Tractability results for weighted Banach spaces of smooth functions

Markus Weimar
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
email: markus.weimar@uni-jena.de

March 11, 2011

Abstract

We study the L_∞-approximation problem for weighted Banach spaces of smooth d-variate functions, where d can be arbitrarily large. We consider the worst case error for algorithms that use finitely many pieces of information from different classes. Adaptive algorithms are also allowed. For a scale of Banach spaces we prove necessary and sufficient conditions for tractability in the case of product weights. Furthermore, we show the equivalence of weak tractability with the fact that the problem does not suffer from the curse of dimensionality.

1 Introduction

The so-called *curse of dimensionality* can often be observed for multivariate approximation problems. That is, the minimal number of information operations needed to compute an ε-approximation of a d-variate problem depends exponentially on the dimension d. The phrase *curse of dimensionality* was already coined by Bellman in 1957. Since the late 1980's there has been a considerable interest in finding optimal algorithms, also concerning the optimal dependence on d and a theory called *information-based complexity* (IBC) has been created, see, e.g., [10]. Since there are different ways to measure the lack of exponential behavior, several kinds of tractability were introduced. A brief history of the studies of multivariate problems, as well as general tractability results and many concrete examples can be found in, e.g., [5, 6, 8].

In this paper we especially consider the L_∞-approximation problem defined on some Banach spaces $\mathcal{F}_d$ of real-valued d-variate functions. In Section 2 we formulate the problem exactly and recall usual error definitions, as well as notions of tractability. Afterwards, in Section 3, we illustrate the hardness of the problem with an example studied by Novak and Woźniakowski [7] and show how weighted spaces can help to improve this negative result. Thereby, we especially concentrate on so-called *product weights*. While there exists a well-developed concept to handle problems defined on Hilbert spaces, we need an essentially new approach to conclude results in the general Banach space setting. These new ideas are presented in Section 4. Using this technique we prove a lower error bound for a very small class of functions, i.e. we consider the space $\mathcal{P}_d^\gamma$ of d-variate polynomials of degree at most one in each coordinate, equipped with some weighted norm. In Section 5 we recall a known result of Kuo, Wasilkowski and Woźniakowski [3] about upper error bounds on a certain weighted reproducing kernel Hilbert space $\mathcal{H}_d^\gamma$. Next, in Section 6, we prove the three main theorems of this paper. That is, we show necessary and sufficient conditions for several kinds of tractability for a whole scale of weighted Banach function spaces $\mathcal{F}_d^\gamma$, where $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma \hookrightarrow \mathcal{H}_d^\gamma$, in terms of the weights γ. In particular, we provide a characterization of weak tractability and the curse of dimensionality. It is shown that for these kinds of tractability results we can restrict ourselves to linear non-adaptive algorithms. We illustrate our results by applying them to selected examples and discuss a typical case of product weights. Finally, in Section 7, we add some remarks about possible extensions of the result to other domains. In addition, we briefly consider the L_p-approximation problem for $1 \leq p < \infty$ and correct a small mistake stated in [7].

2 The approximation problem

We investigate tractability properties of the approximation problem defined on some Banach spaces $\mathcal{F}_d$ of bounded functions $f\colon [0,1]^d \to \mathbb{R}$. We want to minimize the *worst case error*

$$e^{\mathrm{wor}}(A_{n,d}; \mathcal{F}_d) = \sup_{f \in \mathcal{B}(\mathcal{F}_d)} \left\| f - A_{n,d}(f) \,|\, L_\infty([0,1]^d) \right\|$$

with respect to all algorithms $A_{n,d} \in \mathcal{A}_n$ that use n pieces of information in d dimensions from a certain class Λ. Here $\mathcal{B}(\mathcal{F}_d) = \{f \in \mathcal{F}_d \,|\, \|f \,|\, \mathcal{F}_d\| \leq 1\}$ denotes the unit ball of $\mathcal{F}_d$. Hence, we study the *n-th minimal error*

$$e(n, d; \mathcal{F}_d) = \inf_{A_{n,d} \in \mathcal{A}_n} e^{\mathrm{wor}}(A_{n,d}; \mathcal{F}_d)$$

of L_∞-approximation on $\mathcal{F}_d$. An algorithm $A_{n,d} \in \mathcal{A}_n$ is modeled as a mapping $\phi\colon \mathbb{R}^n \to L_\infty([0,1]^d)$ and a function $N\colon \mathcal{F}_d \to \mathbb{R}^n$ such that $A_{n,d} = \phi \circ N$. In detail, the information map N is given by

$$N(f) = (L_1(f), L_2(f), \dots, L_n(f)), \qquad f \in \mathcal{F}_d, \tag{1}$$

where $L_j \in \Lambda$. Here we distinguish certain classes of information operations Λ. In one case we assume that we can compute arbitrary continuous linear functionals. Then $\Lambda = \Lambda^{\mathrm{all}}$ coincides with $\mathcal{F}_d^*$, the dual space of $\mathcal{F}_d$. Often only function evaluations are permitted, i.e. $L_j(f) = f\left(t^{(j)}\right)$ for a certain fixed $t^{(j)} \in [0,1]^d$. In this case $\Lambda = \Lambda^{\mathrm{std}}$ is called *standard information*. If function evaluation is continuous for all $t \in [0,1]^d$ we have $\Lambda^{\mathrm{std}} \subset \Lambda^{\mathrm{all}}$. If L_j depends continuously on f but is not necessarily linear the class is denoted by Λ^{cont}. Note that in this case also N is continuous and we obviously have $\Lambda^{\mathrm{all}} \subset \Lambda^{\mathrm{cont}}$.

Furthermore, we distinguish between *adaptive* and *non-adaptive* algorithms. The latter case is described above in formula (1), where L_j does not depend on the previously computed values $L_1(f), \dots, L_{j-1}(f)$. In contrast, we also discuss algorithms of the form $A_{n,d} = \phi \circ N$ with

$$N(f) = (L_1(f), L_2(f; y_1), \dots, L_n(f; y_1, \dots, y_{n-1})), \qquad f \in \mathcal{F}_d, \tag{2}$$

where $y_1 = L_1(f)$ and $y_j = L_j(f; y_1, \dots, y_{j-1})$ for $j = 2, 3, \dots, n$. If N is adaptive we restrict ourselves to the case where L_j depends linearly on f, i.e. $L_j(\,\cdot\,; y_1, \dots, y_{j-1}) \in \Lambda^{\mathrm{all}}$.

In all cases of information maps, the mapping ϕ can be chosen arbitrarily and is not necessarily linear or continuous. The smallest class of algorithms under consideration is the class of linear, non-adaptive algorithms of the form

$$(A_{n,d}f)(x) = \sum_{j=1}^{n} L_j(f) \cdot g_j(x), \qquad x \in [0,1]^d,$$

with some $g_j \in L_\infty$ and $L_j \in \Lambda^{\mathrm{all}}$ or even $L_j \in \Lambda^{\mathrm{std}}$. We denote the class of all such algorithms by $\mathcal{A}_n^{\mathrm{lin}}$. On the other hand, the most general classes consist of algorithms $A_{n,d} = \phi \circ N$, where ϕ is arbitrary and N either uses non-adaptive continuous or adaptive linear information. We denote the respective classes by $\mathcal{A}_n^{\mathrm{cont}}$ and $\mathcal{A}_n^{\mathrm{adapt}}$.

The minimal number of information operations needed to achieve an error smaller than a given $\varepsilon > 0$,

$$n(\varepsilon, d; \mathcal{F}_d) = \min \{ n \in \mathbb{N}_0 \mid e(n, d; \mathcal{F}_d) \le \varepsilon \},$$

is called *information complexity*.

If for a given problem, like the L_∞-approximation (with respect to a given class of algorithms) considered here, $n(\varepsilon, d; \mathcal{F}_d)$ increases exponentially in the dimension d we say the problem suffers from the *curse of dimensionality*. That is, there exist constants $c > 0$ and $C > 1$ such that for at least one $\varepsilon > 0$ we have

$$n(\varepsilon, d; \mathcal{F}_d) \geq c \cdot C^d,$$

for infinitely many $d \in \mathbb{N}$. More generally, if the information complexity depends exponentially on d or ε^{-1} we call the problem *intractable*. Otherwise we have *weak tractability*, which can be expressed by

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln\left(n(\varepsilon, d; \mathcal{F}_d)\right)}{\varepsilon^{-1}+d} = 0.$$

We want to stress the point that weak tractability implies the absence of the curse of dimensionality, but in general the converse is not true. Since there are many ways to measure the lack of exponential dependence we later distinguish between different types of tractability. The most important type is *polynomial tractability*. We say that the problem is polynomially tractable if there exist constants $c, p, q > 0$ such that

$$n(\varepsilon, d; \mathcal{F}_d) \leq c \cdot \varepsilon^{-p} \cdot d^q \quad \text{for all} \quad d \in \mathbb{N}, \varepsilon > 0.$$

If this inequality holds with $q = 0$, the problem is called *strongly polynomially tractable*. For more specific definitions and relations between these classes of tractability see, e.g., [6].

3 The concept of weighted spaces

In [7] it is shown that the approximation problem defined on $C^\infty([0,1]^d)$ is intractable. In fact, Novak and Woźniakowski considered the linear space of all real-valued infinitely differentiable functions f defined on the unit cube $[0,1]^d$ in d dimensions for which the norm

$$\|f \mid \mathcal{F}_d\| = \sup_{\alpha\in\mathbb{N}_0^d} \|D^\alpha f\|_\infty$$

of $f \in \mathcal{F}_d$ is finite. Here $\|\cdot\|_\infty$ denotes the usual sup-norm over $[0,1]^d$ and $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$, where $|\alpha| = \sum_{j=1}^d \alpha_j$ denotes the length of the multi-index $\alpha \in \mathbb{N}_0^d$.

The *initial error* of this problem is given by $e(0, d; \mathcal{F}_d) = 1$, the norm of the embedding $\mathcal{F}_d \hookrightarrow L_\infty$, since $A_{0,d} \equiv 0$ is a valid choice of an algorithm which does not use any information of f. This means that the problem is well-scaled. In detail, Theorem 1 in [7] yields that for L_∞-approximation defined on $\mathcal{F}_d$ we have

$$e(n, d; \mathcal{F}_d) = 1 \quad \text{for all} \quad n = 0, 1, \ldots, 2^{\lfloor d/2 \rfloor} - 1.$$

Therefore, for all $d \in \mathbb{N}$ and $\varepsilon \in (0,1)$,

$$n(\varepsilon, d; \mathcal{F}_d) \geq 2^{\lfloor d/2 \rfloor}.$$

Hence, the problem suffers from the curse of dimensionality; in particular it is intractable.

One possibility to avoid this exponential dependence on d, i.e. to break the curse, is to shrink the function space $\mathcal{F}_d$. A closer look at the norm yields that for $f \in \mathcal{B}(\mathcal{F}_d)$ we have

$$\|D^\alpha f\|_\infty \leq 1 \quad \text{for all } \alpha \in \mathbb{N}_0^d. \tag{3}$$

Hence, every derivative is equally important. In order to shrink the space, for each $\alpha \in \mathbb{N}_0^d$ we replace the right-hand side of inequality (3) by a weight $0 \leq \gamma_\alpha \leq 1$. For α with $|\alpha| = 1$ this means that we control the importance of every single variable. So, the norm in the weighted space is now given by

$$\|f \mid \mathcal{F}_d^\gamma\| = \sup_{\alpha \in \mathbb{N}_0} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_\infty ,$$

where we demand $D^\alpha f$ to be equal to zero if $\gamma_\alpha = 0$.

The idea to introduce weights directly into the norm of the function space appeared for the first time in a paper of Sloan and Woźniakowski in 1998, see [9]. They studied the integration problem defined over some Sobolev Hilbert space, equipped with so-called *product weights*, to explain the overwhelming success of QMC integration rules. Thenceforth, weighted problems attracted a lot of attention. For example it turned out that tractability of approximation of linear operators between Hilbert spaces can be fully characterized in terms of the weights and singular values of the linear operators if we use information operations from the class Λ^{all}.

Let us have a closer look at product weights. Assume that for every $d \in \mathbb{N}$ there exists an ordered and bounded sequence

$$1 \geq \gamma_{d,1} \geq \gamma_{d,2} \geq \ldots \geq \gamma_{d,d} \geq 0.$$

Then for $d \in \mathbb{N}$, the product weight sequence $\gamma = (\gamma_\alpha)_{\alpha \in \mathbb{N}_0^d}$ is given by

$$\gamma_\alpha = \prod_{j=1}^{d} (\gamma_{d,j})^{\alpha_j} , \quad \alpha \in \mathbb{N}_0^d. \tag{4}$$

Note that the dependence of x_j on f is now controlled by the so-called *generator weight* $\gamma_{d,j}$. Since $\gamma_{d,j} = 0$ for some $j \in \{1, \ldots, d\}$ implies that f does not depend on $x_j, \ldots, x_d$ we assume that $\gamma_{d,d} > 0$ in the rest of the paper. Moreover, the ordering of $\gamma_{d,j}$ is without loss of generality. Later on we will see that tractability of our problem will only depend on summability properties of the generator weights.

Among other things, it turns out that for the L_∞-approximation problem defined on the Banach space with the norm given above and generator weights $\gamma_{d,j} \equiv \gamma_j = \Theta\left(j^{-\beta}\right)$ we have

- intractability for $\beta = 0$,
- weak tractability but no polynomial tractability for $0 < \beta < 1$,
- strong polynomial tractability if $1 < \beta$.

Moreover, we prove that for $\beta = 1$ the problem is not strongly polynomially tractable.

4 Lower bounds

First, we want to describe the main ideas used in the Hilbert space setting. Hence, for a moment, consider the problem of L_2-approximation with respect to linear algorithms defined on a reproducing kernel Hilbert space $\mathcal{H}(K_d)$ of functions $f\colon [0,1]^d \to \mathbb{R}$. Let

$$W_d\colon \mathcal{H}(K_d) \to \mathcal{H}(K_d), \qquad W_d(g) = \int_{[0,1]^d} g(x)\, K_d(\cdot, x)\, dx.$$

We assume that W_d is compact. Then the worst case error is fully characterized by the spectrum of W_d that is also a self-adjoint, and non-negative definite operator. Let $\{(\lambda_{d,j}, \eta_{d,j}) \mid j \in \mathbb{N}\}$ denote a complete orthonormal system of eigenpairs of W_d, indexed according to the non-increasing order of the eigenvalues, i.e.

$$W_d(\eta_{d,j}) = \lambda_{d,j}\, \eta_{d,j} \quad \text{and} \quad \langle \eta_{d,i}, \eta_{d,j} \rangle_{\mathcal{H}(K_d)} = \delta_{ij} \quad \text{with} \quad \lambda_{d,j} \geq \lambda_{d,j+1} \geq 0.$$

For $\lambda_{d,n} > 0$, it is well known that the algorithm

$$A^*_{n,d}(f) = \sum_{j=1}^{n} \langle f, \tilde{\eta}_{d,j} \rangle_{L_2} \cdot \tilde{\eta}_{d,j}, \quad \text{where} \quad \tilde{\eta}_{d,j} = \frac{\eta_{d,j}}{\sqrt{\lambda_{d,j}}}$$

is optimal. Then the n-th minimal error is given by

$$e(n, d; \mathcal{H}(K_d)) = e^{\mathrm{wor}}(A^*_{n,d}; \mathcal{H}(K_d)) = \sqrt{\lambda_{d,n+1}}.$$

For more details see, e.g., [4] and [6], as well as the references in there. For a comprehensive introduction to reproducing kernel Hilbert spaces see, for instance, Chapter 1 in the book of Wahba [11].

In the general Banach space setting this approach obviously doesn't work. Our technique is based on the ideas of Werschulz and Woźniakowski [12], as well as Novak and Woźniakowski [7]. Among other things it uses a result from Banach space theory and nonlinear functional analysis, namely, the theorem of Borsuk-Ulam. The proof of the following proposition can be found in Chapter 1.4.2, [1].

Proposition 1 (Borsuk-Ulam)**.** Let V be a linear normed space over $\mathbb{R}$ with $\dim V = m$ and, moreover, let $N\colon V \to \mathbb{R}^n$ be a continuous mapping for $n < m$. Then there exists an element $f^* \in V$ with $\|f^* \,|\, V\| = 1$ such that $N(f^*) = N(-f^*)$.

The main tool to conclude lower bounds in the Banach space setting now reads as follows.

Lemma 1. Assume that F and G are linear normed spaces such that $F \subseteq G$. Furthermore, suppose that $V \subseteq F$ is a linear subspace of dimension m and there exists a constant $a > 0$ such that

$$\|f \,|\, F\| \leq \frac{1}{a} \|f \,|\, G\| \quad \text{for all} \quad f \in V. \tag{5}$$

Then for every $n < m$ and every $A_n \in \mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$

$$e^{\mathrm{wor}}(A_n; F) = \sup_{f \in \mathcal{B}(F)} \|f - A_n(f) \,|\, G\| \geq a.$$

Proof. For $A_n \in \mathcal{A}_n^{\mathrm{cont}}$ the assertion is a simple conclusion of Proposition 1 and can be found in [7]. On the other hand, if $A_n \in \mathcal{A}_n^{\mathrm{adapt}}$ the proof can be obtained by arguments from linear algebra, which are indicated in the proof of Theorem 3.1 in [12]. In any case we exclusively use norm properties from the space G, no additional structure of G is used. Therefore, this tool is available for any kind of approximation problem, not only for L_∞-approximation. $\square$

In the following we use Lemma 1 to conclude a lower bound for the approximation error for the space

$$\mathcal{P}_d^\gamma = \operatorname{span}\left\{ p_i\colon [0,1]^d \to \mathbb{R},\, p_i(x) = \prod_{j=1}^d (x_j)^{i_j} \;\;\middle|\;\; i = (i_1, \ldots, i_d) \in \{0,1\}^d \right\}$$

of all real-valued d-variate polynomials of degree at most one in each coordinate direction, defined on the unit cube $[0,1]^d$. We equip this linear space with the weighted norm

$$\|f \,|\, \mathcal{P}_d^\gamma\| = \max_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_\infty\,, \qquad f \in \mathcal{P}_d^\gamma,$$

where γ is the product weight sequence described as in Section 3.

Theorem 1. Let $e(n,d;\mathcal{P}_d^\gamma)$ be the n-th minimal error of L_∞-approximation on $\mathcal{P}_d^\gamma$ with respect to the class $\mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$ of all algorithms described in Section 2. Then

$$e(n,d;\mathcal{P}_d^\gamma) \geq 1 \quad \text{for all} \quad n < 2^s,$$

and some integer $s \in [0,d]$ with

$$s > \frac{1}{3} \cdot \left(\sum_{j=1}^{d} \gamma_{d,j} - 2 \right). \tag{6}$$

Proof. The proof of the lower error bound consists of several steps. At first, we construct a partition of the set $\{1,\ldots,d\}$ into $s+1$ parts which we will need later and with s satisfying (6). In a second step, we define a special linear subspace $V \subseteq \mathcal{P}_d^\gamma$ with $\dim V = 2^s$. Step 3 then shows that V satisfies the assumptions of Lemma 1. The proof is completed in Step 4.

Step 1. For $k \in \{0,\ldots,d\}$, we define inductively $m_0 = 0$ and

$$m_k = \inf \left\{ t \in \mathbb{N} \,|\, m_{k-1} < t \leq d, \text{ with } 2 \leq \sum_{j=m_{k-1}+1}^{t} \gamma_{d,j} \right\}$$

with the usual convention $\inf \emptyset = \infty$. Note that the infimum coincides with the minimum in the finite case, since then $m_k \in \mathbb{N}$. Moreover, we set

$$s = \max \{ k \in \{0,\ldots,d\} \,|\, m_k < \infty \}.$$

We denote $I_k = \{m_{k-1}+1, m_{k-1}+2, \ldots, m_k\}$ for $k = 1,\ldots,s$. Thus, this gives a uniquely defined disjoint partition of the set

$$\{1,\ldots,d\} = \left(\bigcup_{k=1}^{s} I_k \right) \cup \{m_s+1,\ldots,d\},$$

and m_k denotes the last element of the block I_k. For all $k = 1,\ldots,s$, we conclude

$$2 \leq \sum_{j\in I_k} \gamma_{d,j} < 2 + \gamma_{d,m_k} < 3.$$

Finally, summation of these inequalities gives

$$\sum_{j=1}^{d} \gamma_{d,j} < \sum_{k=1}^{s} \sum_{j\in I_k} \gamma_{d,j} + 2 < 3s + 2,$$

and (6) follows immediately.

If $s = 0$ we can stop at this point since the initial error is 1 as the norm of the embedding $\mathcal{P}_d^\gamma \hookrightarrow L_\infty$ and the remaining assertion is trivial. Hence, from now on we can assume that $s > 0$ and $m_s \geq 1$.

Step 2. To apply Lemma 1 we have to construct a linear subspace V of $F = \mathcal{P}_d^\gamma$ such that the condition (5) holds for $G = L_\infty([0,1]^d)$ and $a = 1$. First, we restrict ourselves to the set

$$\widetilde{F} = \{ f \in F \mid \ f \text{ depends only on } x_1, \ldots, x_{m_s} \} .$$

By a simple isometric isomorphism we can interpret $\widetilde{F}$ as the space $\mathcal{P}_{m_s}^\gamma$.

We are ready to construct a suitable space V using the partition from Step 1. We define V as the span of all functions $g_i \colon X = [0,1]^{m_s} \to \mathbb{R}$, $i = (i_1, \ldots, i_s) \in \{0,1\}^s$, of the form

$$g_i(x) = \prod_{k=1}^{s} \left(\sum_{j \in I_k} \gamma_{d,j} \cdot x_j \right)^{i_k}, \quad x \in X.$$

Obviously, V is a linear subspace of $\mathcal{P}_{m_s}^\gamma$ and with the interpretation above also a linear subspace of F. Moreover, it is easy to see that we have by construction

$$\left\| g \,|\, \mathcal{P}_{m_s}^\gamma \right\| = \| g \,|\, F \| \quad \text{and} \quad \| g \,|\, L_\infty(X) \| = \left\| g \,|\, L_\infty([0,1]^d) \right\| \quad \text{for} \ \ g \in V.$$

Finally, we note that $\dim V = \#\{0,1\}^s = 2^s$. It remains to show that this subspace is the right choice to prove the claim using Lemma 1.

Step 3. The proof of the needed condition (5),

$$\left\| g \,|\, \mathcal{P}_{m_s}^\gamma \right\| \leq \| g \,|\, L_\infty(X) \| \quad \text{for all} \quad g \in V,$$

is a little bit technical. Due to the special structure of the functions $g \in V$, the left hand side reduces to $\max \{ \gamma_\alpha^{-1} \, \| D^\alpha g \,|\, L_\infty(X) \| \mid \alpha \in \mathcal{M} \}$, where the maximum is taken over the set

$$\mathcal{M} = \left\{ \alpha \in \{0,1\}^{m_s} \mid \sum_{j \in I_k} \alpha_j \leq 1 \text{ for all } k = 1, \ldots, s \right\}.$$

This is simply because for $\alpha \notin \mathcal{M}$ we have $D^\alpha g \equiv 0$ and the inequality is trivial. To simplify the notation let us define

$$T \colon \{0,1\}^{m_s} \to \mathbb{N}_0^s, \quad \alpha \mapsto T(\alpha) = \sigma = (\sigma_1, \ldots, \sigma_s), \quad \text{where} \quad \sigma_k = \sum_{j \in I_k} \alpha_j \quad \text{for} \quad k = 1, \ldots, s.$$

Note that $T(\mathcal{M}) = \{0,1\}^s$. Moreover, for every $g = \sum_{i\in\{0,1\}^s} a_i g_i(\cdot) \in V$ define the function

$$h_g\colon Z = \bigtimes_{k=1}^{s} \left[0, \sum_{j\in I_k} \gamma_{d,j}\right] \to \mathbb{R}, \qquad h_g(z) = \sum_{i\in\{0,1\}^s} a_i \prod_{k=1}^{s} z_k^{i_k}.$$

Hence, $h_g(z) = g(x)$ under the transformation $x \mapsto z$ such that

$$z_k = \sum_{j\in I_k} \gamma_{d,j} x_j \quad \text{for every} \quad k = 1, \ldots, s \;\; \text{and every} \;\; x \in X.$$

The span W of all functions $h\colon Z \to \mathbb{R}$ with this structure also is a linear space. Furthermore, easy calculus yields

$$\left(D_x^\alpha g\right)(x) = \left(\prod_{j=1}^{m_s} \left(\gamma_{d,j}\right)^{\alpha_j}\right) \left(D_z^{T(\alpha)} h_g\right)(z) \quad \text{for all} \quad g \in V,\ \alpha \in \mathcal{M} \quad \text{and} \quad x \in X. \tag{7}$$

Here the x and z in D_x^α and $D_z^{T(\alpha)}$ indicate differentiation with respect to x and z, respectively. Since the mapping $x \mapsto z$ is surjective we obtain $\|D^\alpha g \,|\, L_\infty(X)\| = \gamma_\alpha \left\|D^{T(\alpha)} h_g \,|\, L_\infty(Z)\right\|$ by the form of γ given by (4). Hence,

$$\max_{\alpha\in\mathcal{M}} \frac{1}{\gamma_\alpha} \|D^\alpha g \,|\, L_\infty(X)\| = \max_{\sigma\in\{0,1\}^s} \|D^\sigma h_g \,|\, L_\infty(Z)\|.$$

Note that (7) with $\alpha = 0$ yields $\|g \,|\, L_\infty(X)\| = \|h_g \,|\, L_\infty(Z)\|$. Therefore, the claim reduces to

$$\max_{\sigma\in\{0,1\}^s} \|D^\sigma h_g \,|\, L_\infty(Z)\| \le \|h_g \,|\, L_\infty(Z)\| \quad \text{for every} \quad g \in V.$$

We show this estimate for every $h \in W$, i.e.,

$$\|D^\sigma h \,|\, L_\infty(Z)\| \le \|h \,|\, L_\infty(Z)\| \quad \text{for all} \quad \sigma \in \{0,1\}^s. \tag{8}$$

We start with the special case of one derivative, i.e. $\sigma = e_k$ for a certain $k \in \{1, \ldots, s\}$. Since h is affine in each coordinate we can represent it as $h(z) = a(z^k) \cdot z_k + b(z^k)$ with functions a and b which only depend on $z^k = (z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_s)$. Thus, we have $D^{e_k} h(z) = a(z^k)$ and need to show that

$$\left|a(z^k)\right| \le \max\left\{ \left|b(z^k)\right|, \left|a(z^k) \cdot \sum_{j\in I_k} \gamma_{d,j} + b(z^k)\right| \right\}. \tag{9}$$

This is obviously true for every $z \in Z$ with $a(z^k) = 0$. For $a(z^k) \neq 0$ we can divide by $\left|a(z^k)\right|$ to get

$$1 \leq \max\left\{|t|, \left|\sum_{j \in I_k} \gamma_{d,j} - t\right|\right\}$$

if we set $t = -b(z^k)/a(z^k)$. The last maximum is minimal if both of its entries coincide. This is for $t = \frac{1}{2}\sum_{j\in I_k} \gamma_{d,j}$. Hence, we need to demand

$$2 \leq \sum_{j \in I_k} \gamma_{d,j}$$

to conclude (9) for all admissible $z \in Z$. But this is true for every $k \in \{1, \ldots, s\}$ by definition of the sets I_k in Step 1. This proves (8) for the special case $\sigma = e_k$ for all $k \in \{1, \ldots, s\}$.

The inequality (8) also holds for every $\sigma \in \{0,1\}^s$ by an easy inductive argument on the cardinality of $|\sigma|$. Indeed, if $|\sigma| \geq 2$ then $\sigma = \sigma' + e_k$ with $|\sigma'| = |\sigma| - 1$. We now need to estimate $\left\|D^{\sigma'+e_k} h \,|\, L_\infty(Z)\right\|$. Since $D^{e_k} h(z) = a(z^k)$ has the same structure as the function h itself, we have $\left\|D^{\sigma'+e_k} h \,|\, L_\infty(Z)\right\| = \left\|D^{\sigma'} a(z^k) \,|\, L_\infty(Z)\right\|$ and the proof is completed by the inductive step.

Step 4. For every $g \in V$ we have

$$\begin{aligned} \left\|g \,|\, \mathcal{P}_d^\gamma\right\| &= \left\|g \,|\, \mathcal{P}_{m_s}^\gamma\right\| = \max_{\substack{\alpha \in \{0,1\}^{m_s} \\ T(\alpha) \in \{0,1\}^s}} \frac{1}{\gamma_\alpha} \left\|D^\alpha g \,|\, L_\infty(X)\right\| = \max_{\sigma \in \{0,1\}^s} \left\|D^\sigma h_g \,|\, L_\infty(Z)\right\| \\ &\leq \left\|h_g \,|\, L_\infty(Z)\right\| = \left\|g \,|\, L_\infty(X)\right\| = \left\|g \,|\, L_\infty([0,1]^d)\right\|, \end{aligned}$$

where V is a linear subspace of $F = \mathcal{P}_d^\gamma$ with $\dim V = 2^s$. Therefore, Lemma 1 with $a = 1$ yields that the worst case error of any algorithm $A_{n,d}$ we consider, with $n < \dim V$ pieces of information, is bounded from below by one. That is, $e^{\mathrm{wor}}(A_{n,d}; \mathcal{P}_d^\gamma) \geq 1$. We complete the proof by taking the infimum with respect to $A_{n,d} \in \mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$. $\square$

5 Upper bounds

The approximation problem has been studied in many different settings. We restrict ourselves to the case of L_∞-approximation defined on a special weighted anchored Sobolev Hilbert space $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$.

For $d = 1$ and $\gamma > 0$, this is the space of all absolutely continuous functions $f\colon [0,1] \to \mathbb{R}$ whose first derivatives belong to $L_2([0,1])$. The inner product in the space $\mathcal{H}_1^\gamma$ is defined as

$$\langle f, g\rangle_{\mathcal{H}_1^\gamma} = f(0)g(0) + \gamma^{-1} \int_0^1 f'(x) g'(x)\, dx, \qquad f, g \in \mathcal{H}_1^\gamma,$$

where the derivatives have to be understood in the weak sense. For $\gamma = 0$ the space consists of only constant functions.

It turns out that $\mathcal{H}_1^\gamma$ is a reproducing kernel Hilbert space $\mathcal{H}(K_1^\gamma)$ whose kernel is

$$K_1^\gamma(x,y) = 1 + \gamma \min\{x,y\} \qquad \text{for} \quad x,y \in [0,1].$$

For $d > 1$, the space $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$ is defined as the d-fold tensor product of $\mathcal{H}(K_1^{\gamma_{d,j}})$, where we once again assume product weights, see (4), with

$$1 \geq \gamma_{d,1} \geq \gamma_{d,2} \geq \ldots \geq \gamma_{d,d} \geq 0.$$

Due to the product structure of γ_α, the corresponding reproducing kernel of $\mathcal{H}_d^\gamma$ is a weighted Wiener sheet kernel,

$$K_d^\gamma(x,y) = \prod_{j=1}^{d} \left(1 + \gamma_{d,j} \min\{x_j, y_j\}\right), \qquad x,y \in [0,1]^d.$$

The associated inner product is given by

$$\langle f, g\rangle_{\mathcal{H}_d^\gamma} = \sum_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \int_{[0,1]^{|\alpha|}} \frac{\partial^{|\alpha|} f}{\partial x_\alpha}(x_\alpha, 0) \cdot \frac{\partial^{|\alpha|} g}{\partial x_\alpha}(x_\alpha, 0)\, dx_\alpha, \qquad f,g \in \mathcal{H}_d^\gamma.$$

Here the term (x_α, a) means the d-dimensional vector with $(x_\alpha, a)_j = x_j$ for all coordinates j with $\alpha_j = 1$ and $(x_\alpha, a)_j = a_j$ otherwise. For $\alpha = 0$ we replace the integral by $f(a)g(a)$. Therefore, the point $a = 0 \in [0,1]^d$ is sometimes called an *anchor* of the space.

A closer look at the respective norm justifies to refer to $\mathcal{H}(K_d^\gamma)$ as a Sobolev space of dominating mixed smoothness. For $\gamma_{d,d} > 0$, the space $\mathcal{H}(K_d^\gamma)$ algebraically coincides with the space

$$\left\{ f\colon [0,1]^d \to \mathbb{R} \mid D^\alpha f \in L_2([0,1]^d) \text{ for all } \alpha = (\alpha_1, \ldots, \alpha_d) \text{ with } \max_{j=1,\ldots,d} \alpha_j \leq 1 \right\},$$

where $D^\alpha f$ once again denotes the weak derivative in the Sobolev sense. Equipped with the usual norm, this space is often denoted by $W_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d)$, or $S_2^1 W([0,1]^d)$, respectively. If $\gamma_{d,j} = 0$ for some $j \in \{1, \ldots, d\}$ we obtain a proper subspace of functions that are constant with respect to $x_j, \ldots, x_d$. Therefore, we always assume $\gamma_{d,d} > 0$.

Kuo, Wasilkowski and Woźniakowski [3, 8. Example] showed

Proposition 2. There exists a linear algorithm $A^*_{n,d}$ for L_∞-approximation on $\mathcal{H}^\gamma_d$ such that it uses n non-adaptively chosen linear functionals and for every $\tau \in (1/2, 1)$ there are constants $a_\tau, b_\tau > 0$ independent of γ and d such that

$$e^{\mathrm{wor}}(A^*_{n,d}; \mathcal{H}^\gamma_d) \leq b_\tau \cdot n^{-(1-\tau)/(2\tau)} \cdot \prod_{j=1}^{d} \left(1 + a_\tau \gamma^\tau_{d,j}\right)^{1/(2\tau)}.$$

Furthermore, $A^*_{n,d}$ is close to be optimal in the class $\mathcal{A}^{\mathrm{lin}}_n$.

6 Conclusions and applications

We now combine lower and upper bounds presented before and prove general results for L_∞-approximation on weighted Banach function spaces. More precisely, consider a sequence of Banach spaces $\mathcal{F}^\gamma_d$ of functions $f\colon [0,1]^d \to \mathbb{R}$ which fulfills the following simple assumptions:

(A1) $\mathcal{P}^\gamma_d \hookrightarrow \mathcal{F}^\gamma_d$ with an embedding factor $C_{1,d} \leq 1$ for all d,

(A2) $\mathcal{F}^\gamma_d \hookrightarrow \mathcal{H}^\gamma_d$ with an embedding factor $C_{2,d}$ for all d and

$$C_{2,d} \leq a \cdot \exp\left(b \cdot \sum_{j=1}^{d} (\gamma_{d,j})^t\right)$$

for some constants $a, b \geq 0$ and a parameter $t \in (0, 1]$, independent of d and γ.

By $A \hookrightarrow B$ with an embedding factor C, we mean that the normed linear space A is continuously embedded in the normed linear space B and

$$\|f \mid B\| \leq C\, \|f \mid A\| \quad \text{for all} \quad f \in A.$$

That is, we can take $C = \|\mathrm{id} \mid \mathcal{L}(A, B)\|$ as the (operator-) norm of the identity $\mathrm{id}\colon A \to B$. Moreover, γ is once again a product weight sequence given by formula (4). The spaces $\mathcal{P}^\gamma_d$ and $\mathcal{H}^\gamma_d$ are defined in Section 4 and Section 5, respectively.

To simplify the notation for necessary and sufficient conditions of tractability, we use the commonly known definitions of the so-called *sum exponents* for the weight sequence γ,

$$p(\gamma) = \inf\left\{\kappa \geq 0 \mid \quad P_\kappa(\gamma) = \limsup_{d\to\infty} \sum_{j=1}^{d} (\gamma_{d,j})^\kappa < \infty\right\}$$

and

$$q(\gamma) = \inf\left\{\kappa \geq 0 \,|\quad Q_\kappa(\gamma) = \limsup_{d\to\infty} \frac{\sum_{j=1}^d (\gamma_{d,j})^\kappa}{\ln(d+1)} < \infty\right\},$$

with the convention that $\inf \emptyset = \infty$.

Theorem 2 (**Necessary conditions**)**.** Assume that (A1) holds. Consider L_∞-approximation over $\mathcal{F}_d^\gamma$ with respect to the class of algorithms $\mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$. Then

$$n(\varepsilon, d; \mathcal{F}_d^\gamma) > \exp\left(\frac{1}{3} \cdot \ln 2 \cdot \big(\sum_{j=1}^d \gamma_{d,j} - 2\big)\right) \quad \text{for all} \quad d \in \mathbb{N} \quad \text{and} \quad \varepsilon \in (0,1). \tag{10}$$

Therefore, if the problem is

- *polynomially tractable* then $q(\gamma) \leq 1$,
- *strongly polynomially tractable* then $p(\gamma) \leq 1$.

Proof. Due to (A1), every algorithm $A_{n,d} \in \mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$ for L_∞-approximation defined on $\mathcal{F}_d^\gamma$ also applies to the embedded space $\mathcal{P}_d^\gamma$. Furthermore, $C_{1,d} \leq 1$ implies that the unit ball $\mathcal{B}(\mathcal{P}_d^\gamma)$ is contained in the unit ball $\mathcal{B}(\mathcal{F}_d^\gamma)$. Therefore,

$$e^{\mathrm{wor}}(A_{n,d}; \mathcal{F}_d^\gamma) \geq e^{\mathrm{wor}}(A_{n,d}\big|_{\mathcal{P}_d^\gamma}; \mathcal{P}_d^\gamma) \geq e(n, d; \mathcal{P}_d^\gamma).$$

From Theorem 1 we have

$$e(n, d; \mathcal{P}_d^\gamma) \geq 1 \quad \text{for} \quad n < 2^s,$$

where $s = s(\gamma, d) \in [0, d]$ satisfies (6). Hence, for $d \in \mathbb{N}$ and $\varepsilon \in (0,1)$ we conclude

$$n(\varepsilon, d; \mathcal{F}_d^\gamma) \geq 2^s > \frac{1}{4^{1/3}}\, 2^{1/3 \sum_{j=1}^d \gamma_{d,j}},$$

as claimed in (10).

Suppose now that the problem is polynomially tractable. Then there are non-negative constants C, p and q such that

$$n(\varepsilon, d; \mathcal{F}_d^\gamma) \leq C \varepsilon^{-p} d^q \quad \text{for all} \quad d \in \mathbb{N},\ \varepsilon > 0.$$

Take now an arbitrarily fixed ε in $(0,1)$. Then (10) implies that there is a positive $\tilde{C}$ such that

$$2^{1/3\cdot\sum_{j=1}^{d}\gamma_{d,j}} \leq \tilde{C}\cdot d^{q} \quad \text{for all} \quad d\in\mathbb{N}.$$

This is equivalent to the boundedness of $\sum_{j=1}^{d}\gamma_{d,j}/\ln(d+1)$, and therefore $q(\gamma)\leq 1$, as claimed.

Suppose that the problem is strongly polynomially tractable. Then $q=0$ in the bound above, and $\sum_{j=1}^{d}\gamma_{d,j}$ is uniformly bounded in d. Hence, $p(\gamma)\leq 1$, as claimed. □

Of course, the conditions $q(\gamma)\leq 1$ and $p(\gamma)\leq 1$ are also necessary for polynomial and strong polynomial tractability with respect to smaller classes of algorithms.

We next assume (A2) and show that slightly stronger conditions on the weights γ than in Theorem 2 are sufficient for polynomial and strong polynomial tractability.

Theorem 3 (**Sufficient conditions**)**.** Assume that (A2) holds with a parameter $t\in(0,1]$. Consider L_∞-approximation over $\mathcal{F}_d^\gamma$ with respect to the class of linear algorithms $\mathcal{A}_n^{\mathrm{lin}}$. Then

- $q(\gamma)<t$ implies that the problem is *polynomially tractable*,
- $p(\gamma)<t$ implies that the problem is *strongly polynomially tractable.*

Proof. Due to (A2), the restriction of the algorithm $A_{n,d}^*$ in Proposition 2 from $\mathcal{H}_d^\gamma$ to $\mathcal{F}_d^\gamma$ is a valid linear algorithm for L_∞-approximation over $\mathcal{F}_d^\gamma$. Furthermore, due to linearity of $A_{n,d}^*$ for all $f\in\mathcal{H}_d^\gamma$, we have

$$\left\|f-A_{n,d}^*f\,|\,L_\infty([0,1]^d)\right\|\leq e^{\mathrm{wor}}(A_{n,d}^*;\mathcal{H}_d^\gamma)\cdot\|f\,|\,\mathcal{H}_d^\gamma\|\leq e^{\mathrm{wor}}(A_{n,d}^*;\mathcal{H}_d^\gamma)\cdot C_{2,d}\cdot\|f\,|\,\mathcal{F}_d^\gamma\|\,.$$

Therefore, we can estimate the n-th minimal error by

$$\begin{aligned} e(n,d;\mathcal{F}_d^\gamma) &\leq e^{\mathrm{wor}}(A_{n,d}^*\big|_{\mathcal{F}_d^\gamma};\mathcal{F}_d^\gamma)\leq C_{2,d}\cdot e^{\mathrm{wor}}(A_{n,d}^*;\mathcal{H}_d^\gamma)\\ &\leq a\cdot\exp\left(b\cdot\sum_{j=1}^{d}(\gamma_{d,j})^t\right)\cdot b_\tau\cdot n^{-(1-\tau)/(2\tau)}\cdot\prod_{j=1}^{d}\left(1+a_\tau\gamma_{d,j}^\tau\right)^{1/(2\tau)}, \end{aligned}$$

where τ is an arbitrary number from $(1/2,1)$. Using $1+x\leq e^x$ for $x\geq 0$, we have

$$e(n,d;\mathcal{F}_d^\gamma)\leq a\cdot b_\tau\cdot n^{-(1-\tau)/(2\tau)}\cdot\exp\left(b\sum_{j=1}^{d}(\gamma_{d,j})^t+\frac{a_\tau}{2\tau}\sum_{j=1}^{d}(\gamma_{d,j})^\tau\right).$$

Choosing n such that the right-hand side is at most ε, we obtain an estimate for the information complexity with respect to the class of linear algorithms,

$$n(\varepsilon, d; \mathcal{F}_d^\gamma) \leq c_1 \cdot \varepsilon^{-2\tau/(1-\tau)} \cdot \exp\left(c_2 \sum_{j=1}^{d} (\gamma_{d,j})^t + c_3 \sum_{j=1}^{d} (\gamma_{d,j})^\tau \right), \tag{11}$$

where the positive constants c_1, c_2 and c_3 only depend on τ, a and b.

Suppose that $q(\gamma) < t$. Then $Q_\kappa(\gamma)$ is finite for every $\kappa > q(\gamma)$. Taking $\kappa = t$ we obtain

$$\frac{\sum_{j=1}^{d} (\gamma_{d,j})^t}{\ln(d+1)} \cdot \ln(d+1) \leq (Q_t(\gamma) + \delta) \cdot \ln(d+1) = \ln(d+1)^{Q_t(\gamma)+\delta}$$

for every $\delta > 0$ whenever d is larger than a certain d_δ. This means that the factor $\exp\left(c_2 \sum_{j=1}^{d} (\gamma_{d,j})^t\right)$ in (11) is polynomially dependent on d.

On the other hand, we can choose $\tau \in (\max\{q(\gamma), 1/2\}, 1)$ such that $Q_\tau(\gamma)$ is finite and the factor $\exp\left(c_3 \sum_{j=1}^{d} (\gamma_{d,j})^\tau\right)$ in (11) is also polynomially dependent on d. So, for this value of τ we can rewrite (11) as

$$n(\varepsilon, d; \mathcal{F}_d^\gamma) = \mathcal{O}\left(\varepsilon^{-2\tau/(1-\tau)} \cdot (d+1)^{c_4}\right),$$

with c_4 independent of d and ε. This means that the problem is polynomially tractable, as claimed.

Suppose finally that $p(\gamma) < t$. Then the sums $\sum_{j=1}^{d} (\gamma_{d,j})^t$ and $\sum_{j=1}^{d} (\gamma_{d,j})^\tau$ for $\tau \in (\max\{p(\gamma), 1/2\}, 1)$ are both uniformly bounded in d. Therefore (11) yields strong polynomial tractability, and completes the proof. $\square$

The conditions in Theorem 3 are obviously also sufficient if we consider larger classes of algorithms. Moreover, the proof of Theorem 3 also provides explicit upper bounds for the exponents of tractability.

We now discuss the role of assumptions (A1) and (A2). They are quite different. The assumption (A1) is used to find a lower bound on the information complexity for the space $\mathcal{F}_d^\gamma$ as long the space $\mathcal{P}_d^\gamma$ is continuously embedded in $\mathcal{F}_d^\gamma$ with an embedding factor at most one. Such an embedding can be shown for several different classes of functions.

The assumption (A2) is used to find an upper bound on the information complexity for the space $\mathcal{F}_d^\gamma$ as long as it is continuously embedded in the space $\mathcal{H}_d^\gamma$ with an embedding factor depending exponentially on the sum of some power of the product weights. This considerably restricts the choice of $\mathcal{F}_d^\gamma$. We need this assumption in order to use the linear algorithm $A_{n,d}^*$ defined on the space $\mathcal{H}_d^\gamma$ due to Kuo et al. [3] and the error bound they proved. Obviously, we can replace the space $\mathcal{H}_d^\gamma$ in (A2) by some other space which contains at least $\mathcal{P}_d^\gamma$ and for which we know a linear algorithm using n linear functionals whose worst case error is polynomial in n^{-1} with an explicit dependence on the product weights.

We now show that the assumptions (A1) and (A2) allow us to characterize weak tractability and the curse of dimensionality.

Theorem 4 (**Weak tractability and the curse of dimensionality**)**.** Suppose that (A1) and (A2) with a parameter $t \in (0,1]$ hold. Then for L_∞-approximation defined on the space $\mathcal{F}_d^\gamma$ the following statements are equivalent:

(i) The problem is weakly tractable with respect to the class $\mathcal{A}_n^{\mathrm{lin}}$.

(ii) The problem is weakly tractable with respect to the class $\mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$.

(iii) There is no curse of dimensionality for the class $\mathcal{A}_n^{\mathrm{lin}}$.

(iv) There is no curse of dimensionality for the class $\mathcal{A}_n^{\mathrm{cont}} \cup \mathcal{A}_n^{\mathrm{adapt}}$.

(v) For all $\kappa > 0$ we have $\lim\limits_{d\to\infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.

(vi) There exists $\kappa \in (0,t)$ such that $\lim\limits_{d\to\infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.

Proof. We start by showing that (vi) implies (i), i.e.,

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln\left(n(\varepsilon,d;\mathcal{F}_d^\gamma)\right)}{\varepsilon^{-1}+d} = 0,$$

where the information complexity is taken with respect to linear algorithms $\mathcal{A}_n^{\mathrm{lin}}$. By the arguments used in the proof of Theorem 3 we obtain estimate (11) for all $\varepsilon > 0$, as well as for every $d \in \mathbb{N}$ and all $\tau \in (1/2,1)$, due to assumption (A2). Clearly, for $\kappa \in (0,t)$ as in the hypothesis and $t \in (0,1]$ as in the embedding condition, we can find $\tau \in (1/2,1)$ such that $\kappa < \min\{t,\tau\}$. So, since $\gamma_{d,j} \leq 1$, we can estimate both sums on the right-hand side of (11) from above by $\sum_{j=1}^d (\gamma_{d,j})^{\min\{t,\tau\}} \leq \sum_{j=1}^d (\gamma_{d,j})^\kappa$. Thus,

$$\frac{\ln\left(n(\varepsilon,d;\mathcal{F}_d^\gamma)\right)}{\varepsilon^{-1}+d} \leq \frac{\ln(c_1)}{\varepsilon^{-1}+d} + \frac{2\tau}{1-\tau} \cdot \frac{\ln\left(\varepsilon^{-1}\right)}{\varepsilon^{-1}+d} + \max\{c_2,c_3\} \cdot \frac{\sum_{j=1}^d (\gamma_{d,j})^\kappa}{\varepsilon^{-1}+d}$$

tends to zero when $\varepsilon^{-1}+d$ approaches infinity, as claimed.

Clearly, (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) and (v) $\Rightarrow$ (vi). Hence, we only need to show that (iv) $\Rightarrow$ (v). From (A1) we have estimate (10). Then no curse of dimensionality implies

$$\lim_{d\to\infty} \frac{1}{d} \sum_{j=1}^{d} \gamma_{d,j} = 0.$$

Now, Jensen's inequality yields

$$\frac{1}{d} \sum_{j=1}^{d} \gamma_{d,j} \geq \left(\frac{1}{d} \sum_{j=1}^{d} (\gamma_{d,j})^{\kappa} \right)^{1/\kappa} \quad \text{for} \quad 0 < \kappa \leq 1,$$

since $f(y) = y^{\kappa}$ is a concave function for $y > 0$. Thus,

$$\lim_{d\to\infty} \frac{1}{d} \sum_{j=1}^{d} (\gamma_{d,j})^{\kappa} = 0 \quad \text{for all} \quad 0 < \kappa \leq 1.$$

Finally, for every $\kappa \geq 1$ we can estimate $\gamma_{d,j} \geq (\gamma_{d,j})^{\kappa}$ since $\gamma_{d,j} \leq 1$ for $j = 1, \ldots, d$. Therefore, $\lim_{d\to\infty} d^{-1} \sum_{j=1}^{d} (\gamma_{d,j})^{\kappa} = 0$ also holds for $\kappa > 1$, and the proof is complete. $\square$

In the last part of this section, we give some examples to illustrate the results. In the following we only have to prove the embeddings, i.e. assumptions (A1) and (A2) from the beginning of this section.

Example 1 (Limiting cases $\mathcal{P}_d^{\gamma}$ and $\mathcal{H}_d^{\gamma}$)**.** To begin with, we check the case $\mathcal{F}_d^{\gamma} = \mathcal{P}_d^{\gamma}$. Then (A1) obviously holds with $C_{1,d} = 1$. To prove (A2), note that the algebraical inclusion $\mathcal{F}_d^{\gamma} \subset \mathcal{H}_d^{\gamma}$ is trivial by arguments given in Section 5. For $f \in \mathcal{F}_d^{\gamma} = \mathcal{P}_d^{\gamma}$ we calculate

$$\| f \mid \mathcal{H}_d^{\gamma} \|^2 \leq \sum_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \int_{[0,1]^{|\alpha|}} \| D^{\alpha} f \|_{\infty}^2 \, dx_\alpha \leq \| f \mid \mathcal{F}_d^{\gamma} \|^2 \cdot \sum_{\alpha \in \{0,1\}^d} \gamma_\alpha.$$

Hence, the norm of the embedding $\mathcal{F}_d^{\gamma} \hookrightarrow \mathcal{H}_d^{\gamma}$ is bounded by

$$\left(\sum_{\alpha \in \{0,1\}^d} \gamma_\alpha \right)^{1/2} = \left(\prod_{j=1}^{d} (1 + \gamma_{d,j}) \right)^{1/2} \leq \exp\left(\frac{1}{2} \sum_{j=1}^{d} \gamma_{d,j} \right).$$

So, with $a = 1$, $b = 1/2$ and $t = 1$ also assumption (A2) is fulfilled and we can apply the stated theorems for the space $\mathcal{F}_d^{\gamma} = \mathcal{P}_d^{\gamma}$.

We now turn to the case $\mathcal{F}_d^\gamma = \mathcal{H}_d^\gamma$. Unfortunately, the estimate above indicates that (A1) may not hold for $\mathcal{F}_d^\gamma = \mathcal{H}_d^\gamma$ with $C_{1,d} \leq 1$. Nevertheless, in this case assumption (A2) is true with $C_{2,d} = 1$, i.e., $a = 1$, $b = 0$ and $t = 1$. Therefore, we can apply Theorem 3 for this space. Then the problem is strongly polynomially tractable if $p(\gamma) < 1$. Moreover, we have polynomial tractability if $q(\gamma) < 1$. It is known that these conditions are also necessary, see, e.g., Theorem 12 in [3].

Example 2 **($C^{(1,\ldots,1)}$).** Consider the space

$$\mathcal{F}_d^\gamma = \{f \colon [0,1]^d \to \mathbb{R} \mid f \in C^{(1,\ldots,1)}, \text{ where } \|f \mid \mathcal{F}_d^\gamma\| = \max_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_\infty < \infty\}.$$

Since $\mathcal{P}_d^\gamma$ is a linear subset of $\mathcal{F}_d^\gamma$ and $\|\cdot \mid \mathcal{P}_d^\gamma\|$ is simply the restriction of $\|\cdot \mid \mathcal{F}_d^\gamma\|$ we have $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ with an embedding factor $C_{1,d} = 1$ and (A1) holds. For the factor $C_{2,d}$ of the embedding $\mathcal{F}_d^\gamma \hookrightarrow \mathcal{H}_d^\gamma$, the same estimates hold exactly as in the previous example and, moreover, the set inclusion is obvious. Therefore, also assumption (A2) is fulfilled and we can apply the theorems of this section to the space $\mathcal{F}_d^\gamma$.

Finally, the last example shows that even very high smoothness does not improve the conditions for tractability.

Example 3 **(C^∞).** Assume

$$\mathcal{F}_d^\gamma = \{f \colon [0,1]^d \to \mathbb{R} \mid f \in C^\infty, \text{ where } \|f \mid \mathcal{F}_d^\gamma\| = \sup_{\alpha \in \mathbb{N}_0^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_\infty < \infty\}.$$

Obviously, $\mathcal{P}_d^\gamma \subset C^\infty$, and functions from $\mathcal{P}_d^\gamma$ are at most linear in each coordinate. Hence, $D^\alpha f \equiv 0$ for all $\alpha \in \mathbb{N}_0^d \setminus \{0,1\}^d$. Therefore, once again we have

$$\|f \mid \mathcal{P}_d^\gamma\| = \max_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_\infty = \|f \mid \mathcal{F}_d^\gamma\| \quad \text{for all} \quad f \in \mathcal{P}_d^\gamma.$$

This yields $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ with an embedding factor $C_{1,d} = 1$. In addition, also (A2) can be concluded as in the examples above. So, even infinite smoothness leads to the the same conditions for tractability and the curse of dimensionality as before.

Note that in the last example we do not need to claim a product structure for the weights according to multi-indices $\alpha \in \mathbb{N}_0^d \setminus \{0,1\}^d$. Moreover, this example is a generalization of the space considered in [7]. For $\gamma_\alpha \equiv 1$ we reproduce the intractability result stated there.

In conclusion, we discuss the tractability behavior of L_∞-approximation defined on one of the spaces $\mathcal{F}_d^\gamma$ above using product weights which are independent of the dimension d, i.e.,

$$\gamma_{d,j} \equiv \gamma_j = \Theta(j^{-\beta}) \quad \text{for some} \quad \beta \geq 0.$$

This is a typical example in the theory of product weights, and $p(\gamma)$ is finite if and only if $\beta > 0$. If so then $p(\gamma) = 1/\beta$. See, e.g., Section 5.3.4 in [6].

If $\beta = 0$ then the problem is intractable due to Theorem 4, assertion (v), since $d^{-1}\sum_{j=1}^{d}\gamma_{d,j}$ does not tend to zero. For $\beta \in (0,1)$, easy calculus yields $q(\gamma) > 1$. So, using Theorem 2 we conclude polynomial intractability in this case. On the other hand, for all δ and κ with $0 < \delta < \kappa \le 1$, we have

$$\frac{\sum_{j=1}^{d} j^{-\kappa}}{d} = \frac{\sum_{j=1}^{d} j^{-\kappa} d^{\kappa-(1+\delta)}}{d^{\kappa-\delta}} \le \frac{\sum_{j=1}^{d} j^{-(1+\delta)}}{d^{\kappa-\delta}} \to 0 \quad \text{with} \quad d \to \infty$$

and if $\kappa > 1$ then the fraction obviously tends to zero, too. Hence, condition (vi) of Theorem 4 holds and the problem is weakly tractable if $\beta > 0$.

For $\beta = 1$, we use inequality (10) from Theorem 2 and estimate

$$\sum_{j=1}^{d} \gamma_{d,j} = \sum_{j=1}^{d} j^{-1} \ge c \cdot \ln(d+1)$$

for some positive c. Therefore,

$$n(\varepsilon, d; \mathcal{F}_d^{\gamma}) \ge \frac{1}{2^{2/3}}\,(d+1)^{c/3\cdot\ln(2)} \quad \text{for all} \quad d \in \mathbb{N},\, \varepsilon \in (0,1).$$

Hence, strong polynomial tractability does not hold. Moreover, it is easy to show that for $\beta = 1$ the sufficient condition $q(\gamma) < 1$ for polynomial tractability is not fulfilled. So, we do not know if polynomial tractability holds.

If $\beta > 1$ we easily see that $p(\gamma) = \frac{1}{\beta} < 1 = t$. Hence, Theorem 3 provides strong polynomial tractability in this case.

7 Final remarks

Note that the main result of this paper, the lower bound given in Theorem 1, can be easily transfered from $[0,1]^d$ to more general domains Ω. Indeed, the case $\Omega = [c_1, c_2]^d$, where $c_1 < c_2$, can be immediately obtained using our techniques. It turns out that in this case we have to modify estimate (6) by a constant which depends only on the length of the interval $[c_1, c_2]$. Thus, the general tractability behavior does not change.

Another extension of the results is possible if we consider the L_p-norms ($1 \le p < \infty$) instead of the L_∞-norm. We want to briefly discuss these norms for the unweighted case. Then the modifications for the weighted case are obvious. Following Novak and Woźniakowski [7] let

$$\mathcal{F}_{d,p} = \left\{ f\colon [c_1,c_2]^d \to \mathbb{R} \mid f \in C^\infty \text{ with } \|f \mid \mathcal{F}_{d,p}\| = \sup_{\alpha\in\mathbb{N}_0^d} \|D^\alpha f \mid L_p\| < \infty \right\}$$

for $1 \le p < \infty$ and $d \in \mathbb{N}$. Let $l = c_2 - c_1 > 0$. We want to approximate $f \in \mathcal{F}_{d,p}$ in the norm of L_p, i.e., we consider the n-th minimal error

$$e_p(n,d;\mathcal{F}_{d,p}) = \inf_{A_{n,d}\in\mathcal{A}_n} e_p^{\mathrm{wor}}(A_{n,d};\mathcal{F}_{d,p}) = \inf_{A_{n,d}\in\mathcal{A}_n} \sup_{f\in\mathcal{B}(\mathcal{F}_{d,p})} \left\| f - A_{n,d}(f) \mid L_p([c_1,c_2]^d) \right\| .$$

Without loss of generality we restrict ourselves to the case $[c_1,c_2] = [0,l]$. In order to conclude a lower bound analogue to Theorem 1, i.e., $e_p(n,d;\mathcal{F}_{d,p}) \ge 1$ for $n < 2^s$, we once again use Lemma 1 with $F = \mathcal{F}_{d,p}$ and $G = L_p([0,l]^d)$. The authors of [7] suggest to use the subspace $V_d^{(k)} \subset \mathcal{F}_{d,p}$ defined as

$$V_d^{(k)} = \operatorname{span}\left\{ g_i\colon [c_1,c_2]^d \to \mathbb{R},\, x \mapsto g_i(x) = \prod_{j=1}^{s} \left(\sum_{m=(j-1)k+1}^{jk} x_m \right)^{i_j} \mid i \in \{0,1\}^s \right\},$$

where $s = \lfloor d/k \rfloor$ and $k \in \mathbb{N}$ such that $kl \ge 2(p+1)^{1/p}$. Hence, if $l < 2(p+1)^{1/p}$ we have to use blocks of variables with size $k > 1$ in order to guarantee (5), i.e., to fulfill the condition

$$\|g \mid \mathcal{F}_{d,p}\| \le \|g \mid L_p\| \quad \text{for all} \quad g \in V_d^{(k)}. \tag{12}$$

Therefore, Novak and Woźniakowski defined $k = \lceil 2(p+1)^{1/p}/l \rceil$, but this is too small as the following example shows.

Take $l = 1$, i.e. $[c_1,c_2]^d = [0,1]^d$, and $p = 1$. Then $k = 4$ should be a proper choice, but for $g^*(x) = (x_1 + x_2 + x_3 + x_4) - 2$ we obtain $\|g^* \mid L_1\| = 7/15$ by using Maple, while $\|\partial g^*/\partial x_1 \mid L_1\| = 1$. This contradicts (12).

Proposition 3. Let $1 \le p < \infty$ and $k \in \mathbb{N}$ with

$$k \ge \left\lceil 8(p+1)^{2/p}/l^2 \right\rceil . \tag{13}$$

Then condition (12) holds for $V_d^{(k)} \subset \mathcal{F}_{d,p}$. Hence, the problem remains intractable since $e_p(n,d;\mathcal{F}_{d,p}) \ge 1$ for all $n < 2^{\lfloor d/k \rfloor}$.

Proof. Step 1. Due to the structure of functions g from $V_d^{(k)}$, it suffices to show

$$\left\| D^\alpha g \,|\, L_p([0,l]^{ks}) \right\| \leq \left\| g \,|\, L_p([0,l]^{ks}) \right\| \quad \text{for all} \quad g \in V_d^{(k)} \quad \text{and for every} \quad \alpha \in \mathcal{M}^{(k)},$$

where the set of multi-indices $\mathcal{M}^{(k)}$ is defined by

$$\mathcal{M}^{(k)} = \left\{ \alpha \in \{0,1\}^{ks} \,|\, \sum_{m \in I_j} \alpha_m \leq 1, \;\text{ for all } \; j = 1, \ldots, s \right\}$$

and $I_j = \{(j-1)k+1, \ldots, jk\}$. Similar to the proof of Theorem 1, we only consider the case $\alpha = e_t \in \{0,1\}^{ks}$ with $t \in I_j$. The rest then follows by induction. We can represent $g \in V_d^{(k)}$, as well as $D^{e_t} g$, by functions $a, b\colon [0,l]^{k(s-1)} \to \mathbb{R}$ such that

$$g(x) = a(\tilde{x}) \sum_{m=1}^{k} y_m + b(\tilde{x}) \quad \text{and} \quad D^{e_t} g(x) = a(\tilde{x}),$$

where $x = (x_{I_1}, \ldots, x_{I_{j-1}}, y, x_{I_{j+1}}, \ldots, x_{I_s}) \in [0,l]^{ks}$ and $\tilde{x} = (x_{I_1}, \ldots, x_{I_{j-1}}, x_{I_{j+1}}, \ldots, x_{I_s}) \in [0,l]^{k(s-1)}$, as well as $y = (y_1, \ldots, y_k) \in [0,l]^k$. Here x_{I_j} denotes the k-dimensional vector of components x_m with coordinates $m \in I_j$. Therefore, we can rewrite the inequality $\left\| D^{e_t} g \,|\, L_p([0,l]^{ks}) \right\| \leq \left\| g \,|\, L_p([0,l]^{ks}) \right\|$ as

$$\int_{[0,l]^{k(s-1)}} \int_{[0,l]^k} |a(\tilde{x})|^p \, dy \, d\tilde{x} \leq \int_{[0,l]^{k(s-1)}} \int_{[0,l]^k} \left| a(\tilde{x}) \sum_{m=1}^{k} y_m + b(\tilde{x}) \right|^p \, dy \, d\tilde{x}$$

such that it is enough to prove a point wise estimate of the inner integrals for fixed $\tilde{x} \in [0,l]^{k(s-1)}$ with $a = a(\tilde{x}) \neq 0$. Easy calculus yields

$$\int_{[0,l]^k} \left| a \sum_{m=1}^{k} y_m + b \right|^p \, dy = l^{p+k} \cdot \int_{[-1/2,1/2]^k} \left| a \sum_{m=1}^{k} z_m + b' \right|^p \, dz$$

for some constant $b' \in \mathbb{R}$. The right-hand side is minimized for $b' = 0$. So, we can estimate this integral from below by

$$\begin{aligned} \int_{[0,l]^k} \left| a \sum_{m=1}^{k} y_m + b \right|^p \, dy \quad &\geq \quad l^{p+k} \cdot |a|^p \cdot \int_{[-1/2,1/2]^k} \left| \sum_{m=1}^{k} z_m \right|^p \, dz \\ &= \quad l^p \cdot \int_{[0,l]^k} |a|^p \, dy \cdot \int_{[-1/2,1/2]^k} \left| \sum_{m=1}^{k} z_m \right|^p \, dz. \end{aligned}$$

Hence, it remains to show that the choice of k implies that

$$\int_{[-1/2,1/2]^k} \left| \sum_{m=1}^{k} z_m \right|^p dz \geq l^{-p}.$$

Step 2. In this last part, we will show by arguments from Banach space geometry that

$$\int_{[-1/2,1/2]^k} \left| \sum_{m=1}^{k} z_m \right|^p dz \geq \left(\frac{k}{2}\right)^{p/2} \cdot \frac{1}{2^p(1+p)} = \left(\frac{k}{2}\right)^{p/2} \cdot \int_{-1/2}^{1/2} |x|^p \, dx. \tag{14}$$

Obviously, we only need to prove the inequality for $k \geq 2$ since the equation on the right, as well as the case $k = 1$, are trivial. To abbreviate the notation, we define

$$f \colon \mathbb{R}^k \to \mathbb{R}, \quad z = (z_1, \ldots, z_k) \mapsto \sum_{m=1}^{k} z_m$$

for fixed $k \geq 2$.

For given vectors $z, \xi \in \mathbb{R}^k$, let $\langle z, \xi \rangle$ denote the scalar product $\sum_{m=1}^{k} z_m \xi_m$. In the special case $\xi = 1/\sqrt{k} \cdot (1, \ldots, 1) \in S^{k-1}$ it is $\langle z, \xi \rangle = t$ for a given $t \in \mathbb{R}$, if and only if, $f(z) = t\sqrt{k}$. Furthermore, note that every ξ in the k-dimensional unit sphere S^{k-1} uniquely defines a hyperplane $\xi^\perp = \{z \in \mathbb{R}^k \mid \langle z, \xi \rangle = 0\}$ perpendicular to ξ which contains zero. Therefore, for every $t \in [0, \infty)$, the set $\xi^\perp + t\xi = \{z \in \mathbb{R}^k \mid \langle z, \xi \rangle = t\}$ describes a parallel shifted hyperplane with distance t to the origin. Using Fubini's theorem, this leads to the following representation

$$\int_{[-1/2,1/2]^k} |f(z)|^p \, dz = 2 \cdot \int_{\substack{[-1/2,1/2]^k \\ \langle z, \xi \rangle \geq 0}} f(z)^p \, dz = 2 \cdot k^{p/2} \cdot \int_0^\infty t^p \left(\int_{\substack{[-1/2,1/2]^k \\ \langle z, \xi \rangle = t}} 1 \, dz \right) dt.$$

Now we see that the inner integral describes the $(k-1)$-dimensional volume

$$v(t) = \lambda_{k-1} \left([-1/2, 1/2]^k \cap (\xi^\perp + t\xi) \right)$$

of the parallel section of the unit cube with the hyperplane defined above. Because of Ball's famous theorem we know $v(0) \leq \sqrt{2}$, independent of k, see, e.g., [2, Chapter 7]. Moreover, $\xi^\perp$ provides a central hyperplane section of the unit cube such that we have

$$\int_0^\infty v(t) \, dt = \frac{1}{2} \cdot \lambda_k([-1/2, 1/2]^k) = \frac{1}{2}$$

and, by Brunn's theorem (see Theorem 2.3 in [2]), $v \geq 0$ is non-increasing on $[0, \infty)$. Thus, v is related to the distribution function of a certain non-negative real-valued random variable X, up to a normalizing factor, i.e. $v(t) = v(0) \cdot \mathbb{P}(X \geq t)$. Using Hölder's inequality we obtain $\mathbb{E}(X^{1+p}) \geq (\mathbb{E}X)^{1+p}$ and, respectively,

$$\int_0^\infty t^p v(t)\, dt \geq \frac{1}{v(0)^p (1+p)} \cdot \left(\int_0^\infty v(t)\, dt \right)^{1+p}$$

by integration by parts. Altogether we conclude inequality (14) and, with k bounded from below by (13), even

$$\int_{[-1/2,1/2]^k} |f(z)|^p \, dz \geq l^{-p}.$$

Therefore, the proof is complete. □

Using other methods, we can improve inequality (14) in Step 2 of the last proof. In detail, we can represent the integral on the left as an expectation $\mathbb{E}(|f(Y)|^p)$ with a suitable random vector Y. For $p = 2N$ with $N \in \mathbb{N}$ this can be calculated exactly. Finally, it turns out that it is enough to take

$$k \geq \begin{cases} \lceil 12/l^2 \rceil, & \text{if } 2 \leq p < 4 \\ \lceil 8/l^2 \rceil, & \text{if } 4 \leq p \end{cases}$$

in order to conclude the claimed intractability result for the L_p-approximation problem. Nevertheless, we want to stress the point that also with this improvements the lower bounds on k are not sharp since we know from [7] that in the limit case $p = \infty$ we can take $k = \lceil 2/l \rceil$. On the other hand, upper bounds for the k-dimensional integral, concluded using Hoeffding's inequality, yield that $k^{p/2}$ is the right order.

Acknowledgments

The author thanks A. Hinrichs, E. Novak and H. Woźniakowski for their useful hints and valuable comments on this paper.

References

[1] K. Deimling, Nonlinear Functional Analysis, Springer-Verlag, Berlin, 1985.

[2] A. Koldobsky, Fourier Analysis in Convex Geometry, Amer. Math. Soc., Providence, RI, 2005.

[3] F. Y. Kuo, G. W. Wasilkowski, and H. Woźniakowski, Multivariate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces, J. Approx. Theory, 152, 135–160, 2008.

[4] F. Y. Kuo, G. W. Wasilkowski, and H. Woźniakowski, On the power of standard information for multivariate approximation in the worst case setting, J. Approx. Theory, 158, 97–125, 2009.

[5] E. Novak, I. H. Sloan, J. F. Traub, and H. Woźniakowski, Essays on the Complexity of Continuous Problems, Europ. Math. Soc., Zürich, 2009.

[6] E. Novak, and H. Woźniakowski, Tractability of Multivariate Problems. Vol. I: Linear Information, Europ. Math. Soc., Zürich, 2008.

[7] E. Novak, and H. Woźniakowski, Approximation of infinitely differentiable multivariate functions is intractable, J. Complexity, 25, 398–404, 2009.

[8] E. Novak, and H. Woźniakowski, Tractability of Multivariate Problems. Vol. II: Standard Information for Functionals, Europ. Math. Soc., Zürich, 2010.

[9] I. H. Sloan, and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?, J. Complexity, 14, 1–33, 1998.

[10] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, Information-based Complexity, Academic Press Inc., Boston, 1988.

[11] G. Wahba, Spline Models for Observational Data, Soc. Indust. Appl. Math. (SIAM), Philadelphia, 1990.

[12] A. G. Werschulz, and H. Woźniakowski, Tractability of multivariate approximation over a weighted unanchored Sobolev space, Constr. Approx., 30, 395–421, 2009.