of this argument is also true; see [24]. To this end, let $\{\eta_i\}_{i=1}^{\infty}$ be a given arbitrary sequence of linearly independent functions defined on D such that $\sum_{i=1}^{\infty}\eta_i^2(t) < \infty$ for all $t \in D$. Consider the space $H = \text{span}\{\eta_1,\eta_2,\dots\}$ of functions $f(t) = \sum_{i=1}^{\infty} f_i\eta_i(t)$ with real numbers f_i such that $\sum_{i=1}^{\infty} f_i^2 < \infty$. Observe that $f(t)$ is well defined. For $f \in H$ the coefficients f_i are uniquely determined since the η_i's are linearly independent. The inner product in H is given by requiring that the η_i's be orthonormal, $\langle\eta_i,\eta_j\rangle_H = \delta_{i,j}$. Hence, for $f,g \in H$ we have $\langle f,g\rangle_H = \sum_{i=1}^{\infty} f_i g_i$, with f_i and g_i being the coefficients of f and g, respectively. Then H is a Hilbert space. We claim that

$$K_d(x,t) = \sum_{i=1}^{\infty}\eta_i(x)\eta_i(t)$$

is its reproducing kernel. Indeed, $K_d(\cdot,t) \in H$ for all $t \in D$. Consider the matrix $M = (K_d(x_i,x_j))_{i,j=1}^n$ for arbitrary n and $x_i, x_j \in D$. Clearly, M is symmetric. Moreover, for arbitrary real numbers a_i we have

$$\sum_{i,j=1}^n a_i a_j K_d(x_i,x_j) = \sum_{k=1}^{\infty}\sum_{i,j=1}^n a_i a_j \eta_k(x_i)\eta_k(x_j) = \sum_{k=1}^{\infty}\left(\sum_{i=1}^n a_i\eta_k(x_i)\right)^2 \ge 0.$$

This proves that M is nonnegative definite. Finally, for all $f \in H$ and $t \in D$ we have

$$\langle f, K_d(\cdot,t)\rangle_H = \sum_{i=1}^{\infty} f_i\eta_i(t) = f(t).$$

Hence, K_d is the reproducing kernel of H, as claimed.

Note that the Hilbert space $L_2(D)$ does not have a reproducing kernel, since point evaluation $t \in D \mapsto f(t)$ is not well defined for $L_2(D)$ and thus cannot be continuous. It is easy to see that $H(K_d)$ is continuously embedded in $L_2(D)$ if we assume that

$$\int_D K_d(t,t)\,dt < \infty. \tag{2.3}$$

Indeed, $f^2(t) \le \|f\|^2_{H(K_d)} K_d(t,t)$, and therefore (2.1) holds with

$$c(H(K_d)) = \left(\int_D K_d(t,t)\,dt\right)^{1/2}. \tag{2.4}$$

In this case, $H_d(K)$ is a proper subset of $L_2(D)$, and $K_d(\cdot,t) \in L_2(D)$ for arbitrary $t \in D$.

We add in passing that if (2.3) does not hold, then $H(K_d)$ need not be a subset of $L_2(D)$. Indeed, take a function $h : D \to \mathbb{R}$ that is *not* integrable, say, $h(x) = x_1^{-1}$ for $x_1 \in (0,1]$ and $h(x) = 0$ for $x_1 = 0$. Let $K_d(x,t) = h(x)h(t)$. Then $H(K_d) = \operatorname{span}\{h\}$ is a one-dimensional Hilbert space, and $H(K_d)$ is not a subset of $L_2(D)$. In this case $\int_D K_d(x,t)\,dt = \left(\int_0^1 x^{-2}\,dx\right)^2 = \infty$.

Many examples of reproducing kernel Hilbert spaces can be found in the literature; see, for example, [15, 24]. Here we only mention one example of a tensor product Hilbert space that is often studied for multivariate integration. Let $H(K_d) = W_2^1([0,1]) \otimes W_2^1([0,1]) \otimes \cdots \otimes W_2^1([0,1])$, where $W_2^1([0,1])$ is the Sobolev space of univariate absolutely continuous functions $f : [0,1] \to \mathbb{R}$ for which $f' \in L_2([0,1])$, under the inner product

$$\langle f,g\rangle_{W_2^1([0,1])} \;=\; f(0)g(0) + \int_0^1 f'(t)g'(t)\,dt.$$

The reproducing kernel K_d takes the form

$$K_d(x,t) \;=\; \prod_{i=1}^d (1+\min\{x_i,t_i\})$$

for $x = [x_1,x_2,\dots,x_d]$, $t = [t_1,t_2,\dots,t_d] \in D$. The inner product in $H(K_d)$ is given by

$$\langle f,g\rangle_{H(K_d)} \;=\; f(0)g(0) \;+ \sum_{\emptyset\neq u\subset\{1,2,\dots,d\}} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial x_u} f(x_u,0) \frac{\partial^{|u|}}{\partial x_u} g(x_u,0)\,dx_u.$$

Here, the sum is over all nonempty subsets u of the index set $\{1,2,\dots,d\}$, and therefore we have 2^d-1 terms. Each term is an integral over the $|u|$-dimensional unit cube, where $|u|$ denotes the cardinality of the set u. For $x \in D$, the vector x_u denotes the $|u|$-dimensional vector that consists of the components of x_j with $j \in u$. Finally, the vector $(x_u,0)$ denotes the d-dimensional vector whose jth component is x_j for $j \in u$ and is zero for $j \notin u$. Details can be found in, e.g., [19].

3. Multivariate integration and approximation

For a reproducing kernel Hilbert space $H(K_d)$ with kernel K_d satisfying (2.3), we consider multivariate integration, $\mathrm{INT}_d : H(K_d) \to \mathbb{R}$, defined as

$$\mathrm{INT}_d(f) \;=\; \int_D f(t)\,dt,$$

and multivariate approximation, $\mathrm{APP}_d : H(K_d) \to L_2(D)$, defined as

$$\mathrm{APP}_d(f) \;=\; f.$$

Here, we want to approximate functions f from the space $H(K_d)$ in the weaker norm of the space $L_2(D)$. Thus multivariate approximation is an embedding operator.

For the reader's convenience, we recall some well-known facts about multivariate integration and approximation. It is easy to see that INT_d and APP_d are continuous linear operators. Their norms are defined as

$$\|\mathrm{INT}_d\| := \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} |\mathrm{INT}_d(f)| \qquad \text{and} \qquad \|\mathrm{APP}_d\| := \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} \|f\|_{L_2(D)}.$$

Since $|\mathrm{INT}_d(f)| \le \|f\|_{L_2(D)}$, the relation $\|\mathrm{INT}_d\| \le \|\mathrm{APP}_d\|$ directly follows. Furthermore, $f^2(t) \le \|f\|_{H(K_d)} K_d(t,t)$ and (2.4) imply $\|\mathrm{INT}_d\| \le \|\mathrm{APP}_d\| \le c(H(K_d))$.

We first consider multivariate integration. Letting

$$h_d(t) = \int_D K_d(x,t)\,dx \qquad \text{for all } t \in D,$$

it is easy to verify that h_d is the Riesz representer for multivariate integration, i.e., that

$$\mathrm{INT}_d(f) = \langle f, h_d\rangle_{H(K_d)} \qquad \text{for all } f \in H(K_d)$$

and

$$\|\mathrm{INT}_d\| = \|h_d\|_{H(K_d)} = \left(\int_D \int_D K_d(x,t)\,dx\,dt\right)^{1/2} = \left(\int_D h_d(x)dx\right)^{1/2}.$$

We approximate $\mathrm{INT}_d(f)$ by a quadrature

$$Q_{n,d}(f) = \sum_{i=1}^n a_i f(x_i),$$

with weights $a_i \in \mathbb{R}$ and sample points $x_i \in D$. Note that $Q_{n,d}$ is also a continuous linear functional since

$$Q_{n,d}(f) = \left\langle f, \sum_{i=1}^n a_i K_d(\cdot, x_i)\right\rangle_{H(K_d)}.$$

Observe that

$$\|Q_{n,d}\| = \left\|\sum_{i=1}^n a_i K_d(\cdot, x_i)\right\|_{H(K_d)} = \left(\sum_{i,j=1}^n a_i a_j K_d(x_i, x_j)\right)^{1/2}.$$

The error $\mathrm{INT}_d(f) - Q_{n,d}(f)$ can be represented as

$$\mathrm{INT}_d(f) - Q_{n,d}(f) = \left\langle f, h_d - \sum_{i=1}^n a_i K_d(\cdot, x_i)\right\rangle_{H(K_d)}.$$

The worst case integration error of $Q_{n,d}$ is defined as the largest error for all f in the unit ball in $H(K_d)$,

(3.1) $$e^{\mathrm{INT}}(Q_{n,d}, K_d) = \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} |\mathrm{INT}_d(f) - Q_{n,d}(f)|.$$

Then it is clear that

$$
\begin{aligned}
e^{\mathrm{INT}}(Q_{n,d},K_d)^2 &= \|\mathrm{INT}_d - Q_{n,d}\|^2 = \Big\| h_d - \sum_{i=1}^n a_i K_d(\cdot,x_i)\Big\|^2_{H(K_d)} \\
&= \|h_d\|^2_{H(K_d)} - 2\sum_{i=1}^n a_i h_d(x_i) + \Big\|\sum_{i=1}^n a_i K_d(\cdot,x_i)\Big\|^2_{H(K_d)} \\
&= \int_D\int_D K_d(x,y)\,dx dy - 2\sum_{i=1}^n a_i \int_D K_d(x_i,y)\,dy \\
&\quad + \sum_{i,j=1}^n a_i a_j K(x_i,x_j). \qquad (3.2)
\end{aligned}
$$

In this way, we obtain a well-known explicit formula for the worst case integration error.

Next, we consider multivariate approximation $\mathrm{APP}_d : H(K_d) \to L_2(D)$. We first have to find the adjoint operator $\mathrm{APP}_d^* : L_2(D) \to H(K_d)$. Observe that for $f \in H(K_d)$ and $g \in L_2(D)$ we have

$$
\begin{aligned}
\langle f, \mathrm{APP}_d^*(g)\rangle_{H(K_d)} = \langle \mathrm{APP}_d(f), g\rangle_{L_2(D)} &= \int_D f(t)g(t)\,dt \\
&= \int_D \langle f, K_d(\cdot,t)\rangle_{H(K_d)}\, g(t)\,dt = \Big\langle f, \int_D K_d(\cdot,t)g(t)\,dt\Big\rangle_{H(K_d)}.
\end{aligned}
$$

Hence,

$$\mathrm{APP}_d^*(g)\,(x) = \int_D K_d(x,t)g(t)\,dt \qquad \text{for all } g \in L_2(D).$$

For $f \in H(K_d)$, this yields

$$\|f\|^2_{L_2(D)} = \langle \mathrm{APP}_d(f), \mathrm{APP}_d(f)\rangle_{L_2(D)} = \langle W_d(f), f\rangle_{H(K_d)},$$

where

$$W_d = \mathrm{APP}_d^* \circ \mathrm{APP}_d : H(K_d) \to H(K_d) \qquad (3.3)$$

is given by

$$W_d(f)\,(x) = (\mathrm{APP}_d^* \circ \mathrm{APP}_d f)\,(x) = \mathrm{APP}_d^*(f)\,(x) = \int_D K_d(x,t)f(t)\,dt \qquad \text{for all } f \in H(K_d).$$

Clearly, W_d is a self-adjoint nonnegative definite operator satisfying

$$\|\mathrm{APP}_d\| = \|W_d\|^{1/2}.$$

The norm of W_d is equal to the largest eigenvalue of W_d and is bounded by $\int_D K_d(t,t)\,dt$.

We approximate $\mathrm{APP}_d(f)$ by linear algorithms

$$A_{n,d}(f) = \sum_{i=1}^n w_i f(x_j), \qquad (3.4)$$

where, $x_i \in D$ as before, but the weights w_i are now *functions* from $L_2(D)$. Note that $A_{n,d}$ is a continuous linear operator. The worst case approximation error of

$A_{n,d}$ is defined as

$$(3.5)\quad e^{\mathrm{APP}}(A_{n,d},K_d) = \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} \|\mathrm{APP}_d(f) - A_{n,d}(f)\|_{L_2(D)} \;=\; \|\mathrm{APP}_d - A_{n,d}\|.$$

The operator norm of $\|\mathrm{APP}_d - A_{n,d}\|$ is equal to the square of the largest eigenvalue of the operator $(\mathrm{APP}_d - A_{n,d})^*(\mathrm{APP}_d - A_{n,d})$, for which no explicit formula is known. Hence we do not have an explicit formula analogous to (3.2) giving the worst case error for multivariate approximation. But it is known that the error of $A_{n,d}$ is at least equal to the square of the $(n+1)$st largest eigenvalue of the operator W_d; see, e.g., [22].

However, it is straightforward to show that multivariate approximation is not easier than multivariate integration. Indeed, for any approximation algorithm $A_{n,d}$ of the form (3.4), we can define a quadrature $Q^A_{n,d}$ by integrating $A_{n,d}(f)\,(t)$ over t, so that

$$Q^A_{n,d}(f) \;=\; \sum_{i=1}^{n}\left(\int_D w_i(t)\,dt\right) f(x_i).$$

Then

$$\mathrm{INT}_d(f) - Q^A_{n,d}(f) \;=\; \int_D\left(f(t) - \sum_{i=1}^{n} w_i(t)f(x_i)\right)dt$$

and

(3.6)

$$|\mathrm{INT}_d(f) - Q^A_{n,d}(f)| \;\le\; \left[\int_D\left(f(t) - \sum_{i=1}^{n} w_i(t)f(x_i)\right)^2 dt\right]^{1/2} \;=\; \|f - A_{n,d}\|_{L_2(D)} \qquad \text{for all } f\in H(K_d).$$

Hence,

$$(3.7)\qquad e^{\mathrm{INT}}(Q^A_{n,d},K_d) \;\le\; e^{\mathrm{APP}}(A_{n,d},K_d)$$

for all approximation algorithms $A_{n,d}$.

4. Global smoothness

We want to find universal algorithms that will be efficient for a (possibly large) class of Hilbert spaces with reproducing kernels. We determine these classes by the global smoothness of their kernels. To this end, for a nonnegative integer r and $\beta\in[0,1]$ we define $\mathcal{K}_{d,r,\beta}$ to be the class of reproducing kernels K_d such that

$$(4.1)\qquad K_d \;\in\; C^{r,r}(D\times D),$$

$$(4.2)\qquad K_d^{(\alpha,\alpha)}(x,x) - 2K_d^{(\alpha,\alpha)}(x,y) + K_d^{(\alpha,\alpha)}(y,y) \le\; L_{K_d}\|x-y\|^{2\beta} \qquad \text{for } |\alpha|\le r \text{ and } x,y\in D.$$

Here, we use standard multi-index notation: for $\alpha = [\alpha_1,\alpha_2,\dots,\alpha_d]$ with nonnegative integers α_j, we write $D^{(\alpha)} = \partial^{|\alpha|}/\partial x_1^{\alpha_1}\cdots\partial x_d^{\alpha_d}$ and $|\alpha| := \alpha_1+\alpha_2+\cdots+\alpha_d \le r$. We let $C^{r,r}(D\times D)$ denote the space of functions on $D\times D$ whose partial derivatives of order up to r are continuous, with $K_d^{(\alpha,\alpha^*)}(x,y) \;=\; D_x^{\alpha}D_y^{\alpha^*}K(x,y)$ for multi-indices α,α^* and $x,y\in D$. Note that the nonnegative number L_{K_d} in (4.2) depends only on the kernel K_d. Thus we assume that K_d is r times continuously differentiable and all its rth derivatives satisfy a Hölder-type condition with the exponent 2β. Note that the specific choice of the vector norm in (4.2) is not important since all vector norms are equivalent.

Kernels having this kind of smoothness have been well studied; see, e.g., [15] and the references cited therein. In particular, for any $f \in H(K_d)$ and any α with $|\alpha| \le r$, we have

$$\max_{x\in D} |f^{(\alpha)}(x)| \;\le\; \|f\|_{H(K_d)} \left\{ \max_{t\in D} K_d^{(\alpha,\alpha)}(t,t) \right\}^{1/2}. \tag{4.3}$$

Formula (4.3) can be derived as follows. Differentiation of $f(t) = \langle f, K_d(\cdot,t)\rangle_{H(K_d)}$ results in

$$f^{(\alpha)}(t) \;=\; \left\langle f, K_d^{(0,\alpha)}(\cdot,t)\right\rangle_{H(K_d)}.$$

In particular, for $f = K_d^{(0,\alpha)}(\cdot,x)$ we have

$$K_d^{(\alpha,\alpha)}(t,x) = (K_d^{(0,\alpha)}(\cdot,x), K_d^{(0,\alpha)}(\cdot,t))_{H(K_d)}.$$

For $x = t$ we obtain

$$K_d^{(\alpha,\alpha)}(t,t) = \|K_d^{(0,\alpha)}(\cdot,t)\|^2_{H(K_d)},$$

so that

$$|f^{(\alpha)}(t)| \le \|f\|_{H(K_d)} K_d^{(\alpha,\alpha)}(t,t)^{1/2},$$

which yields (4.3).

Furthermore, for $|\alpha| = r$ we have

$$f^{(\alpha)}(x) - f^{(\alpha)}(y) \;=\; \left\langle f, K_d^{(0,\alpha)}(\cdot,x) - K_d^{(0,\alpha)}(\cdot,y)\right\rangle_{H(K_d)}$$

and

$$\begin{aligned} |f^{(\alpha)}(x) - f^{(\alpha)}(y)| \;&\le\; \|f\|_{H(K_d)}\, \|K_d^{(0,\alpha)}(\cdot,x) - K_d^{(0,\alpha)}(\cdot,y)\|_{H(K_d)} \\ &=\; \|f\|_{H(K_d)} \left(K_d^{(\alpha,\alpha)}(x,x) - 2K_d^{(\alpha,\alpha)}(x,y) + K_d^{(\alpha,\alpha)}(y,y)\right)^{1/2} \\ &\le\; \|f\|_{H(K_d)}\, L_{K_d}^{1/2}\, \|x-y\|^{\beta}, \end{aligned}$$

where we used (4.2) in the last estimate.

This shows that for $K_d \in \mathcal{K}_{d,r,\beta}$ we have $H(K_d) \subset C^{r,\beta}(D)$, where $C^{r,\beta}(D)$ denotes the space of all functions from $C^r(D)$ whose partial derivatives up to order r satisfy a Hölder condition with the exponent β. The norm in the space $C^{r,\beta}(D)$ is given as

$$\|f\|_{C^{r,\beta}(D)} \;=\; \max\left\{ \max_{\alpha:\,|\alpha|\le r}\, \max_{x\in D} |f^{(\alpha)}(x)|,\quad \sup_{\substack{\alpha:\,|\alpha|=r\\ x,y\in D,\, x\ne y}} \frac{|f^{(\alpha)}(x) - f^{(\alpha)}(y)|}{\|x-y\|^{\beta}} \right\}.$$

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

$$\|f\|_{C^{r,\beta}(D)} \le c(K_d,r,\beta)\|f\|_{H(K_d)} \tag{4.4}$$

$$\text{with}\quad c(K_d,r,\beta) \;=\; \max\left\{ \max_{\alpha:\,|\alpha|\le r}\, \max_{t\in D} K^{(\alpha,\alpha)}(t,t)^{1/2},\, L_{K_d}^{1/2} \right\}.$$

5. Algorithms for $C^{r,\beta}(D)$ and $H(K_d)$ with $K_d \in \mathcal{K}_{d,r,\beta}$

For the class $C^{r,\beta}(D)$ it is known that the optimal error bounds of algorithms for multivariate integration and approximation are $\Theta(n^{-(r+\beta)/d})$; see [10, p. 34]. Thus, the optimal rate of convergence is $(r+\beta)/d$. We now briefly discuss algorithms achieving this optimal rate of convergence.

For $k = 1, 2, \ldots$, we subdive the unit cube D into k^d sub-cubes $D_{k,j}$ of side-length $1/k$. The sub-cubes $D_{k,j}$ are nonoverlapping with edges parallel to the coordinate axes. Let $l = r-1$ if $\beta = 0$, and $l = r$ if $\beta \in (0,1]$. Let $\nu = \binom{d+l}{l}$ be the dimension of the linear space $P_{d,l}$ of polynomials of d variables of degree at most l. For each sub-cube $D_{k,j}$ we use ν sample points $x^*_{k,j,1}, x^*_{k,j,2}, \ldots, x^*_{k,j,\nu}$ from $D_{k,j}$, and ν polynomials $w^*_{k,j,1}, w^*_{k,j,1}, \ldots, w^*_{k,j,\nu}$ of d variables of degree at most l such that

$$p = \sum_{i=1}^{\nu} p(x^*_{k,j,i})\, w^*_{k,j,i} \qquad \forall\, p \in P_{d,l}.$$

We consider algorithms for which the total number of sample points is $n_k = \nu\, k^d$. For multivariate approximation, we define the algorithm

$$A^*_{n_k,d}(f) = \sum_{j=1}^{k^d} \sum_{i=1}^{\nu} w^*_{k,j,i} f(x^*_{k,j,i}),$$

and for multivariate integration, we define the quadrature

$$Q^*_{n_k,d}(f) = \int_D A^*_{n_k,d}(f)\,(t)\,dt = \sum_{j=1}^{k^d} \sum_{i=1}^{\nu} a^*_{k,j,i} f(x^*_{k,j,i})$$
$$\text{with} \quad a^*_{k,j,i} = \int_D w^*_{k,j,i}(t)\,dt.$$

Both algorithms involve n_k sample points when $k = 1, 2, \ldots$. For an arbitrary positive integer n we proceed as follows. For $n < \nu$, we set $A^*_{n,d} = Q^*_{n,d} = 0$. For $n \geq \nu$, there exists a unique k such that $n \in [n_k, n_{k+1})$. We define

$$A^*_{n,d}(f) = A^*_{n_k,d}(f) \qquad \text{and} \qquad Q^*_{n,d}(f) = Q^*_{n_k,d}(f). \tag{5.1}$$

The approximation error of $A^*_{n_k,d}$ in the space $C^{r,\beta}(D)$ is of order $k^{-(r+\beta)} = \Theta(n_k^{-(r+\beta)/d})$ which can be proven by using Taylor's theorem on each sub-cube $D_{k,j}$. Since $n = \Theta(n_k)$, the approximation error of $A^*_{n,d}$ in the space $C^{r,\beta}(D)$ is of order $n^{-(r+\beta)/d}$, as claimed. Obviously, the same holds for multivariate integration, due to (3.7). Hence, we have

$$\left(\mathrm{APP}_d(f) - A^*_{n,d}(f)\right)(x) = O\left(n^{-(r+\beta)/d}\, \|f\|_{C^{r,\beta}(D)}\right), \quad \text{for all } x \in D, \tag{5.2}$$

$$\|\mathrm{APP}_d(f) - A^*_{n,d}(f)\|_{L_2(D)} = O\left(n^{-(r+\beta)/d}\, \|f\|_{C^{r,\beta}(D)}\right), \tag{5.3}$$

$$|\mathrm{INT}_d(f) - Q^*_{n,d}(f)| = O\left(n^{-(r+\beta)/d}\, \|f\|_{C^{r,\beta}(D)}\right). \tag{5.4}$$

Here, the factors in the O-notation do not depend on x, n, and f but may depend on d, r, and β.

We emphasize that for different values of n the algorithms $A^*_{n,d}$ and $Q^*_{n,d}$ may use different sample points. For example, consider the case $n = n_{k+1} - 1$. If we increase n by 1, then we use a different partition of the cube D, and the new sample points $x^*_{k+1,j,i}$ may not be related to the previous sample points $x^*_{k,j,i}$. This is not a desirable property, since the number of function values used by, say, $A^*_{1,d}, A^*_{2,d}, \dots, A^*_{n,d}$ may be of order n^2. The same holds for the quadrature case. It would be better to have "nested" algorithms $A^{\rm nes}_{n,d}$ and $Q^{\rm nes}_{n,d}$, i.e., algorithms of the form

$$A^{\rm nes}_{n,d}(f) \;=\; \sum_{j=1}^{n} w_{n,j} f(x_j) \qquad \text{and} \qquad Q^{\rm nes}_{n,d}(f) \;=\; \sum_{j=1}^{n} a_{n,j} f(x_j)$$

for some functions $w_{n,j}$, numbers $a_{n,j}$, and some sequence $x_1, x_2, \dots, x_n, \dots$ of sample points. The main idea behind nested algorithms is that the function values used by, say, $A^{\rm nes}_{n,d}$ are reused by $A^{\rm nes}_{n+1,d}$, so that we only need to compute at most one extra function value $f(x_{n+1})$ to obtain $A^{\rm nes}_{n+1,d}(f)$. Hence, the number of function values used by the sequence $A^{\rm nes}_{1,d}, A^{\rm nes}_{2,d}, \dots, A^{\rm nes}_{n,d}$ is n. The same holds in the quadrature case for the sequence $Q^{\rm nes}_{1,d}, Q^{\rm nes}_{2,d}, \dots, Q^{\rm nes}_{n,d}$.

We now show that it is possible to have nested algorithms for multivariate approximation and integration with the same order of convergence as $A^*_{n,d}$ and $Q^*_{n,d}$, respectively. Let X^*_n be the set of the sample points used by $A^*_{n,d}$ and $Q^*_{n,d}$. We form the sequence $\{x_j\}$ of sample points by taking

$$X^*_{2^0}, X^*_{2^1}, \dots, X^*_{2^m}, \dots \tag{5.5}$$

with an arbitrary ordering of the elements in the sets $X^*_{2^m}$. Observe that the cardinality of $\bigcup_{l=0}^m X^*_{2^l}$ is at most $c_m \;=\; 2^{m+1} - 1$. We are ready to define the nested algorithms $A^{\rm nes}_{n,d}$ and $Q^{\rm nes}_{n,d}$. For an arbitrary $n \ge 1$, there exists a unique m such that $n \in [c_m, c_{m+1})$. We set

$$A^{\rm nes}_{n,d}(f) \;=\; A^*_{2^m,d}(f) \qquad \text{and} \qquad Q^{\rm nes}_{n,d}(f) \;=\; Q^*_{2^m,d}(f). \tag{5.6}$$

Thus, the algorithms $A^{\rm nes}_{n,d}$ and $Q^{\rm nes}_{n,d}$ are the same for $n \in \{c_m, c_m+1, \dots, c_{m+1}-1\}$, coinciding with the algorithms $A^*_{2^m,d}$ and $Q^*_{2^m,d}$ that use the sample points from the set $X^*_{2^m}$. So $A^{\rm nes}_{n,d}$ and $Q^{\rm nes}_{n,d}$ use the first n elements of the sequence (5.5) as sample points; the weights associated to the sample points not from the set $X^*_{2^m}$ are zero. Clearly, $n = \Theta(2^m)$ and therefore

$$\begin{aligned} \|{\rm APP}_d(f) - A^{\rm nes}_{n,d}(f)\|_{L_2(D)} &= \|{\rm APP}_d(f) - A^*_{2^m,d}(f)\|_{L_2(D)} \\ &= \; O(2^{-m(r+\beta)/d}) \;=\; O(n^{-(r+\beta)/d}), \end{aligned}$$

as claimed. Obviously, the same results holds for multivariate integration.

We now comment on the errors of $A^{\rm nes}_{n,d}$ and $Q^{\rm nes}_{n,d}$ for $f \in H(K_d)$ with $K_d \in \mathcal{K}_{d,r,\beta}$. We know that $H(K_d) \subset C^{r,\beta}(D)$. Also, formula (4.4) states that $\|f\|_{C^{r,\beta}(D)} \le c(K_d, r, \beta)\, \|f\|_{H(K_d)}$ for all $f \in H(K_d)$. Therefore for all $f \in H(K_d)$ with $K_d \in \mathcal{K}_{d,r,\beta}$, the estimates (5.2), (5.3), and (5.4) yield

$$\begin{aligned} \left({\rm APP}_d(f) - A^{\rm nes}_{n,d}(f)\right)(x) \;&= O\left(n^{-(r+\beta)/d} c(K_d, r, \beta) \|f\|_{H(K_d)}\right), \quad \text{for all } x \in D, \\ \|{\rm APP}_d(f) - A^{\rm nes}_{n,d}(f)\|_{L_2(D)} \;&= O\left(n^{-(r+\beta)/d} c(K_d, r, \beta) \|f\|_{H(K_d)}\right), \\ |{\rm INT}_d(f) - Q^{\rm nes}_{n,d}(f)| \;&= O\left(n^{-(r+\beta)/d} c(K_d, r, \beta) \|f\|_{H(K_d)}\right). \end{aligned}$$

Furthermore, for the worst case approximation and integration errors of $A_{n,d}^{\mathrm{nes}}$ and $Q_{n,d}^{\mathrm{nes}}$ we obtain

$$e^{\mathrm{APP}}(A_{n,d}^{\mathrm{nes}}, K_d) = O(c(K_d, r, \beta)\, n^{-(r+\beta)/d})$$

and

$$e^{\mathrm{INT}}(Q_{n,d}^{\mathrm{nes}}, K_d) = O(c(K_d, r, \beta)\, n^{-(r+\beta)/d})$$

for all $K_d \in \mathcal{K}_{d,r,\beta}$ with $c(K_d, r, \beta)$ given by (4.4). The factors in the O-notation in the last five formulas are independent of x, n, and K_d, but may depend on d, r, and β.

6. Universal algorithms

We are now ready to define the optimal rate of convergence for universal algorithms for multivariate integration and approximation over $H(K_d)$, with K_d from the class $\mathcal{K}_{d,r,\beta}$.

We begin with multivariate integration. Let $Q = \{Q_{n,d}\}$ be an arbitrary sequence of quadratures $Q_{n,d}(f) = \sum_{j=1}^{n} a_{n,j} f(x_{n,j})$ for some numbers $a_{n,j}$ and some sample points[1] $x_{n,j}$ from D. We want to find *universal* quadratures $Q_{n,d}$, i.e., quadratures that approximate integrals for *all* functions from the space $H(K_d)$ and for *all* kernels K_d from the class $\mathcal{K}_{d,r,\beta}$ with the optimal (largest possible) rate of convergence. We stress that these quadratures may only use the smoothness property of the kernels; i.e., they may depend on the number of variables d and the smoothness parameters r and β, but they must be independent of the specific form of the kernel K_d. The optimal rate of convergence is defined as

$$p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}) = \sup\Big\{ p \ge 0 \ : \exists\, Q = \{Q_{n,d}\} \text{ such that } \forall\, K_d \in \mathcal{K}_{d,r,\beta} \ \ \forall\, f \in H(K_d) \ \lim_{n\to\infty} n^p \,|\mathrm{INT}_d(f) - Q_{n,d}(f)| = 0 \Big\}. \tag{6.1}$$

We say that $Q = \{Q_{n,d}\}$ is a *universal* quadrature for the class $\mathcal{K}_{d,r,\beta}$ if

$$\lim_{n\to\infty} n^p \,|\mathrm{INT}_d(f) - Q_{n,d}(f)| = 0, \qquad \forall\, p < p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}),\ \forall\, f \in H(K_d),\ \forall\, K_d \in \mathcal{K}_{d,r,\beta}.$$

For multivariate approximation, we proceed similarly and define the optimal order of convergence as

$$p^{\mathrm{APP}}(\mathcal{K}_{d,r,\beta}) = \sup\Big\{ p \ge 0 \ : \exists\, A = \{A_{n,d}\} \text{ such that } \forall\, K_d \in \mathcal{K}_{d,r,\beta} \ \ \forall\, f \in H(K_d) \ \lim_{n\to\infty} n^p \,\|\mathrm{APP}_d(f) - A_{n,d}(f)\|_{L_2(D)} = 0 \Big\}. \tag{6.2}$$

[1] In fact, we can even consider more general quadratures of the form $Q_{n,d}(f) = \phi(f(x_{n,1}), f(x_{n,2}), \dots, f(x_{n,n}))$ for some nonlinear mapping ϕ with an adaptive choice of points $x_{n,j}$; i.e., $x_{n,j}$ can be an arbitrary function of the already computed function values $f(x_{n,1}), f(x_{n,2}), \dots, f(x_{n,j-1})$.

We say that $A = \{A_{n,d}\}$ is a *universal* approximation algorithm for the class $\mathcal{K}_{d,r,\beta}$ if

$$\lim_{n\to\infty} n^p \, \|\mathrm{APP}_d(f) - A_{n,d}(f)\|_{L_2(D)} \;=\; 0,$$
$$\forall\, p < p^{\mathrm{APP}}(\mathcal{K}_{d,r,\beta}),\ \forall\, f \in H(K_d),\ \forall\, K_d \in \mathcal{K}_{d,r,\beta}.$$

From (3.6) it follows that $p^{\mathrm{APP}}(\mathcal{K}_{d,r,\beta}) \;\le\; p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta})$.

We are now ready to prove the main theorem of this paper.

Theorem 1. *The nested quadrature* $Q^{\mathrm{nes}} = \{Q^{\mathrm{nes}}_{n,d}\}$ *and the nested approximation algorithm* $A^{\mathrm{nes}} = \{A^{\mathrm{nes}}_{n,d}\}$ *defined by* (5.6) *are universal quadrature and universal approximation algorithms for the class* $\mathcal{K}_{d,r,\beta}$*. Their optimal rates of convergence are*

$$p^{\mathrm{APP}}(\mathcal{K}_{d,r,\beta}) \;=\; p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}) \;=\; \frac{r+\beta}{d}.$$

Proof. From the considerations in §5, we know that the errors of $Q^{\mathrm{nes}}_{n,d}(f)$ and $A^{\mathrm{nes}}_{n,d}(f)$ are of order $n^{-(r+\beta)/d}$ for all $f \in H(K_d)$ and all $K_d \in \mathcal{K}_{d,r,\beta}$. Hence, $p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}) \ge p^{\mathrm{APP}}(\mathcal{K}_{d,r,\beta}) \ge (r+\beta)/d$. Thus, it remains to show that $p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}) \le (r+\beta)/d$.

The proof of this estimate will use a technique due to Bakhvalov [4] for finding lower bounds for multivariate integration for the class $C^{r,\beta}(D)$; see also [10], as well as the result of Trojan [23] for analyzing the asymptotic setting for linear problems. See also [22, Chapter 10].

Define $g(x) = (x(1-x))^{r+1}$ for $x \in [0,1]$ and $g(x) = 0$ for $x \in \mathbb{R} \setminus [0,1]$. Clearly $g \in C^{r,\beta}(\mathbb{R})$, and the support of g is $[0,1]$. We partition the domain $D = [0,1]^d$ as before into k^d sub-cubes $D_{j,k}$ of side-length $1/k$. For each of the k^d sub-cubes $D_{k,j}$, let $t_j = [t_{j,1}, t_{j,2}, \ldots, t_{j,d}]$ denote the lower left corner of $D_{k,j}$, and for $x = [x_1, x_2, \ldots, x_d] \in D$ let

$$g_{k,j}(x) \;=\; k^{-(r+\beta)} \prod_{i=1}^{d} g\left(k(x_i - t_{j,i})\right) \quad \text{for } j = 1, 2, \ldots, k^d.$$

Note that the sub-cube $D_{k,j}$ is the support of $g_{k,j}$. For $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_d]$ with $|\alpha| \le r$ we have

$$g^{(\alpha)}_{k,j}(x) \;=\; k^{-(r+\beta-|\alpha|)} \prod_{i=1}^{d} g^{(\alpha_i)}(k(x_i - t_{j,i})).$$

Thus, $|g^{(\alpha)}_{k,j}(x)| = O(1)$ with the factor in the O-notation independent of x, k, and j.

Choose a multi-index α with $|\alpha| = r$. Suppose that x and y lie in the same sub-cube, say $D_{k,j}$. We have

$$g^{(\alpha)}_{k,j}(x) - g^{(\alpha)}_{k,j}(y) \;=\; k^{-\beta}\left(\prod_{i=1}^{d} g^{(\alpha_i)}(k(x_i - t_{j,i})) - \prod_{i=1}^{d} g^{(\alpha_i)}(k(y_i - t_{j,i}))\right).$$

It can be checked by induction on d that

$$\prod_{i=1}^{d} a_i - \prod_{i=1}^{d} b_i \;=\; \sum_{j=1}^{d} a_1 a_2 \cdots a_{j-1}(a_j - b_j) b_{j+1} b_{j+2} \cdots b_d$$

for any choice of numbers a_i and b_i. Using this identity, the fact that all $g^{(\alpha_i)}(y)$ are of order 1 for all y, and using the property that the $g^{(\alpha_i)}$ satisfy the Lipschitz condition for $\alpha_i \le r$, we have

$$\begin{aligned} g_{k,j}^{(\alpha)}(x) - g_{k,j}^{(\alpha)}(y) &= O\left(k^{1-\beta}\sum_{j=1}^{d}|x_j - y_j|\right) = O\left(k^{1-\beta}\|x-y\|_1^{1-\beta}\|x-y\|_1^{\beta}\right) \\ &= O\left(k^{1-\beta}k^{-(1-\beta)}\|x-y\|_1^{\beta}\right) = O(\|x-y\|_1^{\beta}). \end{aligned}$$

Here $\|x\|_1 = \sum_{j=1}^{d}|x_j|$ is the l_1-norm.

Keeping in mind the fact that $|\alpha| = r$, we now suppose that x and y belong to different sub-cubes. Let $x \in D_{k,j}$. Then there is a point z lying on the boundary of the sub-cube $D_{k,j}$ and belonging to the interval $\{\, cx + (1-c)y : c \in [0,1] \,\}$ for which $\|x-z\|_1 \le \|x-y\|_1$. Since y is outside of $D_{k,j}$, we have $g_{k,j}^{(\alpha)}(y) = g_{k,j}^{(\alpha)}(z) = 0$ and

$$g_{k,j}^{(\alpha)}(x) - g_{k,j}^{(\alpha)}(y) = g_{k,j}^{(\alpha)}(x) - g_{k,j}^{(\alpha)}(z) = O(\|x-z\|_1) = O\|x-y\|_1^{\beta}).$$

Hence, the Hölder condition holds for the l_1-norm, and since all vector norms are equivalent, this also holds for any other vector norm. Therefore $g_{k,j} \in C^{r,\beta}(D)$ and $\|g_{k,j}\|_{C^{r,\beta}} = O(1)$, with the factor in the O-notation being independent of k and j. Furthermore,

$$\int_D g_{k,j}(t)\,dt = \int_{D_{k,j}} g_{k,j}(t)\,dt = k^{-(d+r+\beta)}\int_D g(t)\,dt = \Theta(k^{-(d+r+\beta)}).$$

We now partition the cube D into sub-cubes $D_{k,j}$ with $k = 2^i$ for $i = 1, 2, \ldots$. Let

$$G = \{\, g_{2^i,j} \; : \; j = 1, 2, \ldots, 2^{id}, i = 1, 2, \ldots \} \tag{6.3}$$

be the set of functions $g_{2^i,j}$ obtained by all such partitions of D. We claim that the elements of the set G are linearly independent. To this end, take arbitrary numbers $c_{i,j}$ for which

$$\sum_{i=1}^{\infty}\sum_{j=1}^{2^{id}} c_{i,j}g_{2^i,j}(x) = 0 \qquad \text{for all } x \in D.$$

We need to show that this can happen only for $c_{i,j} = 0$. Let

$$\Gamma_i = \bigcup_{j=1}^{2^{id}} \partial D_{2^i,j}$$

be the union of the boundary points of all sub-cubes $D_{2^i,j}$. Clearly, Γ_i is a proper subset of Γ_{i+1}. For $i^* = 1, 2, \ldots$ and $j^* = 1, 2, \ldots, 2^{id}$, choose a point $x \in \Gamma_{i^*+1}$ belonging to the interior of $D_{2^{i^*},j^*}$. Then $x \in \Gamma_i$ for all $i \ge i^* + 1$. Since the support of each $g_{2^i,j}$ is the interior of $D_{2^i,j}$, we have $g_{2^i,j}(x) = 0$ for all $i \ge i^* + 1$ and all $j = 1, 2, \ldots, 2^{id}$. For such x, we obtain

$$0 = \sum_{i=1}^{\infty}\sum_{j=1}^{2^{id}} c_{i,j}g_{2^i,j}(x) = \sum_{i=1}^{i^*}\sum_{j=1}^{2^{id}} c_{i,j}g_{2^i,j}(x).$$

Since the supports of $g_{2^i,j}$ are disjoint, the sum over j may have at most one nonzero term for $j = j(i)$ with $x \in D_{2^i,j(i)}$. Thus

$$0 \;=\; \sum_{i=1}^{i^*} c_{i,j(i)} g_{2^i,j(i)}(x). \tag{6.4}$$

If $i^* = 1$, then $j(1) = j^*$ and $c_{1,j^*} = 0$. Since j^* can be any integer from $[1, 2^d]$ we see that all $c_{1,j} = 0$. Then we use induction on i^* and assume that all $c_{i,j} = 0$ for $i = 1, 2, \ldots, i^* - 1$ and $j = 1, 2, \ldots, 2^{id}$. From formula (6.4), we conclude that $c_{i^*,j^*} g_{2^{i^*},j^*}(x) = 0$. Thus, $c_{i^*,j^*} = 0$, which holds for all $j^* = 1, 2, \ldots, 2^{i^*d}$. Hence the $c_{i,j}$ are all zero, as claimed.

Now consider an arbitrary quadrature $Q = \{Q_{n,d}\}$, and let $x_{n,1}, x_{n,2}, \ldots, x_{n,n}$ be the sample points used by $Q_{n,d}$. Let us assume that for some positive p we have

$$\lim_{n\to\infty} n^p \left|\mathrm{INT}_d(f) - Q_{n,d}(f)\right| \;=\; 0, \quad \forall\, f \in H(K_d),\ \forall\, K_d \in \mathcal{K}_{d,r,\beta}.$$

We want to show that $p \le (r + \beta)/d$.

As before, we can switch to a nested quadrature $\bar{Q}_{n,d}$ that uses the first n points of the sequence $\{x_i\}$ of sample points given by

$$x_{1,1}, x_{2,1}, x_{2,2}, \ldots, x_{2^m,1}, x_{2^m,2}, \ldots, x_{2^m,2^m}, \ldots,$$

with $\bar{Q}_{n,d} = Q_{2^m,d}$ for $n \in [2^m, 2^{m+1})$. Since $n = \Theta(2^m)$, we also have

$$\lim_{n\to\infty} n^p \left|\mathrm{INT}_d(f) - \bar{Q}_{n,d}(f)\right| \;=\; 0, \quad \forall\, f \in H(K_d),\ \forall\, K_d \in \mathcal{K}_{d,r,\beta}.$$

For $i = 1, 2, \ldots,$ define $n_i = 2^{id-1}$. For $j = 1, 2, \ldots, n_i$, the sample point x_j used by the algorithm $\bar{Q}_{n_i,d}$ belongs to $D_{2^i,l(j)}$ for some index $l(j) \in [1, 2^{id}]$. Let

$$J_i \;=\; \{1, 2, \ldots, 2^{id}\} \setminus \{l(1), l(2), \ldots, l(n_i)\}.$$

Clearly $|J_i| \in [2^{id-1}, 2^{id})$. Define

$$g_{2^i}(t) \;=\; \sum_{j \in J_i} g_{2^i,j}(t)$$

as the sum of functions $g_{2^i,j}$ whose supports $D_{2^i,j}$ do not contain the sample points used by $\bar{Q}_{n_i,d}$. Therefore[2] $g_{2^i}(x_j) = 0$ for $j = 1, 2, \ldots, n_i$, and $\bar{Q}_{n_i,d}(g_{2^i}) = 0$. We also have

$$g_{2^i}(t) \;=\; O\left(2^{-i(r+\beta)}\right) \tag{6.5}$$

and

$$\begin{aligned}\int_D g_{2^i}(t)\,dt = &\ \sum_{j\in J_i} \int_D g_{2^i,j}(t)\,dt \\ = &\ 2^{-i(d+r+\beta)} |J_i| \int_D g(t)\,dt = \ \Theta(2^{-i(r+\beta)}) \;=\; \Theta(n_i^{-(r+\beta)/d}).\end{aligned}$$

The function g_{2^i} also belongs to $C^{r,\beta}$, and $\|g_{2^i}\|_{C^{r,\beta}(D)} = O(1)$ with the factor in the O-notation independent of i.

Let δ be an arbitrary positive number. Consider the sequence of functions $\{2^{-i\delta} g_{2^i}\}$. These functions are linearly independent since they are from the set

[2]For a more general form of quadrature, it is enough to work with $x_{n,j}$ for which we obtain the zero function values $f(x_{n,j}) = 0$. Then $\bar{Q}_{n,d}(g_n) = \phi(0, 0, \ldots, 0)$. It is easy to see that the best choice of $\phi(0, 0, \ldots, 0)$ is zero, since otherwise the error for $\mathrm{sign}(-\phi(0, 0, \ldots, 0))g_n$ is even larger.

$G = \{g_{2^i,j}\}$ (see (6.3)) of linearly independent functions as shown before. Due to (6.5), we find that

$$\sum_{i=1}^{\infty} \left(2^{-i\delta} g_{2^i}(t)\right)^2 = O\left(\sum_{i=1}^{\infty} 2^{-2i(r+\beta+\delta)}\right) < \infty.$$

As shown in §2, the Hilbert space

$$H = \operatorname{span}\left\{ 2^{-\delta} g_2, 2^{-2\delta} g_4, \ldots, 2^{-i\delta} g_{2^i}, \ldots \right\},$$

with inner product $\left\langle 2^{-i\delta} g_{2^i}, 2^{-j\delta} g_{2^j} \right\rangle_H = \delta_{i,j}$, has the reproducing kernel

$$K_d(x, y) = \sum_{i=1}^{\infty} 2^{-i\delta} g_{2^i}(x)\, 2^{-i\delta} g_{2^i}(y) \qquad \text{for all } x, y \in D.$$

We now show that $K_d \in \mathcal{K}_{d,r,\beta}$. Indeed, for $|\alpha| \le r$ we have

$$K_d^{(\alpha,\alpha)}(x, y) = \sum_{i=1}^{\infty} 2^{-i\delta} g_{2^i}^{(\alpha)}(x)\, 2^{-i\delta} g_{2^i}^{(\alpha)}(y),$$

where the last series is convergent because

$$2^{-i\delta} g_{2^i}^{(\alpha)}(x)\, 2^{-i\delta} g_{2^i}^{(\alpha)}(y) = O(2^{-2i(r+\beta-|\alpha|+\delta)})$$

with $r+\beta-|\alpha|+\delta \ge \delta > 0$. Hence, $K_d \in C^{r,r}(D\times D)$, and (4.1) holds. Furthermore, for $|\alpha| = r$ we have

$$K_d^{(\alpha,\alpha)}(x, x) - 2K_d^{(\alpha,\alpha)}(x, y) + K_d^{(\alpha,\alpha)}(y, y) = \sum_{i=1}^{\infty} 2^{-2i\delta} \left(g_{2^i}^{(\alpha)}(x) - g_{2^i}^{(\alpha)}(y)\right)^2.$$

Since $g_{2^i}^{(\alpha)}$ satisfies the Hölder condition with the exponent β and a constant, say c, of order 1 independent of i, the condition (4.2) holds with the exponent 2β and $L_{K_d} \le c/(1 - 2^{-2\delta})$. Hence, $K_d \in \mathcal{K}_{d,r,\beta}$.

For the space $H(K_d)$ we first compute the worst case error $e^{\text{wor}}(\bar{Q}_{n,d})$ of the quadrature $\bar{Q}_{n,d}$; see (3.1). It is well known (see, e.g., [22, p. 76]) that

$$e^{\text{wor}}(Q_{n,d}) \ge e_{n,d} := \sup_{\substack{f \in H(K_d) \\ \|f\|_{H(K_d)} \le 1 \\ f(x_1) = \cdots = f(x_n) = 0,}} |\text{INT}_d(f)|. \tag{6.6}$$

For a given integer n, take an integer i such that

$$2^{(i-1)d} < 2n \le 2^{id} \qquad \text{and} \qquad f = 2^{-i\delta} g_{2^i}.$$

Then $f \in H(K_d)$ and $\|f\|_{H(K_d)} = 1$. Furthermore, by the construction of the function g_{2^i} and since $n \le n_i = 2^{id-1}$, we have $f(x_j) = 0$ for $j = 1, 2, \ldots, n$. We also have $2^{-d} n_i \le n \le n_i$ and $n = \Theta(n_i)$. Finally,

$$e_{n,d} \ge \text{INT}_d(f) = \Theta\left(2^{-i(\delta+r+\beta)}\right) = \Theta(n^{-(\delta+r+\beta)/d}).$$

We now apply Trojan's theorem (see [22, p. 384]), which states that for any positive δ and for any nested quadrature $Q = \{Q_{n,d}\}$, the set

$$\left\{ f \in H(K_d) : \lim_{n\to\infty} \frac{|\text{INT}_d(f) - Q_{n,d}(f)|}{n^{-\delta/d} e_{n,d}} = 0 \right\}$$

is nowhere dense. We apply this theorem to the nested quadrature $\bar{Q}_{n,d}$. Hence, there exists a function f in $H(K_d)$ for which $|\text{INT}_d(f) - \bar{Q}_{n,d}(f)|$ does not go to zero faster than $n^{-(2\delta+r+\beta)/d}$. Therefore $p \le (2\delta+r+\beta)/d$. Since the positive number δ can be chosen arbitrarily small, we conclude that $p \le (r+\beta)/d$. Since $\bar{Q} = \{\bar{Q}_{n,d}\}$ is an arbitrarily chosen nested quadrature, we have $p^{\mathrm{INT}}(\mathcal{K}_{d,r,\beta}) \le (r+\beta)/d$, as claimed. This completes the proof. $\square$

7. Nonuniversal algorithms

Universal algorithms work for all Hilbert spaces $H(K_d)$ for $K_d \in \mathcal{K}_{r,d,\beta}$, and their optimal rate of convergence is $(r+\beta)/d$. We now want to study how this optimal rate can be improved if we allow nonuniversal algorithms that depend on the specific kernel K_d; i.e., we consider algorithms of the form $\sum_{i=1}^n b_{n,i} f(x_{n,i})$ with $b_{n,i}$ and $x_{n,i}$ depending on K_d. Here, $b_{n,i} = a_{n,i} \in \mathbb{R}$ for multivariate integration, and $b_{n,i} = w_{n,i} \in L_2(D)$ for multivariate approximation. Let

$$\begin{aligned} p^{\mathrm{nonuni-INT}}(\mathcal{K}_{d,r,\beta}) = \sup \Big\{ p \ge 0 \ :\ &\forall\, K_d \in \mathcal{K}_{d,r,\beta} \ \ \exists\, Q = \{Q_{n,d}\} \\ &\text{such that } \ \forall\, f \in H(K_d) \\ &\lim_{n\to\infty} n^p \left|\mathrm{INT}_d(f) - Q_{n,d}(f)\right| = 0 \Big\}, \\ p^{\mathrm{nonuni-APP}}(\mathcal{K}_{d,r,\beta}) = \sup \Big\{ p \ge 0 \ :\ &\forall\, K_d \in \mathcal{K}_{d,r,\beta} \ \ \exists\, A = \{A_{n,d}\} \\ &\text{such that } \ \forall\, f \in H(K_d) \\ &\lim_{n\to\infty} n^p \left\|\mathrm{APP}_d(f) - A_{n,d}(f)\right\|_{L_2(D)} = 0 \Big\} \end{aligned}$$

denote the optimal rates of convergence that can be achieved by using nonuniversal algorithms for multivariate integration and approximation for the spaces $H(K_d)$ with $K_d \in \mathcal{K}_{d,r,\beta}$. As before, (3.6) implies that

$$p^{\mathrm{nonuni-APP}}(\mathcal{K}_{d,r,\beta}) \le p^{\mathrm{nonuni-INT}}(\mathcal{K}_{d,r,\beta}).$$

Theorem 2. *We have*

$$\begin{aligned} p^{\mathrm{nonuni-INT}}(\mathcal{K}_{d,r,\beta}) &= \frac{1}{2} + \frac{r+\beta}{d}, \\ p^{\mathrm{nonuni-APP}}(\mathcal{K}_{d,r,\beta}) &= a + \frac{r+\beta}{d} \quad \textit{with} \ \ a \in \left[\frac{1}{4+4(r+\beta)/d}, \frac{1}{2}\right]. \end{aligned}$$

Proof. Let $\tau := 1/2 + (r+\beta)/d$. We first consider multivariate integration. Using known results, we can easily see that $p^{\mathrm{nonuni-INT}}(\mathcal{K}_{d,r,\beta}) \ge \tau$. Indeed, it was proved in [17] (see also [15, p. 136]) that for any $K_d \in \mathcal{K}_{d,r,\beta}$ there are sample points $x_{n,1}, x_{n,2}, \dots, x_{n,n}$ from D and real numbers $a_{n,1}, a_{n,2}, \dots, a_{n,n}$ all depending on K_d such that the quadrature $Q_{n,d}(f) = \sum_{j=1}^n a_{n,j} f(x_{n,j})$ has a worst case error of the order $n^{-\tau}$. The proof of this fact is nonconstructive. Its idea is as follows: the worst case error of a quadrature for the unit ball of $H(K_d)$ is the same as the average case error of the same quadrature for the space of continuous functions equipped with a zero mean Gaussian measure with the covariance function K_d. We stress that the covariance function and the reproducing kernel are the same. Then the average case error is bounded by roughly $n^{-1/2}$ times the average case error for the approximation problem in the L_2 norm for the same space and the same measure; see [25]. Furthermore we know from [28] that the average case error for this approximation problem is bounded by the worst case error of the approximation problem for the unit ball of $H(K_d)$ in the L_∞ norm. Finally, since in our case $H(K_d) \subset C^{r,\beta}(D)$ for $K \in \mathcal{K}_{d,r,\beta}$, and since the approximation problem in the L_∞ norm has the rate $n^{-(r+\beta)/d}$ (see [10]), the result follows.

We now prove that $p^{\text{nonuni}-\text{INT}}(\mathcal{K}_{d,r,\beta}) \le \tau$. Again, it is known that for even r (see [16] and also [15, p. 153]), and for an arbitrary integer r and $\beta \in (0,1)$ (see [15, p. 140]), there is a kernel K_d from $\mathcal{K}_{d,r,\beta}$ for which the worst case error of any quadrature using n sample points is of order $n^{-\tau}$. Applying Trojan's theorem we conclude that $p^{\text{nonuni}-\text{INT}}(\mathcal{K}_{d,r,\beta}) \le \tau$.

To show the same upper bound for all integers r and $\beta \in [0,1]$, we use the construction from the proof of Theorem 1. For an arbitrary positive δ, we take the set G of linearly independent functions defined by (6.3) and form the Hilbert space

$$H = \text{span}\left\{ \eta_{i,j} := 2^{-i\delta} g_{2^i,j} \ : \ i = 1,2,\dots, \ j = 1,2,\dots,2^{id} \right\} \tag{7.1}$$

with orthonormal $\eta_{i,j}$. The reproducing kernel of H is

$$K_d(x,y) = \sum_{i=1}^{\infty} \sum_{j=1}^{2^{id}} \eta_{i,j}(x)\eta_{i,j}(y) \qquad \text{for all } x,y \in D.$$

We emphasize that the inner sum over j has at most one nonzero term, and therefore

$$\sum_{i=1}^{\infty} \sum_{j=1}^{2^{id}} \eta_{i,j}^2(x) = O\left(\sum_{i=1}^{\infty} 2^{-2i(\delta+r+\beta)} \right) < \infty.$$

As before, we can show that $K_d \in \mathcal{K}_{d,r,\beta}$.

Now take an arbitrary quadrature $Q_{n,d}(f) = \sum_{k=1}^{n} a_{n,k} f(x_{n,k})$. We choose an integer i such that $2^{(i-1)d} < 2n \le 2^{id}$. The sample point $x_{n,k}$ belongs to a sub-cube $D_{2^i,j(k)}$ for some index $j(k) \in [1, 2^{id}]$. Let

$$J_n = \{1,2,\dots,2^{id}\} \setminus \{j(1), j(2), \dots, j(n)\}.$$

Then $|J_n| \in [2^{id} - n, 2^{id} - 1] \subset [2^{id-1}, 2^{id} - 1]$, and obviously $|J_n| = \Theta(n)$. Consider the function

$$f = \frac{1}{|J_n|^{1/2}} \sum_{j \in J_n} \eta_{i,j}.$$

Since the $\eta_{i,j}$ are orthonormal, we have $\|f\|_{H(K_d)} = 1$. Furthermore, $f(x_{n,k}) = 0$ for $k = 1,2,\dots,n$, since the supports of the functions $\eta_{i,j}$ for $j \in J_n$ do not contain the sample points $x_{n,k}$. We also have

$$\text{INT}_d(f) = \frac{1}{|J_n|^{1/2}} |J_n| \, 2^{-i\delta} \, 2^{-i(d+r+\beta)} \int_D g(t)\, dt = \Theta\left(\left(\frac{1}{n} \right)^{1/2 + (r+\beta)/d - \delta/d} \right).$$

From (6.6) we see that $e^{\text{wor}}(Q_{n,d}) \ge e_{n,d} = \Theta(n^{-\tau-\delta/d})$. Applying Trojan's theorem and letting δ tend to zero, we conclude that $p^{\text{nonuni}-\text{INT}}(\mathcal{K}_{d,r,\beta}) \le \tau$, as claimed.

We now turn to multivariate approximation. Obviously, $p^{\text{nonuni}-\text{APP}}(\mathcal{K}_{d,r,\beta}) \le \tau$ since multivariate approximation is not easier than multivariate integration. This proves that $a \le 1/2$. We now show that

$$p^{\text{nonuni}-\text{APP}}(\mathcal{K}_{d,r,\beta}) \ge \tau^2/(\tau + 1/2),$$

which is equivalent to $a \ge 1/(4 + 4(r+\beta)/d)$.

We know from the proof of Theorem 1 that for any integer n there are sample points $x_i = x_{i,n}$ and functions $w_i = w_{i,n}$ from $L_2(D)$ such that the algorithm $A_n(f) = \sum_{i=1}^n w_i f(x_i)$ approximates f pointwise with error of order $n^{-(r+\beta)/d}$. Thus, for any $f \in H(K_d)$ we have

$$g_{f,n}(x) := f(x) - A_n(f)(x) = O\left(n^{-(r+\beta)/d} \|f\|_{H(K_d)}\right),$$

with the factor in the O-notation independent of x, n and f. Obviously, $g_{f,n} \in L_2(D)$. Observe that $g_{f,n}(x) = \langle f, h_n(\cdot, x)\rangle_{H(K_d)}$ with

$$h_n(\cdot, x) = K_d(\cdot, x) - \sum_{i=1}^n w_i(x) K_d(\cdot, x_i) \in H(K_d).$$

Taking $f = h_n(\cdot, x)/\|h_n(\cdot, x)\|_{H(K_d)}$ we conclude that

$$\|h_n(\cdot, x)\|_{H(K_d)} = O\left(n^{-(r+\beta)/d}\right) \qquad \text{for all } x \in D, \tag{7.2}$$

with an O-factor independent of x and n but depending on d, r and β.

To improve the error bound of the algorithm A_n for the space $H(K_d)$, we need to consider the operator W_d defined by (3.3). We denote its eigenpairs by (λ_j, η_j); i.e., $W_d \eta_j = \lambda_j \eta_j$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, and $\langle \eta_i, \eta_j\rangle_{H(K_d)} = \delta_{i,j}$. Without loss of generality we assume that λ_i is positive, and set $\zeta_i = \lambda_i^{-1/2} \eta_i$. Then for $f \in H(K_d)$ we have

$$\begin{aligned} \langle f, \eta_i\rangle_{H(K_d)} \eta_i &= \frac{1}{\lambda_i} \langle f, W_d \eta_i\rangle_{H(K_d)} \eta_i \\ &= \frac{1}{\lambda_i} \langle \mathrm{APP}_d f, \mathrm{APP}_d \eta_i\rangle_{L_2(D)} \eta_i = \langle f, \zeta_i\rangle_{L_2(D)} \zeta_i. \end{aligned} \tag{7.3}$$

Putting $f = \eta_j$, we conclude from the known facts that

$$\langle \eta_i, \eta_j\rangle_{L_2(D)} = \lambda_i \delta_{i,j} \qquad \text{and} \qquad \langle \zeta_i, \zeta_j\rangle_{L_2(D)} = \delta_{i,j}. \tag{7.4}$$

We now show that

$$\lambda_n = O(n^{-2\tau}) \qquad \text{for all } n. \tag{7.5}$$

It is known (see, e.g., [22, p. 234]) that $\left(\sum_{j=n+1}^\infty \lambda_i\right)^{1/2}$ is the minimal average case error of algorithms (using at most n linear evaluations of f) for the multivariate approximation problem defined on the space of continuous functions equipped with a zero mean Gaussian measure with the covariance function K_d. It is proved in [17] (see also [15, p. 135]) that the average case error for algorithms using at most n function values is of order $n^{-(r+\beta)/d}$. Therefore,

$$n \lambda_{2n} \leq \sum_{i=n+1}^\infty \lambda_i = O\left(n^{-2(r+\beta)/d}\right).$$

This yields $\lambda_{2n} = O(n^{-(1+2(r+\beta)/d)}) = O(n^{-2\tau})$, as claimed.

Now let m be an integer, which will be specified later, and define the projection

$$P_m f = \sum_{i=1}^m \langle f, \zeta_i\rangle_{L_2(D)} \zeta_i \qquad \text{for all } f \in L_2(D).$$

Let $t_1, t_2, \ldots, t_n$ be sample points from D that will be also specified later, and define the algorithm

$$A_{2n,m}(f) \;=\; P_m A_n(f) \,+\, B_{n,m}(f),$$

where

$$B_{n,m} \;=\; \sum_{i=1}^{m} \left(\frac{1}{n} \sum_{j=1}^{n} g_{f,n}(t_j)\, \zeta_i(t_j) \right) \zeta_i.$$

Observe that the algorithm $A_{2n,m}$ uses at most $2n$ function values; $f(x_1), f(x_2), \ldots,$ $f(x_n)$ are used by $A_n(f)$ and $f(t_1), f(t_2), \ldots, f(t_n)$ are used by $B_{n,m}$.

Let $f \in H(K_d)$. Then (7.3) implies that $P_m f = \sum_{i=1}^{m} \langle f, \eta_i \rangle_{H(K_d)} \eta_i$. Therefore

$$f \;=\; \sum_{i=1}^{\infty} \langle f, \eta_i \rangle_{H(K_d)} \eta_i \;=\; P_m f + P_m^{\perp} f,$$

where $P_m^{\perp} f = \sum_{i=m+1}^{\infty} \langle f, \eta_i \rangle_{H(K_d)} \eta_i$. The relations (7.4) and (7.5) yield

$$\|P_m^{\perp} f\|_{L_2(D)}^2 \;=\; \sum_{i=m+1}^{\infty} \lambda_i \langle f, \eta_i \rangle_{H(K_d)}^2 \;\le\; \lambda_{m+1} \|f\|_{H(K_d)}^2 \;=\; O\left(m^{-2\tau} \|f\|_{H(K_d)}^2 \right).$$

We now estimate the error of $A_{2n,m}$ for the space $H(K_d)$ with $K_d \in \mathcal{K}_{d,r,\beta}$. Clearly,

$$f - A_{2n,m}(f) \;=\; e_{n,m}(f) \,+\, P_m^{\perp} f,$$

where

$$e_{n,m}(f) \;=\; P_m \left(f - A_n(f) \right) - B_{n,m}(f).$$

We have

$$\begin{aligned} e_{n,m}(f) \;&=\; \sum_{i=1}^{m} \left(\langle g_{f,n}, \zeta_i \rangle_{L_2(D)} - \frac{1}{n} \sum_{j=1}^{n} g_{f,n}(t_j) \zeta_i(t_j) \right) \zeta_i \\ &=\; \sum_{i=1}^{m} \left\langle f, \int_D h_n(\cdot, x) \zeta_i(x)\, dx - \frac{1}{n} \sum_{j=1}^{n} h_n(\cdot, t_j) \zeta_i(t_j) \right\rangle_{H(K_d)} \zeta_i. \end{aligned}$$

Therefore

$$\|e_{n,m}(f)\|_{L_2(D)}^2 \;\le\; \|f\|_{H(K_d)}^2 \sum_{i=1}^{m} \alpha_i,$$

where

$$\begin{aligned} \alpha_i \;&=\; \left\| \int_D h_n(\cdot, x) \zeta_i(x)\, dx - \frac{1}{n} \sum_{j=1}^{n} h_n(\cdot, t_j) \zeta_i(t_j) \right\|_{H(K_d)}^2 \\ &=\; \int_{D^2} \zeta_i(x) \zeta_i(t) \langle h_n(\cdot, x), h_n(\cdot, t) \rangle_{H(K_d)} \, dx\, dt \\ &\quad -\; \frac{2}{n} \sum_{j=1}^{n} \zeta_i(t_j) \int_D \zeta_i(x) \langle h_n(\cdot, x), h_n(\cdot, t_j) \rangle_{H(K_d)} \, dx \\ &\quad +\; \frac{1}{n^2} \sum_{k,j=1}^{n} \zeta_i(t_k) \zeta_i(t_j) \langle h_n(\cdot, t_k), h_n(\cdot, t_j) \rangle_{H(K_d)}. \end{aligned}$$

Observe that $\alpha_i = \alpha_i(t_1, t_2, \ldots, t_n)$. We take the sample points t_j as independent random variables distributed uniformly over D and compute the average value of $\alpha(t_1,t_2,\dots,t_n) := \sum_{i=1}^m \alpha_i(t_1,t_2,\dots,t_n)$. Using the same analysis as for the classical Monte Carlo method (see, e.g., [10, 15, 22]), we obtain

$$\int_{D^n} \alpha(t_1,t_2,\dots,t_n)\,dt_1\cdots dt_n$$
$$= \frac{1}{n}\sum_{i=1}^m \biggl(\int_D \zeta_i^2(x)\,\|h_n(\cdot,x)\|^2_{H(K_d)}\,dx$$
$$- \int_{D^2} \zeta_i(x)\zeta_i(t)\,\langle h_n(\cdot,x),h_n(\cdot,t)\rangle_{H(K_d)}\,dx\,dt \biggr).$$

Since $\int_{D^2} \zeta_i(x)\zeta_i(t)\,\langle h_n(\cdot,x),h_n(\cdot,t)\rangle_{H(K_d)}\,dx\,dt = \|\int_D \zeta_i(x)h_n(\cdot,x)\|^2_{H(K_d)} \ge 0$, we obtain

$$\int_{D^n} \alpha(t_1,t_2,\dots,t_n)dt_1\cdots dt_n \le \frac{1}{n}\sum_{i=1}^m \int_D \zeta_i^2(x)\,\|h_n(\cdot,x)\|^2_{H(K_d)}\,dx.$$

From (7.2) and the second part of (7.4) we obtain

$$\int_{D^n} \alpha(t_1,t_2,\dots,t_n)dt_1\cdots dt_n = O\left(\frac{m}{n^{1+2(r+\beta)/d}}\right).$$

From the mean value theorem applied to the integral of α, we conclude that there are sample points $t_1,t_2,\dots,t_n$ which may depend on K_d, for which

$$\alpha(t_1,t_2,\dots,t_n) = O\left(\frac{m}{n^{1+2(r+\beta)/d}}\right).$$

Combining all these estimates and using the algorithm $A_{2n,m}$ with the sample points from the mean value theorem, we obtain

$$\|f - A_{2n,m}(f)\|_{L_2(D)} \le \|e_{n,m}(f)\|_{L_2(D)} + \|P_m^\perp f\|_{L_2(D)}$$
$$= O\left(\left(m^{1/2}\left(\frac{1}{n}\right)^\tau + \left(\frac{1}{m}\right)^\tau\right)\|f\|_{H(K_d)}\right).$$

We choose m to minimize the expression in the O-notation. Then $m=\Theta\left(n^{\tau/(\tau+1/2)}\right)$ and

$$\|f - A_{2n,m}(f)\|_{L_2(D)} = O\left(\left(\frac{1}{n}\right)^{\tau^2/(\tau+1/2)}\|f\|_{H(K_d)}\right). \tag{7.6}$$

This proves that $p^{\text{nonuni}-\text{APP}}(\mathcal{K}_{d,r,\beta}) \ge \tau^2/(\tau+1/2)$, as claimed, and completes the proof. $\square$

We now comment on Theorem 2. For multivariate integration, the optimal rate of convergence for nonuniversal algorithms is $1/2$ larger than for universal algorithms. As already mentioned in the Introduction, this is important when d is large relative to $r+\beta$.

For multivariate approximation, we only presented bounds on the optimal rate of convergence for nonuniversal algorithms. Weaker bounds on $p^{\text{nonuni}-\text{APP}}(\mathcal{K}_{d,r,\beta})$ can be concluded from the known results. For example,

$$p^{\text{nonuni}-\text{APP}}(\mathcal{K}_{d,r,\beta}) \ge \tau/(2\tau+1)$$

follows from [26, Corollary 1] and (7.5). Note that our bound $\tau^2/(\tau+1/2)$ is larger than $\tau/(2\tau+1)$ for $\tau > 1/2$, which holds iff $r+\beta > 0$. In fact, more can be said about the optimal rate of convergence of nonuniversal algorithms using a proof technique similar to that in [8], where this problem is studied in the average case setting. We leave further bounds on $p^{\text{nonuni-APP}}(\mathcal{K}_{d,r,\beta})$ for future research.

It is easy to check that $\tau \geq 1/2$ implies

$$p^{\text{nonuni-APP}}(\mathcal{K}_{d,r,\beta}) \geq \frac{1}{4}.$$

Altogether, the optimal rate of convergence of nonuniversal algorithms cannot be smaller than $1/2$ for multivariate integration and $1/4$ for multivariate approximation. This is in sharp contrast with the optimal rate of convergence of universal algorithms that can be arbitrarily small depending on the values of d, r and β.

On the other hand, if we consider the favorable situation where the sum of the smoothness parameters $r+\beta$ is much larger than d, then $\tau/(\tau+1/2) \approx 1$ and

$$\begin{aligned} p^{\text{nonuni-APP}}(\mathcal{K}_{d,r,\beta}) &\approx p^{\text{nonuni-INT}}(\mathcal{K}_{d,r,\beta}) \\ &\approx p^{\text{APP}}(\mathcal{K}_{d,r,\beta}) = p^{\text{INT}}(\mathcal{K}_{d,r,\beta}) = \frac{r+\beta}{d}. \end{aligned}$$

In this case, the difference between the optimal rates of convergence between universal and nonuniversal algorithms for multivariate integration and approximation disappears.

8. Smooth product kernels

So far, we have studied kernels K_d with given smoothness whose structure is unspecified. In this section we assume additional knowledge about the kernel, namely that K_d is of product form. To be precise, we consider

$$K_d(x,y) = \prod_{j=1}^{d} K_{r_j,\beta_j}(x_j,y_j), \tag{8.1}$$

where the kernel $K_{r_j,\beta_j} : [0,1]^2 \to \mathbb{R}$ corresponds to the Hilbert space $H(K_{r_j,\beta_j})$ of univariate functions $f : [0,1] \to \mathbb{R}$. As before, to guarantee some smoothness of functions from $H(K_d)$, we assume that for each $j = 1,2,\ldots,d$, we have $K_{r_j,\beta_j} \in \mathcal{K}_{1,r_j,\beta_j}$ for some nonnegative integer r_j and some real $\beta_j \in [0,1]$. This means that the space $H(K_d)$ is the tensor product of spaces $H(K_{r_j,\beta_j})$ of univariate functions. If $\text{span}\{\eta_{j,1},\eta_{j,2},\ldots\}$ is an orthonormal basis of $H(K_{r_j,\beta_j})$, then $\text{span}\{\eta_{\vec{k}}\}$, with $\vec{k} = [k_1,k_2,\ldots,k_d]$ and $\eta_{\vec{k}}(x) = \eta_{1,k_1}(x_1)\eta_{2,k_2}(x_2)\cdots\eta_{d,k_d}(x_d)$, is an orthonormal basis of the space $H(K_d)$. The space $H(K_d)$ consists of functions

$$f(x) = \sum_{k_1,k_2,\ldots,k_d=1}^{\infty} c_{k_1,k_2,\ldots,k_d}\eta_{1,k_1}(x_1)\eta_{2,k_2}(x_2)\cdots\eta_{d,k_d}(x_d) \quad \text{with} \sum_{k_1,k_2,\ldots,k_d=1}^{\infty} c_{k_1,k_2,\ldots,k_d}^2 < \infty.$$

Let $\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}$ denote the class of all such kernels K_d having the product form (8.1) with $K_{r_j,\beta_j} \in \mathcal{K}_{1,r_j,\beta_j}$ for $j = 1,2,\ldots,d$. We study universal and nonuniversal algorithms in the class $\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}$ with fixed $\vec{r} = [r_1,r_2,\ldots,r_d]$ and $\vec{\beta} = [\beta_1,\beta_2,\ldots,\beta_d]$. The optimal rates of convergence are defined as in (6.1) and (6.2), the only difference being that the kernels K_d are now taken from the class

$\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}$. In this way we obtain the optimal rate $p^{\text{prod}-\text{INT}}$ of convergence for universal algorithms and $p^{\text{nonuni}-\text{prod}-\text{INT}}$ for nonuniversal algorithms for multivariate integration, and correspondingly $p^{\text{prod}-\text{APP}}$ and $p^{\text{nonuni}-\text{prod}-\text{APP}}$ for multivariate approximation. As before, $p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \le p^{\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}})$. Using previously known results, as well as the results of this article, it is easy to establish the following theorem.

Theorem 3. *Let* $q_d := \min_{j=1,2,\dots,d}\{r_j + \beta_j\}$. *Then*

$$\begin{aligned} p^{\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) &= q_d, \\ p^{\text{nonuni}-\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) &= \frac{1}{2} + q_d, \\ p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) &= q_d, \\ p^{\text{nonuni}-\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) &= a + q_d \qquad \text{with } a \in \left[\frac{1}{4+4q_d}, \frac{1}{2}\right]. \end{aligned}$$

Proof. For universal algorithms, since

$$p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \le p^{\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}),$$

it is enough to prove that

$$p^{\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \le q_d \qquad \text{and} \qquad p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \ge q_d.$$

Let j^* be such that

$$r_{j^*} + \beta_{j^*} = \min_{j=1,2,\dots,d}\{r_j + \beta_j\}.$$

To show that

$$p^{\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \le q_d,$$

consider an arbitrary algorithm $Q_{n,d}(f) = \sum_{i=1}^n a_{i,n} f(x_{i,n})$ for some $a_{i,n} \in \mathbb{R}$ and $x_{i,n} \in D = [0,1]^d$. We now apply this algorithm to the space $H(K_d)$ with $K_d(x,y) = \prod_{j=1}^d K_{r_j,\beta_j}(x_j,y_j)$, and with an arbitrary $K_{r_{j^*},\beta_{j^*}}$ from $\mathcal{K}_{1,r_{j^*},\beta_{j^*}}$ and $K_{r_j,\beta_j} = 1$ for all $j \neq j^*$. For $j \neq j^*$, we have $H(K_{r_j,\beta_j}) = \text{span}\{1\}$, and $H(K_d)$ consists of functions depending only on the j^*th variable. Hence, $Q_{n,d}(f) = \sum_{i=1}^n a_{i,n} f(x_{i,n,j^*})$, where x_{i,n,j^*} denotes the j^*th component of the sample point $x_{i,n}$, and f is a univariate function from $H(K_{r_{j^*},\beta_{j^*}})$. We now apply Theorem 1 with $d=1$ and conclude that the optimal rate of convergence is at most q_d.

To prove that $q_d \le p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}})$, we observe from §4 that $H(K_{r_j,\beta_j}) \subset C^{r_j,\beta_j}([0,1])$ and that $\|f\|_{C^{r_j,\beta_j}([0,1])} \le c(K_{r_j,\beta_j}, r_j, \beta_j)\|f\|_{H(K_{r_j,\beta_j})}$; see (4.4). Furthermore, from §5 we know there are algorithms $A_{n,1,j}^{\text{nes}}$ for multivariate approximation such that

$$\|\text{APP}_d(f) - A_{n,1,j}^{\text{nes}}(f)\|_{L_2(D)} = O\left(n^{-(r_j+\beta_j)} c(K_{r_j,\beta_j})\|f\|_{H(K_{r_j,\beta_j})}\right).$$

It is now enough to apply Smolyak's algorithm (see [18]), using the univariate algorithms $A_{n,1,j}^{\text{nes}}$ for $j = 1,2,\dots,d$ as its components. It is known that the error of Smolyak's algorithm is of order n^{-q_d} times a logarithmic factor $\log n$ raised to a power that is linear in $d-1$; see, e.g., [27, Remark 2] and the literature cited therein. (Further details on various aspects of Smolyak's algorithm can be found in [6, 7, 11, 12, 21].) This proves that $p^{\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \ge q_d$, as needed.

The bounds on the optimal rates of convergence for nonuniversal algorithms can be shown similarly. For multivariate integration, the bound

$$p^{\text{nonuni}-\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \le \frac{1}{2} + q_d$$

follows from the construction of K_d as above and from Theorem 2. The opposite inequality $p^{\text{nonuni}-\text{prod}-\text{INT}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \ge \frac{1}{2} + q_d$ follows from the application of Smolyak's algorithm for the univariate algorithms with errors of order $n^{-(1/2+r_j+\beta_j)}$ whose existence is established in Theorem 2. For multivariate approximation, we only need to prove that $p^{\text{nonuni}-\text{prod}-\text{APP}}(\mathcal{K}_{\text{prod},d,\vec{r},\vec{\beta}}) \ge 1/(4+4q_d)+q_d$. This again follows from the application of Smolyak's algorithm for the univariate algorithms with errors of order $n^{-(a_j+r+j+\beta_j)}$ with $a_j = 1/(4+4(r_j+\beta_j))$, whose existence is established in Theorem 2. Since

$$\min_{j=1,2,\ldots,d} \left\{ \frac{1}{4+4(r_j+\beta_j)} + r_j + \beta_j \right\} = \frac{1}{4+4q_d} + q_d,$$

the result follows. □

As already mentioned in the Introduction, the number d of variables plays a much weaker role for product kernels. If we assume that

$$\min_{d=1,2,\ldots} \min_{j=1,2,\ldots,d} \{r_j + \beta_j\} > 0,$$

then even the optimal rates of convergence for universal algorithms do *not* suffer from the curse of dimensionality. This shows that the product structure of reproducing kernels is a powerful property that breaks the curse of dimensionality present in the nonproduct case.

9. Conclusion and open problems

In this paper we studied the optimal rate of convergence of universal and nonuniversal algorithms for multivariate integration and approximation. We considered functions from a reproducing kernel Hilbert space $H(K_d)$ with an arbitrary K_d, all of whose partial derivatives up to order r satisfy a Hölder-type condition with exponent 2β. For universal algorithms, the weights and sample points may depend on d, r and β, but are independent of the specific kernel K_d. For nonuniversal algorithms, the weights and sample points may additionally depend on K_d. We proved that for universal algorithms, the optimal rate of convergence is $(r+\beta)/d$ for both multivariate integration and approximation, whereas for nonuniversal algorithms, the optimal rate of convergence is $1/2 + (r+\beta)/d$ for multivariate integration and $a + (r+\beta)/d$ with $a \in [1/(4+4(r+\beta)/d), 1/2]$ for multivariate approximation. Thus, universal algorithms are applicable to wide classes of functions without a priori knowledge of the specific form of reproducing kernels, but they suffer from the curse of dimensionality; i.e., for fixed smoothness r and β, their optimal rate of convergence goes to zero with increasing d. This is the price we have to pay for universality. For nonuniversal algorithms, we know the reproducing kernel of a specific Hilbert space from which the data stem. This knowledge may then be exploited in the design of an algorithm, and leads to an optimal rate of convergence that does not suffer from the curse of dimensionality. Unfortunately, in practical applications, this additional knowledge may often not be available. If $r+\beta$ is large relative to d, then both the universal and the nonuniversal algorithms exhibit approximately the same optimal rates of convergence.

We also considered the case of a kernel K_d having product structure. This case is closely related to spaces of bounded mixed derivatives. Then the optimal rates of convergence of both universal and nonuniversal algorithms for multivariate integration and approximation only depend weakly on d. Furthermore, if the minimal smoothness of product kernels does not deteriorate with d, then the curse of dimensionality is not present. However, we stress that in all cases, the multiplicative factors in the order estimates are unknown functions of d, which means that they could possibly grow exponentially (if not faster).

Finally, let us mention a few open problems.

- We were only able to give bounds on the optimal rate of convergence for nonuniversal algorithms and multivariate approximation. This involves a value a from the interval $[1/(4+4(r+\beta)/d), 1/2]$ in the general case and $[1/(4+4q_d), 1/2]$ with $q_d = \min_{j=1,..,d}\{r_j + \beta_j\}$ in the product kernel case, respectively. Clearly, if we can improve our estimates of a, then we will attain tighter bounds.
- In this paper we restricted ourselves to two multivariate problems, namely integration and approximation. It is natural to study universal (and nonuniversal) algorithms for other practically important problems such as partial differential equations or general linear problems. Such problems are in some cases closely related to multivariate approximation, and we therefore hope that it will be possible to apply our multivariate approximation results here as well.
- As already often mentioned in this article, we do not control the dependence on d of the factors in the order estimates. It is a natural question to also study this question and to check on the optimal order of universal algorithms for which the dependence on d of these factors is only polynomial. This problem is related to the problem of tractability that so far has only been studied for nonuniversal algorithms; see [13] for a survey. Here, it seems natural to also consider universal algorithms for classes of weighted spaces, with limited knowledge about smoothness and weights of reproducing kernels.
- Universality of algorithms may also be studied in the randomized, average case, and probabilistic settings. To illustrate this point, let us stress that we have only considered deterministic algorithms in this article. In the randomized setting, for example, randomized algorithms can be employed, such as the classical Monte Carlo method for multivariate integration. It would be natural to study the optimal rate of convergence for both universal and nonuniversal randomized algorithms in the classes $\mathcal{K}_{d,r,\beta}$ and $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$.
- For multivariate approximation and related problems, it is also reasonable to consider algorithms that use information that is more general than function values. An example is given by algorithms that use arbitrary linear functionals $L_i(f)$ for $i = 1, 2, \ldots, n$ for some finite n. It would be useful to determine the optimal rates of convergence for universal algorithms using this more general information and to compare them to the optimal rates of convergence obtained in this paper for the classes $\mathcal{K}_{d,r,\beta}$ and $\mathcal{K}_{\mathrm{prod},d,\vec{r},\vec{\beta}}$.

These and similar problems concerning universal and nonuniversal algorithms will be the subject of future research.

Acknowledgments

The research reported in this article was initiated when the two authors were guests of Ian Sloan and his group at the UNSW, Sydney, in Spring 2005. The hospitality and the support of the School of Mathematics and the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (MASCOS) at UNSW is greatly acknowledged.

We are grateful to S. Heinrich, E. Novak, K. Ritter, I. Sloan, J. F. Traub, G. W. Wasilkowski, and A. G. Werschulz for valuable comments and suggestions on how to improve the presentation of this paper.

References

[1] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.*, 68, 337–404, 1950. MR0051437 (14:479c)

[2] K. I. Babenko, On the phenomenon of saturation in numerical analysis, *Dokl. Akad. Nauk SSSR*, 241, 505-509, 1978. Translation in *Soviet Math. Dokl.*, 19, 859-863, 1978. MR0504216 (80c:65116)

[3] I. Babuska, Über universal optimale Quadraturformeln, *Appl. Mat.*, 13, 304–338, 338-404, 1968. MR0244680 (39:5994)

[4] N. S. Bakhvalov, On approximate calculation of integrals (in Russian), *Vestnik MGV, Ser. Mat. Mekh. Astron. Fiz.*, 4, 3–18, 1959. MR0111138 (22:2002)

[5] H. Brass, Universal quadrature rules in the space of periodic functions, in Numerical Integration III, eds. H. Brass and G. Hämmerlin, ISNM, vol. 85, Birkhäuser, 16–24, 1988. MR1021519 (91a:65057)

[6] H.-J. Bungartz and M. Griebel, Sparse grids, *Acta Numerica*, 13, 147–269, 2004.

[7] T. Gerstner and M. Griebel, Numerical integration using sparse grids, *Numer. Alg.*, 18, 209-232, 1998. MR1669959 (99m:65042)

[8] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimension, *Adv. in Comput. Math.*, 12, 25–58, 2000. MR1758946 (2001d:65017)

[9] V. Motornyj, On the best quadrature formula of the form $\sum_{k=1}^{n} p_k f(x_k)$ for some classes of periodic differentiable functions, *Ivz. Akad. Nauk USSR Ser. Mat.*, 38, 538–614, 1974.

[10] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lectures Notes in Mathematics, No. 1349, Springer, New York, 1988. MR0971255 (90a:65004)

[11] E. Novak and K. Ritter, High dimensional integration of smooth functions over cubes, *Numerische Mathematik*, 75, 79–97, 1996. MR1417864 (97k:65057)

[12] E. Novak and K. Ritter, The curse of dimension and a universal method for numerical integration, in *Multivariate Approximation and Splines*, G. Nürnberger, J. Schmidt and G. Walz, eds., International Series in Numerical Mathematics, Birkhäuser, Basel, 177-188, 1998. MR1485004 (98g:65022)

[13] E. Novak and H. Woźniakowski, When are integration and discrepancy tractable?, Foundation of Computational Mathematics, Oxford, 1999, R. A. DeVore, A. Iserles, and E. Süli, eds., *Cambridge University Press*, Cambridge, 211-266, 2001. MR1836619 (2002h:68083)

[14] K. Petras, On the universality of the Gaussian formula, *East J. Approx.*, 2, 427–438, 1996. MR1426714 (97k:41050)

[15] K. Ritter, *Average-Case Analysis of Numerical Problems*, Lectures Notes in Mathematics, No. 1733, Springer, New York, 2000. MR1763973 (2001i:65001)

[16] K. Ritter and G. W. Wasilkowski, Integration and L_2-approximation: Average case setting with isotropic Wiener measure for smooth functions, *Rocky Mountain J. Math.*, 26, 1541–1557, 1996. MR1447602 (98c:65212)

[17] K. Ritter, G. W. Wasilkowski and H. Woźniakowski, On multivariate integration for stochastic processes, in Numerical Integration IV, eds. H. Brass and G. Hämmerlin, ISNM, vol. 112, Birkhäuser, 331–347, 1993. MR1248414 (94j:65031)

[18] S. Smolyak, Quadrature and interpolation formulas for tensor product of certain classes of functions, *Soviet Math. Dokl.*, 4, 240-243, Russian original in *Dokl. Akad. Nauk. SSSR*, 148, 1042-1045, 1963.

[19] I. H. Sloan and H. Woźniakowski, When are Quasi-Monte Carlo algorithms efficient for high dimensional integrals, *J. Complexity*, 14, 1–33, 1998. MR1617765 (99d:65384)
[20] V. Temlyakov, On universal cubature formulae, *Soviet Math. Dokl.*, 43, 39–42, 1991. MR1102770 (92e:65030)
[21] V. Temlyakov, *Approximation of Periodic Functions*, Nova Science, New York, 1994, MR1373654 (96j:41001)
[22] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-based Complexity*, Academic Press, New York, 1988. MR0958691 (90f:68085)
[23] J. Trojan, Asymptotic setting for linear problems, unpublished manuscript, available as Chapter 10 in *Information-based Complexity* by J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, Academic Press, 1988.
[24] G. Wahba, *Spline models for observational data*, CBSM-NSF Regional Conf. Ser. Appl. Math., 59, SIAM, Philadelphia. MR1045442 (91g:62028)
[25] G. W. Wasilkowski, Integration and approximation of multivariate functions: average case setting with isotropic Wiener measure, *Bull. Amer. Math. Soc*, 28, 308-314, 1993. Full version *J. Approx. Theory*, 77, 212-227, 1994. MR1184000 (93i:65136); MR1275937 (95c:65234)
[26] G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for weighted approximation, *Journal of FoCM*, 4, 417–438, 2001. MR1857723 (2002g:41052)
[27] G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, *J. Complexity*, 11, 1–56, 1995. MR1319049 (95k:65138)
[28] H. Woźniakowski, Average case complexity of linear multivariate problems, Part I: Theory, Part II: Applications, *J. Complexity*, 8, 337–372, 373–392, 1992. MR1195257 (94a:65076b); MR1195258 (94a:65076c)

Institut für Numerische Simulation, Universität Bonn, Wegelerstrasse 6, D-53113 Bonn, Germany
E-mail address: griebel@ins.uni-bonn.de

Department of Computer Science, Columbia University, New York, NY 10027, USA; and Institute of Applied Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland
E-mail address: henryk@cs.columbia.edu