Optimal Order of Convergence and (In)Tractability of Multivariate Approximation of Smooth Functions

Erich Novak
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
email: novak@mathematik.uni-jena.de

Henryk Woźniakowski *
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
email: henryk@cs.columbia.edu

November 12, 2008

Abstract

We study the approximation problem for C^∞ functions $f : [0,1]^d \to \mathbb{R}$ with respect to a $W_p^{\mathbf{m}}$-norm. Here, $\mathbf{m} = [m, m, \dots, m]$, d times, with the norm of the target space defined in terms of up to m partial derivatives with respect to all d variables. The optimal order of convergence is infinite, hence excellent, but the problem is still intractable and suffers from the curse of dimensionality if $m \ge 1$. This means that the order of convergence supplies incomplete information concerning the computational difficulty of a problem. For $m = 0$ and $p = 2$, we prove that the problem is not polynomially tractable, but that it is weakly tractable.

1 Introduction

The (optimal) order of convergence, or rate of convergence, is an important concept of numerical analysis and approximation theory. The order of convergence measures how fast the minimal error $e(n)$ of algorithms using n function values or linear functionals goes to zero. Roughly speaking, if $e(n) = \Theta(n^{-\alpha})$ then to guarantee that the error is ε, we must take $n = \Theta(\varepsilon^{-1/\alpha})$ for $\varepsilon \to 0$. Hence, asymptotically in ε, the larger the order of convergence the easier the problem. However, it is not clear what this means for a fixed $\varepsilon > 0$; how long do we have to wait for the asymptotic behavior to occur?

*This author was partially supported by the National Science Foundation under Grant DMS-0608727 and by the Humboldt Research Award at the University of Jena.

In this paper we assume that the functions $f : [0,1]^d \to \mathbb{R}$ are infinitely many times differentiable with respect to all variables and that the sum of all normalized derivatives is bounded in the L_p-norm, for some $p \in [1,\infty]$. We approximate such functions with respect to the Sobolev norm $W_p^{\mathbf{m}}([0,1]^d)$ for $\mathbf{m} = [m,m,\dots,m]$, d times, for some non-negative integer m. This space is defined in Section 2. Here we only mention that the norm in the target space is defined in terms of up to m partial derivatives with respect to all d variables. We consider the worst case setting along with algorithms using arbitrary linear functionals as information operations on f. Here, d can be arbitrarily large. To stress the importance of d, we denote the minimal error $e(n)$ by $e(n,d)$.

Since the smoothness of the functions is unbounded, the optimal convergence rate of this multivariate approximation problem is infinite. That is, for any d and arbitrarily large r we have

$$e(n,d) \,=\, \mathcal{O}(n^{-r}) \quad \text{as} \quad n \,\to\, \infty.$$

Despite this excellent asymptotic speed of convergence, we prove that

$$e(n,d) \,=\, 1 \quad \text{for all} \quad n \,=\, 0,1,\dots,(m+1)^d - 1.$$

Let $n(\varepsilon,d)$ denote the smallest number of linear functionals that is needed to find an algorithm with error at most ε. The last result means that

$$n(\varepsilon,d) \,\ge\, (m+1)^d \quad \text{for all} \;\; \varepsilon \,\in\, (0,1) \text{ and } d \,\in\, \mathbb{N} := \{1,2,\dots,\}.$$

For $m \ge 1$, we have exponential dependence on d, which is called the curse of dimensionality. Hence, for $m \ge 1$, the multivariate approximation problem is intractable.

So, we see that the only hope of non-exponential dependence on d is when $m = 0$. In this case, we restrict ourselves to $p = 2$ and study multivariate approximation simply in the L_2-norm. Now the multivariate approximation problem is *weakly tractable*, i.e., $n(\varepsilon,d)$ does not depend exponentially on $\varepsilon^{-1} + d$. More precisely, we have

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

However, the problem remains *polynomially intractable*, i.e., no matter how large we choose C, p and q, the inequality

$$n(\varepsilon,d) \,\le\, C\,\varepsilon^{-p}\,d^{\,q}$$

does *not* hold for some $\varepsilon \in (0,1)$ and $d \in \mathbb{N}$.

These results illustrate that the optimal order of convergence does not tell us everything about the difficulty of solving the problem. We may have an excellent order of convergence, but an exponential dependence on d. Or equivalently, we must wait exponentially long to enjoy the excellent asymptotic behavior.

We add in passing that similar results hold also for some other multivariate problems. For example, consider multivariate integration studied in [5] for the Korobov space with the smoothness parameter $\alpha > 1$. In this case, algorithms can use only function values. Then

$$e(n,d) \,=\, \mathcal{O}\left(n^{-p}\right) \quad \text{as} \;\; n \to \infty, \qquad \text{for all} \;\; p \,<\, \alpha.$$

However,

$$e(n,d) \;=\; 1 \qquad \text{for} \;\; n = 0, 1, \dots, 2^d - 1, \tag{1}$$

which implies that

$$n(\varepsilon, d) \;\geq\; 2^d \quad \text{for all} \;\; \varepsilon \in (0,1).$$

That is, even for arbitrarily large α, despite an excellent order of convergence, this integration problem is *intractable*. Further examples can be found in [9].

We finally add that this paper has been written before the publication of the book [3]. We included the essence of this paper in [3] as Example 3.1.4 of Chapter 3 with the information that this example is based on a paper submitted for publication. In meantime, our book has been published sooner than the publication of this paper.

2 The problem

We consider functions from the class $C^\infty([0,1]^d)$ of infinitely differentiable functions defined on the d-dimensional cube $[0,1]^d$. Let $f \in C^\infty([0,1]^d)$. Obviously for any multi-index

$$\beta \;=\; [\beta_1, \beta_2, \dots, \beta_d] \;\in\; \mathbb{N}_0^d \qquad \text{with} \;\; \mathbb{N}_0 \;:=\; \{0, 1, 2, \dots\},$$

the function

$$D^\beta f \;:=\; \frac{\partial^{\beta_1+\beta_2+\cdots+\beta_d}}{\partial^{\beta_1} x_1 \partial^{\beta_2} x_2 \cdots \partial^{\beta_d} x_d}\; f$$

also belongs to $C^\infty([0,1]^d)$. For any $p \in [1,\infty]$ we also have $\|D^\beta f\|_{L_p} < \infty$, where L_p is the classical space of functions defined on $[0,1]^d$, i.e., for $p \in [1,\infty)$ we have

$$\|f\|_{L_p} \;=\; \left(\int_{[0,1]^d} |f(x)|^p \,\mathrm{d}x \right)^{1/p},$$

whereas for $p = \infty$, we have

$$\|f\|_{L_\infty} \;=\; \operatorname*{ess\,sup}_{x \in [0,1]^d} |f(x)|.$$

We restrict the class $C^\infty([0,1]^d)$ by taking the linear space

$$F \;=\; F_{d,p} \;:=\; \left\{ f \;\in\; C^\infty([0,1]^d) \;\middle|\; \|f\|_F \;:=\; \left(\sum_{\beta \in \mathbb{N}_0^d} \frac{1}{\beta!} \, \|D^\beta f\|_{L_p}^p \right)^{1/p} \;<\; \infty \right\},$$

with $\beta! = \prod_{j=1}^d \beta_j!$.

Hence, we deal with infinitely differentiable functions for which the sum of all normalized derivatives is bounded in L_p. This class is nonempty since $f \equiv 1$ belongs to F. Furthermore, all multivariate polynomials belong to F since the series with respect to β for a polynomial consists of only finitely many positive terms. In any case, we hope the reader agrees that F seems to be a "very small" set of functions.

For a given non-negative integer m, we consider the space $G = G_{d,m,p}$ given by

$$G = \left\{ f \in W_p^{\mathbf{m}}([0,1]^d) \;\middle|\; \|f\|_G := \left(\sum_{\beta\in\mathbb{N}_0^d:\ |\beta_j|\le m} \frac{1}{\beta!}\, \|D^\beta f\|_{L_p}^p \right)^{1/p} < \infty \right\}.$$

Hence, G is the Sobolev space $W_p^{\mathbf{m}}([0,1]^d)$ of functions whose partial derivatives up to order m in each variable belong to $L_p([0,1]^d)$. Note that for $m = 0$, the space $G_{d,0,p}$ is just $L_p([0,1]^d)$.

For any m and for all $f \in F$ we have $\|f\|_G \le \|f\|_F$. Let $P_{d,m}$ denote the linear space of polynomials of d variables that are of degree at most m in each variable. Clearly, $\dim(P_{d,m}) = (m+1)^d$ and

$$\|f\|_F = \|f\|_G \quad \text{for all} \quad f \in P_{d,m}.$$

Hence, the norms in F and G are the same for this $(m+1)^d$-dimensional subspace. As we shall see this property will be very important for our analysis.

For the classes $F_{d,p}$ and $G_{d,m,p}$, we consider the multivariate approximation problem APP_d with $\mathrm{APP}_d : F_{d,p} \to G_{d,m,p}$ given by

$$\mathrm{APP}_d f = f.$$

This is clearly a well-defined problem. Since

$$\|\mathrm{APP}_d\| := \sup_{f\in F_{d,p},\, \|f\|_F \le 1} \|\mathrm{APP}_d f\|_{G_{d,m,p}} = 1,$$

it is properly normalized. We approximate $\mathrm{APP}_d f$ by algorithms A_n that may now use not only function values but also arbitrary linear functionals, i.e.,

$$A_n(f) = \varphi_n\left(L_1(f), L_2(f), \dots, L_d(f)\right), \tag{2}$$

where $\varphi_n : \mathbb{R}^n \to G_{d,m,p}$ is some linear or non-linear mapping, and L_j is an arbitrary continuous linear functional whose choice may adaptively depend on the already computed values $L_1(f), L_2(f), \dots, L_{j-1}(f)$. The worst case error of A_n is defined by

$$e^{\mathrm{wor}}(A_n) = \sup_{f\in F_{d,p},\, \|f\|_{F_{d,p}}\le 1} \|\mathrm{APP}_d f - A_n(f)\|_{G_{d,m,p}}.$$

The minimal number of information operations needed to solve the problem to within ε is given by

$$n(\varepsilon, d) = n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p}) = \min\{\, n \,:\, \exists\, A_n \ \text{such that}\ e^{\mathrm{wor}}(A_n) \le \varepsilon \,\}.$$

Tractability means that $n(\varepsilon,d)$ does not depend exponentially on ε^{-1} and d. More precisely, we call a problem *weakly tractable* if

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

holds and *intractable* if this relation does not hold. Furthermore, a problem is *polynomially tractable* if there exist non-negative numbers C, p and q such that

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

If $q = 0$ above, then the problem is *strongly polynomially tractable.* For detailed discussion of tractability, the reader is referred to [3].

3 On the order of convergence

We first discuss the optimal order of convergence. Let

$$e(n,d) \;=\; \inf_{A_n} e^{\rm wor}(A_n)$$

be the minimal worst case error that can be achieved by using algorithms A_n of the form (2) based on n arbitrary linear functionals.

It is easy to see that for any $d \in \mathbb{N}$ and any $r > 0$ we have

$$e(n,d) \;=\; \mathcal{O}\left(n^{-r}\right) \quad \text{as} \;\; n \;\to\; \infty. \tag{3}$$

To prove this, consider first the spaces $C^s := C^s([0,1]^d)$ of s times continuously differentiable functions with the norm

$$\|f\|_s = \max_{x\in[0,1]^d} \max_{\beta\colon |\beta|\le s} |D^\beta f(x)|,$$

where $|\beta| = \sum_{i=1}^d \beta_i$. Now take

$$s_2 \;=\; d(r+m), \quad \text{and} \quad s_1 \;=\; d\,m.$$

Note that the norm of the space C^{s_1} is stronger than the norm of $G_{d,m,p}$. That is, $C^{s_1} \subseteq G_{d,m,p}$ and there exists a number C dependent on d, m and p such that $\|f\|_{G_{d,m,p}} \le C\,\|f\|_{s_1}$ for all $f \in C^{s_1}$.

Note that for any positive k, the class $F_{d,p}$ is a subset of the Sobolev space $W_p^{\mathbf{k}}([0,1]^d)$. If the embedding condition $k - s_2 > d/p$ holds then $W_p^{\mathbf{k}}([0,1]^d)$ and $F_{d,p}$ can both be regarded as subsets of C^{s_2}.

It is well-known that we can approximate functions from C^{s_2} in the norm of C^{s_1}, and then in the norm of $G_{d,m,p}$, by algorithms using n function values with worst case error of order $n^{-(s_2-s_1)/d}$. This result was probably first observed by Bakhvalov [1] for $m = 0$, which gives $s_1 = 0$. For general s_1, which is needed for $m \ge 1$, this result can be found, for instance, in the book of Triebel [8, p. 348].

Take $k = s_2 + 1 + d/p$. Then we conclude that functions from $F_{d,p}$ can be approximated in the norm of $G_{d,m,p}$ with worst case error of order n^{-r}, as claimed.

Since r can be arbitrarily large, the optimal order of convergence of the multivariate approximation problem for the class $F_{d,p}$ is formally infinite. This implies that for a fixed d, the minimal number of information operations goes to infinity slower then any power of ε^{-1}. That is, for any fixed d and any positive η we have

$$n^{\rm wor}(\varepsilon, {\rm APP}_d, F_{d,p}, G_{d,m,p}) \;=\; o(\varepsilon^{-\eta}) \quad \text{as} \;\; \varepsilon \to 0.$$

Again, this is very encouraging but one could say that this is possible only because the class $F_{d,p}$ is so small.

4 Intractability for $m \geq 1$

But how about tractability? How long do we have to wait to see this nice convergence of $e(n,d)$ to zero? We claim that

$$e(n,d) \;=\; 1 \qquad \text{for all} \;\; n = 0, 1, \dots, (m+1)^d - 1. \tag{4}$$

First of all, observe that the zero algorithm $A_n(f) = 0$ has worst case error at most 1 since APP_d is properly normalized. Hence, $e(n,d) \leq 1$. To prove the reverse inequality, take an arbitrary algorithm $A_n(f) = \varphi_n(L_1(f), \dots, L_n(f))$ that uses adaptive linear functionals L_j. We now show that $e^{\mathrm{wor}}(A_n) \geq 1$.

For $b = [b_1, b_2, \dots, b_d] \in \{0, 1, \dots, m\}^d$, define the functions

$$f_b(x) \;=\; \prod_{j=1}^{d} \left(x_j - \frac{1}{2}\right)^{b_j}.$$

The functions f_b are polynomials of degree at most m in each variable. Each b yields a new polynomial f_b and the set $\{f_b\}$ consists of $(m+1)^d$ linearly independent polynomials. Note also that $\|f_b\|_F = \|f_b\|_G$ since all terms $D^\beta f_b$ are zero if there is an index $\beta_j > m$, and hence the summation for the F-norm is the same as the summation for the G-norm. Let

$$g(x) \;=\; \sum_{b \in \{0,1,\dots,m\}^d} a_b \, f_b(x)$$

for some real numbers a_b. Again for any choice of a_b we have $\|g\|_F = \|g\|_G$.

We choose the scalars a_b such that $L_1(g) = 0$. Based on this zero value, the second linear functional L_2 is chosen, and we add the second equation for the scalars a_b by requiring that $L_2(g) = 0$. We do the same for all chosen linear functionals L_j based on the zero information, and we have n homogeneous linear equations for $\{a_b\}$,

$$\sum_{b \in \{0,1,\dots,m\}^d} a_b \, L_j(f_b) \;=\; 0 \quad \text{for} \;\; j \;= 1, 2, \dots, n\,.$$

Since we have $(m+1)^d > n$ unknowns, we can choose a non-zero vector $\{a_b\} = \{a_b^*\}$ satisfying these n equations. The function g with a_b^* is non-zero since the f_b's are linearly independent. Then $\|g\|_{F_{d,p}}$ is well-defined and positive. We finally define two functions

$$f_k \;=\; (-1)^k \, \frac{g}{\|g\|_{F_{d,p}}} \quad \text{for} \;\; k \;\in\; \{0, 1\}.$$

Note that

$$f_k \;\in\; F_{d,p} \quad \text{and} \quad \|f_k\|_{F_{d,p}} \;=\; \|f_k\|_{G_{d,m,p}} \;=\; 1.$$

Furthermore, $L_j(f_k) = 0$ for all $j = 1, 2, \dots, n$ and therefore $A_n(f_k) = \varphi(0, \dots, 0)$ does not depend on k. Hence,

$$\begin{aligned} e^{\mathrm{wor}}(A_n) \;&\geq\; \max_{f_0, f_1} \left(\|f_0 - \varphi(0,0,\dots,0)\|_G, \|f_1 - \varphi(0,0,\dots,0)\|_G\right) \\ &\geq\; \tfrac{1}{2} \left(\|f_0 - \varphi(0,0,\dots,0)\|_G + \|f_1 - \varphi(0,0,\dots,0)\|_G\right) \\ &\geq\; \tfrac{1}{2} \|f_0 - f_1\|_G \;=\; 1. \end{aligned}$$

This completes the proof of (4).

The essence of (4) is that the use of fewer than $(m+1)^d$ arbitrary linear functionals is not enough to reduce the error. Hence, if we want to guarantee that the error is at most $\varepsilon < 1$, then we have to use at least $(m+1)^d$ linear functionals. This means that

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p}) \ \ge\ (m+1)^d \qquad \text{for all} \ \ \varepsilon \in [0,1). \tag{5}$$

Hence if $m \ge 1$ then we have the curse of dimensionality and the multivariate approximation problem for the classes $F_{d,p}$ and $G_{d,m,p}$ is *intractable*. This means that the set $F_{d,m}$ is *not* so small after all.

5 Multivariate approximation for $m = 0$

Since for $m \ge 1$ we have intractability and the curse of dimensionality, we consider $m = 0$ with the hope that the curse of dimensionality is no longer present. In this case we restrict ourselves to $p = 2$ and analyze the multivariate approximation problem in detail.

We will need a couple of known general results, see, e.g., the books [3, 6, 7] where these results can be found. For $m = 0$ and $p = 2$, the space $G_{d,0,2}$ is just the Hilbert space $L_2 = L_2([0,1]^d)$ with the inner product

$$\langle f, g\rangle_{L_2} \ =\ \int_{[0,1]^d} f(x)g(x)\,\mathrm{d}x,$$

whereas $F = F_{d,2}$ is the unit ball of the Hilbert space with the inner product

$$\langle f, g\rangle_F \ =\ \sum_{\beta\in\mathbb{N}_0^d} \frac{1}{\beta!} \left\langle D^\beta f, D^\beta g\right\rangle_{L_2}.$$

Let $W_d = \mathrm{APP}_d^*\,\mathrm{APP}_d : F_{d,2} \to F_{d,2}$, where $\mathrm{APP}_d^* : L_2 \to F_{d,2}$ is the adjoint operator of APP_d. Obviously W_d is a self-adjoint positive semi-definite operator. It is well-known that $\lim_{n\to\infty} e(n,d) = 0$ iff W_d is compact. Since we already know that the limit of $e(n,d)$ is zero, we conclude that W_d is compact. Hence, $F_{d,2}$ has an orthonormal basis of the eigenfunctions $\eta_{d,j}$ of W_d, i.e.,

$$W_d\eta_{d,j} \ =\ \lambda_{d,j}\,\eta_{d,j}$$

with $\langle \eta_{d,j}, \eta_{d,k}\rangle_F = \delta_{j,k}$. We may assume that the non-negative eigenvalues $\lambda_{d,j}$ are ordered, i.e.,

$$\lambda_{d,1} \ \ge\ \lambda_{d,2} \ \ge\ \cdots \ge\ \lambda_{d,n} \ \ge\ \cdots\ \ge\ 0.$$

Obviously $\lim_{n\to\infty} \lambda_{d,n} = 0$. Since we now allow algorithms using arbitrary linear functionals, it is well-known that

$$e(n,d) \ =\ \sqrt{\lambda_{d,n+1}} \quad \text{for all} \ \ n = 0,1,\dots,$$

and that the algorithm

$$A_n(f) \ =\ \sum_{j=1}^n \langle f, \eta_{d,j}\rangle_F \ \eta_{d,j}$$

has worst case error equal to $e(n,d)$. We stress that although the algorithm A_n is linear and uses non-adaptive information, it minimizes the worst case error in the class of all non-linear algorithms using n arbitrary adaptive linear functionals.

5.1 Periodic case

We now restrict our attention to functions from the space $F = F_{d,2}$ that are periodic. By a periodic function $f \in F_{d,2}$ we mean that for $d = 1$ we have $f^{(\beta)}(1) = f^{(\beta)}(0)$ for all $\beta \in \mathbb{N}_0$, whereas for $d \geq 1$, we have $(D^\beta f)(x) = (D^\beta f)(y)$ if $|x_i - y_i| \in \{0,1\}$ for all i. That is, the values of all derivatives are the same if a component $x_i = 0$ of x is changed into $x_i = 1$. Hence, let

$$F_{d,2}^{\text{per}} = \{ f \in F_{d,2} \mid f \text{ is periodic} \}.$$

The space $F_{d,2}^{\text{per}}$ is equipped with the same norm as $F_{d,2}$. For example, for $j \in \mathbb{N}_0^d$, the functions

$$\prod_{k=1}^{d} \eta_{j_k}(x_j) \quad \text{with} \quad \eta_{j_k}(x) = \sin(2\pi j_k x) \ \text{ or } \ \eta_{j_k}(x) = \cos(2\pi j_k x)$$

belong to $F_{d,2}^{\text{per}}$. Note that the approximation problem is still properly normalized for the subspace $F_{d,2}^{\text{per}}$ since $\|\text{APP}_d\|_{F_{d,2}^{\text{per}} \to L_2} = 1$.

The subspace $F_{d,2}^{\text{per}}$ is much smaller than $F_{d,2}$. So if we establish a negative result for $F_{d,2}^{\text{per}}$, then the same result will be also true for the larger class $F_{d,2}$. Obviously, positive results for $F_{d,2}^{\text{per}}$ do not have to be true for $F_{d,2}$.

To verify tractability of the approximation problem defined over the subspace $F_{d,2}^{\text{per}}$, we need to find the eigenpairs of W_d. It will be instructive to consider first the univariate case $d = 1$. Define $\eta_1(x) = 1$, and for $k = 1, 2, \dots,$ define

$$\eta_{2k}(x) = \sqrt{2}\, e^{-2(\pi k)^2} \sin(2\pi\, k\, x), \quad \eta_{2k+1}(x) = \sqrt{2}\, e^{-2(\pi k)^2} \cos(2\pi\, k\, x).$$

It is easy to check that the sequence $\{\eta_k\}$ is orthonormal in the subspace $F_{1,2}^{\text{per}}$, i.e., $\langle \eta_k, \eta_s \rangle_{F_{1,2}} = \delta_{k,s}$. Define

$$K_1(x,y) = \sum_{j=1}^{\infty} \eta_j(x)\eta_j(y) \quad \text{for} \quad x, y \in [0,1].$$

We claim that K_1 is the *reproducing kernel* of $F_{1,2}^{\text{per}}$. That is, in particular, $K_1(\cdot, y) \in F_{1,2}^{\text{per}}$ for all $y \in [0,1]$, and $f(y) = \langle f, K_1(\cdot,y)\rangle_{F_{1,2}}$ for all $f \in F_{1,2}^{\text{per}}$ and all $y \in [0,1]$. Indeed, it is enough to check the last property. Observe that for arbitrary $\beta \in \mathbb{N}$ and $k \geq 1$, we have

$$\begin{aligned}\left\langle f^{(\beta)}, \eta_{2k}^{(\beta)} \right\rangle_{L_2} \eta_{2k}(y) \; &+ \; \left\langle f^{(\beta)}, \eta_{2k+1}^{(\beta)} \right\rangle_{L_2} \eta_{2k+1}(y) \\ &= \; (2\pi k)^{2\beta} \left(\langle f, \eta_{2k} \rangle_{L_2} \eta_{2k}(y) + \langle f, \eta_{2k+1} \rangle_{L_2} \eta_{2k+1}(y) \right).\end{aligned}$$

Therefore,

$$\begin{aligned}
\langle f, K_1(\cdot,y)\rangle_{F_{1,2}} &= \sum_{j=1}^{\infty} \langle f,\eta_j\rangle_{F_{1,2}}\,\eta_j(y) = \sum_{j=1}^{\infty}\sum_{\beta=0}^{\infty} \frac{1}{\beta!}\left\langle f^{(\beta)}, \eta_j^{(\beta)}\right\rangle_{L_2} \eta_j(y) \\
&= \sum_{j=1}^{\infty} \langle f,\eta_j\rangle_{L_2}\,\eta_j(y) + \sum_{\beta,k=1}^{\infty} \frac{\left\langle f^{(\beta)}, \eta_{2k}^{(\beta)}\right\rangle_{L_2} \eta_{2k}(y) + \left\langle f^{(\beta)}, \eta_{2k+1}^{(\beta)}\right\rangle_{L_2} \eta_{2k+1}(y)}{\beta!} \\
&= \langle f,\eta_1\rangle_{L_2} + \sum_{k=1}^{\infty} e^{(2\pi k)^2}\left(\langle f,\eta_{2k}\rangle_{L_2}\eta_{2k}(y) + \langle f,\eta_{2k+1}\rangle_{L_2}\eta_{2k+1}(y)\right) \\
&= \langle f,1\rangle_{L_2} + 2\sum_{k=1}^{\infty} \langle f, \sin 2\pi k\cdot\rangle_{L_2} \sin(2\pi k y) + \langle f, \cos 2\pi k\cdot\rangle_{L_2} \cos(2\pi k y).
\end{aligned}$$

The last series is the Fourier series for f evaluated at y. Since f is periodic and differentiable, this is equal to $f(y)$.

It now follows that the sequence $\{\eta_k\}$ is an orthonormal basis of the subspace $F_{1,2}^{\rm per}$. Indeed, it is enough to show that if $f \in F_{1,2}^{\rm per}$ and $\langle f,\eta_j\rangle_{F_{1,2}} = 0$ for all j then $f = 0$. Orthogonality of f to all η_j implies that $\langle f, K_1(\cdot,y)\rangle_{F_{1,2}} = 0$, and therefore $f(y) = 0$. Since this holds for all $y \in [0,1]$, we have $f = 0$, as claimed.

Note that for $k \neq s$, we have

$$0 = \langle \eta_k,\eta_s\rangle_{L_2} = \langle {\rm APP}_1\eta_k, {\rm APP}_1\eta_s\rangle_{L_2} = \langle \eta_k, {\rm APP}_1^*{\rm APP}_1\eta_s\rangle_F = \langle \eta_k, W_1\eta_s\rangle_F\,.$$

This means that $W_1\eta_s$ is orthogonal to all η_k except $k = s$. Hence,

$$W_1\eta_s = \lambda_s\eta_s,$$

and $\lambda_s = \langle \eta_s,\eta_s\rangle_{L_2}$. This yields

$$\lambda_1 = 1 \quad\text{and}\quad \lambda_{2k} = \lambda_{2k+1} = e^{-(2\pi\, k)^2} \quad \text{for} \quad k = 1,2,\dots\,.$$

For $d \geq 2$, it is easy to see that $F_{d,2}^{\rm per}$ is the tensor product of d copies of $F_{1,2}^{\rm per}$ and W_d is the d-fold tensor product of W_1. This implies that the eigenpairs of W_d are

$$W_d\,\eta_{d,j} = \lambda_{d,j}\eta_{d,j},$$

where $j = [j_1, j_2, \dots, j_d] \in \mathbb{N}^d$ and[1]

$$\eta_{d,j}(x) = \prod_{k=1}^{d} \eta_{j_k}(x_k) \quad\text{and}\quad \lambda_{d,j} = \prod_{k=1}^{d} \lambda_{j_k}.$$

Hence, the eigenvalues for the d-dimensional case are given as the products of the eigenvalues for the univariate case. To find out the nth optimal error $e(n,d)$, we must order the sequence

[1] Knowing the eigenvalues of W_d, it is possible to apply general results, such as Theorems 5.1 and 6.1 of [2] and Theorem 3.1 of [10], to conclude the behavior of $n(\varepsilon,d)$. We prefer, however, to derive the bounds on $n(\varepsilon,d)$ directly and obtain sharp bounds and asymptotic constants for a fixed d.

$\{\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}\}_{j\in\mathbb{N}^d}$. The square root of the $(n+1)$st largest eigenvalue is then $e(n,d)$. Thus, $e(n,d)\le\varepsilon$ iff n is at least the cardinality of the set of all eigenvalues $\lambda_{d,j}>\varepsilon^2$. If we denote $n(\varepsilon,d):=n^{\rm wor}(\varepsilon,{\rm APP}_d,F_{d,2}^{\rm per},L_2)$ as the minimal number of linear functionals needed to solve the problem to within ε, then

$$n(\varepsilon,d)\;=\;\big|\big\{\,j=[j_1,j_2,\dots,j_d]\,\in\,\mathbb{N}^d\;:\;\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}\;>\;\varepsilon^2\,\big\}\big|.$$

Clearly, $n(\varepsilon,d)=0$ for all $\varepsilon\ge1$ since the largest eigenvalue is 1. It is also easy to see that $n(\varepsilon,d)=1$ for all $\varepsilon\in(e^{-2\pi^2},1)$ since the second largest eigenvalue is $\lambda_2=e^{-4\pi^2}$. For $d=1$, note that $e^{-(2\pi k)^2}>\varepsilon^2$ iff $k\le\lceil\sqrt{2\ln\varepsilon^{-1}}/(2\pi)\rceil\,-\,1$. This yields that

$$n(\varepsilon,1)\;=\;2\,\left\lceil\frac{1}{2\pi}\sqrt{2\,\ln\,\varepsilon^{-1}}\right\rceil\;-\;1\;=\;\frac{\sqrt{2}}{\pi}\sqrt{\ln\,\frac{1}{\varepsilon}}\;+\;\mathcal{O}(1)\qquad\text{as }\;\varepsilon\;\to\;0.$$

For $d\ge1$, we have the formula

$$n(\varepsilon,d+1)\;=\;\sum_{j=1}^{\infty}n\left(\frac{\varepsilon}{\sqrt{\lambda_j}},d\right)\;=\;n(\varepsilon,d)\;+\;2\sum_{k=1}^{\infty}n\left(\varepsilon\,e^{2(\pi k)^2},d\right),$$

which relates the cases for $d+1$ and d. The last two series are only formally infinite, since for large j and k the corresponding terms are zero. More precisely, to obtain a positive $n(\varepsilon\,e^{2(\pi k)^2},d)$ we need to assume that $\varepsilon\,e^{2(\pi k)^2}<1$. Let

$$k_\varepsilon\;=\;\left\lceil\frac{\sqrt{2}}{2\pi}\sqrt{\ln\,\frac{1}{\varepsilon}}\right\rceil\;-\;1.$$

Then

$$n(\varepsilon,d+1)\;=\;n(\varepsilon,d)\;+\;2\sum_{k=1}^{k_\varepsilon}n\left(\varepsilon\,e^{2(\pi k)^2},d\right).$$

We now show by induction on d that

$$n(\varepsilon,d)\;=\;\Theta\left(\left(\ln\,\frac{1}{\varepsilon}\right)^{d/2}\right)\qquad\text{as }\;\varepsilon\;\to\;0.\tag{6}$$

This is clearly true for $d=1$. If it is true for d, then using the formula for $n(\varepsilon,d+1)$ we easily see that we can bound $n(\varepsilon,d+1)$ from above by $\mathcal{O}((\ln1/\varepsilon)^{(d+1)/2})$ since k_ε is of order $(\ln\,1/\varepsilon)^{1/2}$. We can estimate $n(\varepsilon,d+1)$ from below by taking $k_\varepsilon/2$ terms and using the lower bound on $n(\varepsilon\,e^{2(\pi k)^2},d)$, which again yields an estimate of order $(\ln\,1/\varepsilon)^{(d+1)/2}$.

Let us pause and ask what (6) means. From one point of view, this estimate of $n(\varepsilon,d)$ is quite positive since we have weak dependence on ε only through $\ln\,1/\varepsilon$. But if d is large, (6) may suggest that we have an exponential dependence on d, and the problem may be intractable. As we already know, the factors in the big theta notation are very important and so we can claim nothing based solely on (6). We need more information about how $n(\varepsilon,d)$ behaves. We now prove that

$$C_d\;:=\;\lim_{\varepsilon\to0}\;\frac{n(\varepsilon,d)}{(\ln\,1/\varepsilon)^{d/2}}\;=\;\frac{1}{(2\pi)^{d/2}\,\Gamma(1+d/2)},\tag{7}$$

establishing the asymptotic behavior of $n(\varepsilon, d)$ as ε tends to zero.

For $d = 1$, we have already shown the formula $C_1 = \sqrt{2}/\pi$. Assume that C_d is the asymptotic constant for d, and consider the case $d+1$. For every positive δ there exists $\varepsilon_d = \varepsilon_{d,\delta} \in (0,1)$ such that for all $\varepsilon \in (0, \varepsilon_d]$, we have

$$n(\varepsilon, d) \;=\; C_d(1+g(\varepsilon)) \left(\ln \frac{1}{\varepsilon}\right)^{d/2} \quad \text{with } |g(\varepsilon)| \;\le\; \delta, \quad \text{and} \quad \lim_{\varepsilon\to 0} g(\varepsilon) \;=\; 0.$$

Define

$$k_\varepsilon^* \;=\; \left\lceil \frac{\sqrt{2}}{2\pi} \sqrt{\ln \frac{\varepsilon_d}{\varepsilon}} \right\rceil \;-\; 1.$$

Note that $k_\varepsilon - k_\varepsilon^* = \mathcal{O}(1)$ as $\varepsilon \to 0$. We have

$$\begin{aligned} n(\varepsilon, d+1) \;&=\; C_d\,(1+g(\varepsilon)) \left(\ln \frac{1}{\varepsilon}\right)^{d/2} \\ &\quad +2\,C_d \sum_{k=1}^{k_\varepsilon^*} \left((1+g\left(\varepsilon\, e^{2(\pi k)^2}\right)\right) \left(\ln \frac{1}{\varepsilon} \;-\; 2(\pi\, k)^2\right)^{d/2} \\ &\quad +2 \sum_{k=k_\varepsilon^*+1}^{k_\varepsilon} n(\varepsilon\, e^{2(\pi k)^2}, d). \end{aligned}$$

Note that for $k \in [k_\varepsilon^*+1, k_\varepsilon]$ we have $n(\varepsilon\, e^{2(\pi k)^2}, d) \le n(\varepsilon_d, d)$, and therefore

$$\sum_{k=k_\varepsilon^*+1}^{k_\varepsilon} n(\varepsilon\, e^{2(\pi k)^2}, d) \;\le\; (k_\varepsilon - k_\varepsilon^*)\; n(\varepsilon_d, d) \;=\; \mathcal{O}((\ln \varepsilon^{-1})^{d/2}).$$

Now consider the terms for which $k \in [1, k_\varepsilon^*]$. Then $\varepsilon\, e^{2(\pi k)^2} \le \varepsilon_d$. For ε tending to zero, we have

$$\begin{aligned} \sum_{k=1}^{k_\varepsilon^*} \left(1+g\left(\varepsilon\, e^{2(\pi k)^2}\right)\right) \left(\ln \frac{1}{\varepsilon} \;-\; 2(\pi k)^2\right)^{d/2} \;&=\; (1+o(1)) \int_1^{k_\varepsilon^*} \left(\ln \frac{1}{\varepsilon} \;-\; 2(\pi x)^2\right)^{d/2} \,\mathrm{d}x \\ &=\; \frac{1+o(1)}{\sqrt{2}\,\pi} \left(\ln \frac{1}{\varepsilon}\right)^{(d+1)/2} \int_0^1 (1-x^2)^{d/2}\,\mathrm{d}x \\ &=\; \frac{1+o(1)}{\sqrt{2}\,\pi} \left(\ln \frac{1}{\varepsilon}\right)^{(d+1)/2} \tfrac{1}{2}\, B(\tfrac{1}{2}, 1+d/2), \end{aligned}$$

where $B(x,y)$ is the beta function and is related to the Gamma function by $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. This proves that

$$n(\varepsilon, d+1) \;=\; C_{d+1}(1+o(1))(\ln\, 1/\varepsilon)^{(d+1)/2}$$

as ε goes to zero, with

$$C_{d+1} \;=\; \frac{B(\frac{1}{2}, 1+d/2)\; C_d}{\sqrt{2}\,\pi} \;=\; \frac{\Gamma(1/2)\,\Gamma(1+d/2)}{\sqrt{2}\,\pi\,\Gamma(1+(d+1)/2)}\, C_d.$$

Solving this recurrence, we obtain

$$C_{d+1} \;=\; \frac{\Gamma(\frac{1}{2})^d\,\Gamma(\frac{3}{2})}{(\sqrt{2}\,\pi)^d\,\Gamma(1+(d+1)/2)}\,C_1 \;=\; \frac{1}{(2\pi)^{(d+1)/2}\,\Gamma(1+(d+1)/2)},$$

which agrees with the asymptotic formula (7).

We stress that the asymptotic constant C_d in (7) is super exponentially small in d due to the presence of $\Gamma(1+d/2)$ in the denominator. This property raises our hopes that we can beat the apparent exponential dependence on d. Indeed, assume for a moment that the limit in (7) is uniform in d. That is, suppose that there exists a positive ε_0 such that for all $\varepsilon \in (0,\varepsilon_0)$ and all d, we have

$$n(\varepsilon,d) \;\le 2\,C_d\left(\ln\frac{1}{\varepsilon}\right)^{d/2} \;=\; \frac{2\,\left(\ln\varepsilon^{-1}\right)^{d/2}}{(2\pi)^{d/2}\,\Gamma(1+d/2)}.$$

It can be easily checked that $x^{d/2}/\Gamma(1+d/2) \le \exp(x)$ for all $x \ge 1$. Therefore

$$n(\varepsilon,d) \;\le\; \frac{2}{(2\pi)^{d/2}}\,\frac{1}{\varepsilon}.$$

Hence, we have *strong polynomial tractability* if (7) holds uniformly in d.

We now return to the proof of (7) with the new task of checking whether ε_d can be uniformly bounded from below. Unfortunately, this is *not* true. It is enough to take $\varepsilon^2 \in [\lambda_4,\lambda_2)$ to realize that we can take $d-1$ indices $j_i = 1$ and the remaining index $j_i = 2$ to obtain the eigenvalue $\lambda_{d,j} = \lambda_2 > \varepsilon^2$. Hence $n(\varepsilon,d) \ge d$, which contradicts strong polynomial tractability.

In fact, polynomial tractability does not even hold, i.e., there is no upper bound of the form

$$n(\varepsilon,d) \le C\,\varepsilon^{-p}\,d^q \qquad \forall\;\varepsilon\in(0,1),\; d\in\mathbb{N}.$$

This follows from the general observation that as long as the largest eigenvalue is 1 and the second largest eigenvalue λ_2 for $d=1$ is positive, then there is *no* polynomial tractability. Indeed, for an arbitrary integer k and arbitrary $d > k$, consider the eigenvalues $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}$ with $d-k$ indices j_i equal to 1 and k indices j_i equal to 2. Then we have at least $\binom{d}{k} = \Theta(d^k)$ eigenvalues equal to λ_2^k. It is enough to take, say, $\varepsilon^2 = \lambda_2^k/2$ to realize that $n(\sqrt{\lambda_2^k/2},d)$ is at least of order d^k. Since k can be arbitrary, this contradicts not only strong polynomial tractability, but also polynomial tractability.

Well, we are back to square one. Despite the exponentially small asymptotic constants, we have *polynomial intractability* of the multivariate problem for $m=0$. Hence, the only remaining hope for a positive result is to weaken the notion of polynomial tractability, and in particular, to check if weak tractability holds. Here we will finally report good news.

As in [10], let $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d} > \varepsilon^2$ and let k be the number of indices $j_i \ge 2$. Then $(d-k)_+$ indices are equal to 1. Note that $\lambda_2^k > \varepsilon^2$ implies that

$$k \le a(\varepsilon) \;:=\; \left\lceil \frac{2\,\ln\varepsilon^{-1}}{\ln\,\lambda_2^{-1}} \right\rceil \;-\;1.$$

So we have at least $(d-a(\varepsilon))_+$ indices equal to 1. Observe also that $j_i \le n(\varepsilon,1)$. Thus

$$\binom{d}{(d-a(\varepsilon))_+} \le n(\varepsilon,d) \le \binom{d}{(d-a(\varepsilon))_+} n(\varepsilon,1)^{a(\varepsilon)}.$$

For a fixed ε and for d tending to infinity, we have

$$n(\varepsilon,d) = \Theta\left(d^{\lceil 2 \ln \varepsilon^{-1}/\ln \lambda_2^{-1}\rceil - 1}\right)$$

with the factors in the big-theta notation depending now on ε^{-1}.

For arbitrary d and $\varepsilon \in (0,1)$ we conclude that

$$n(\varepsilon,d) \le \frac{(d+a(\varepsilon))^{a(\varepsilon)}}{a(\varepsilon)!} \left(2\left\lceil \frac{1}{2\pi}\sqrt{2\ln\frac{1}{\varepsilon}}\right\rceil - 1\right)^{a(\varepsilon)}.$$

This implies that

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

which means that weak tractability indeed holds.

Hence, we have mixed news for the periodic case of the approximation problem. We have polynomial intractability, which obviously implies polynomial intractability for the original non-periodic case. But we have weak tractability for the periodic case, but it is not yet clear whether this good property extends to the non-periodic case.

5.2 Weak tractability

We now show that weak tractability holds not only for the original non-periodic case, but it also holds for a much larger space of less smooth functions. Namely, define

$$F_{d,2}^1 = G_{d,1,2} = W_2^{\mathbf{1}}([0,1]^d)$$

as the Sobolev space of functions whose partial derivatives up to order one in each variable belong to $L_2 = L_2([0,1]^d)$. The norm in $F_{d,2}^1$ is defined as in $G_{d,1,2}$. Clearly

$$F_{d,2} \subseteq F_{d,2}^1 \quad \text{and} \quad \|f\|_{F_{d,2}^1} \le \|f\|_{F_{d,2}} \quad \text{for all} \quad f \in F_{d,2}.$$

Again, consider first the case $d=1$, and the subspace $\tilde{F}_{1,2}^1$ of periodic functions from $F_{d,2}^1$. Periodicity now means that $f(1)=f(0)$. Proceeding as before, it is easy to check that the functions $\eta_1 = 1$, and

$$\eta_{2k}(x) = \frac{\sqrt{2}}{\sqrt{1+(2\pi k)^2}}\sin(2\pi kx), \quad \eta_{2k+1}(x) = \frac{\sqrt{2}}{\sqrt{1+(2\pi k)^2}}\cos(2\pi kx)$$

are orthonormal in $\tilde{F}_{1,2}^1$, and the function

$$K_1(x,y) = \sum_{j=1}^{\infty} \eta_j(x)\eta_j(y)$$

is the reproducing kernel of $\tilde{F}^1_{1,2}$. Therefore the sequence $\{\eta_j\}$ forms a basis of $\tilde{F}^1_{1,2}$. The eigenvalues $\lambda_j^{\rm per}$ of $W_1 = {\rm APP}_1^*{\rm APP}_1 : \tilde{F}^1_{1,2} \to \tilde{F}^1_{1,2}$ are

$$\lambda_1^{\rm per} = 1 \quad\text{and}\quad \lambda_{2k}^{\rm per} \;=\; \lambda_{2k+1}^{\rm per} \;=\; \frac{1}{1+(2\pi k)^2} \;\text{ for }\; k \;=\; 1,2,\ldots\,.$$

We now turn to the space $F^1_{1,2}$ of non-periodic functions. Define

$$g(x) = x - \tfrac{1}{2} \,+\, \sum_{k=1}^{\infty} \frac{1}{\pi k(1+(2\pi k)^2)}\,\sin(2\pi k x) \quad\text{for}\;\; x \in [0,1].$$

It is easy to check that g belongs to $F^1_{1,2}$ and is orthogonal to all η_j. Note that $g(1) = -g(0) = \frac{1}{2}$, hence $g \notin \tilde{F}^1_{1,2}$. For $f \in F^1_{1,2}$, let

$$h_f(x) \;=\; f(x) - [f(1)-f(0)]\; g(x).$$

Then $h_f \in \tilde{F}^1_{1,2}$. Hence,

$$f \;=\; [f(1)-f(0)]\,g \,+\, h_f \;\text{ for all }\; f \,\in F^1_{1,2}. \tag{8}$$

This proves that the reproducing kernel of $F^1_{1,2}$ is

$$K_1^{\rm non-per}(x,y) = \frac{g(x)}{\|g\|_{F^1_{1,2}}}\,\frac{g(y)}{\|g\|_{F^1_{1,2}}} \,+\, \sum_{j=1}^{\infty} \eta_j(x)\eta_j(y).$$

The decomposition (8) suggests that we first compute $L_1(f) = f(1) - f(0)$ and then approximate the function $h_f = f - L_1(f)g$. Note that

$$\langle f,\eta_j\rangle_{F^1_{1,2}} \;=\; L_1(f)\,\langle g,\eta_j\rangle_{F^1_{1,2}} \,+\, \langle h_f,\eta_j\rangle_{F^1_{1,2}} \;=\; \langle h_f,\eta_j\rangle_{F^1_{1,2}}$$

and

$$\|f\|^2_{F^1_{1,2}} \;=\; L_1(f)^2\|g\|^2_{F^1_{1,2}} \,+\, \|h_f\|^2_{F^1_{1,2}}.$$

Hence, approximation of functions from the unit ball of $F^1_{1,2}$ with n information evaluations is not harder than approximation of periodic functions from the unit ball of $\tilde{F}^1_{1,2}$ with $n-1$ information evaluations, and not easier than the periodic case with n evaluations.

Let $\lambda_j^{\rm non-per}$ denote the ordered sequence of eigenvalues of

$$W_1^{\rm non-per} = {\rm APP}_1^*{\rm APP}_1 : F^1_{1,2} \to F^1_{1,2}$$

for the non-periodic case. We have

$$W_1 f\,(x) = \int_0^1 K_1^{\rm non-per}(x,y) f(y)\,{\rm d}y,$$

which easily yields from the form of $K_1^{\text{non-per}}$ that all eigenvalues λ_j^{per} for the periodic case are also the eigenvalues for the non-periodic case, and we have one extra eigenvalue

$$\lambda = \frac{\|g\|^2_{L_2([0,1])}}{\|g\|^2_{L_2([0,1])} + \|g'\|^2_{L_2([0,1])}} < 1.$$

Therefore $\lambda_1^{\text{non-per}} = 1$, and $\lambda_2^{\text{non-per}} < \lambda_1^{\text{non-per}}$, as well as

$$\lambda_j^{\text{per}} \le \lambda_j^{\text{non-per}} \le \lambda_{j-1}^{\text{per}} \quad \text{for all } j \ge 2.$$

Hence, $\lambda_j^{\text{non-per}} = \Theta(j^{-1})$.

We turn to the case $d \ge 2$. Since $F^1_{d,2}$ is the d fold tensor product of $F^1_{1,2}$, the eigenvalues of $W_d = \text{APP}_d^*\text{APP}_d : F^1_{d,2} \to F^1_{d,2}$ are products of $\lambda_{j_1}^{\text{non-per}} \cdots \lambda_{j_d}^{\text{non-per}}$ for $j_i \in \mathbb{N}$. In Chapter 5 of [3] we prove that linear tensor product problems are weakly tractable as long as the eigenvalues for $d = 1$ satisfy the following two conditions:

- the second largest eigenvalue is smaller than the largest eigenvalue,
- the nth largest eigenvalue goes to zero faster than $(\ln n)^{-2}(\ln\ln n)^{-2}$.

These two assumptions hold in our case, and therefore the approximation problems for the space $F^1_{d,2}$ as well as for the smaller space $F_{d,2}$ are weakly tractable.

In fact, in this case we can say more on $n(\varepsilon, d)$ due to Theorems 5.1 and 6.1 of [2], see also Chapter 8 of [3]. Namely, there exist positive numbers t and C_t such that

$$n(\varepsilon, d) \le C_t \exp\big(t \ln(\varepsilon^{-1})\,(1 + \ln(d))\big) \quad \text{for all} \quad \varepsilon \in (0,1],\ d \in \mathbb{N}.$$

Furthermore, the infimum over t for which the last estimate holds is

$$\max\left(2, \frac{2}{\ln(1/\lambda_2^{\text{non-per}})}\right).$$

This obviously also implies weak tractability, and in addition tells us much more about the behavior of $n(\varepsilon, d)$.

6 Summary

We summarize our results for the multivariate approximation problem

$$\text{APP}_d : F_{d,p} \to G_{d,m,p}, \qquad d \in \mathbb{N}.$$

Theorem 1

- The order of convergence is infinite for any p and m, see (3).
- For $m \geq 1$ and any p, the problem is intractable, see (5).
- For $m = 0$ and $p = 2$, the problem is not polynomially tractable, but it is weakly tractable.

We end this paper with an open problem which is also presented in [3]: For $m \geq 1$, we have intractability for any p, whereas for $m = 0$ and $p = 2$ we have weak tractability but polynomial intractability. This leaves the case $m = 0$ and $p \neq 2$.

Furthermore, the partially positive result for $m = 0$ and $p = 2$ was obtained by assuming that algorithms use arbitrary linear functionals. It is not clear what happens if we allow algorithms that can only use function values.

- Consider multivariate approximation as before with $m = 0$ and $p \neq 2$, and algorithms using arbitrary linear functionals. Verify whether weak or polynomial tractability hold.
- Consider multivariate approximation as before with $m = 0$ and arbitrary $p \in [1, \infty]$, and algorithms using only function values. Verify whether weak tractability holds.

This problem for $p = \infty$ has been recently solved in [4] by proving that multivariate approximation even for arbitrary linear functionals is still intractable.

References

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

[2] M. Gnewuch and H. Woźniakowski, Generalized tractability for multivariate problems, Part II: Linear tensor product problems, linear information, unrestricted tractability, to appear in *Found. Comput. Math.*.

[3] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume 1, Linear Information, European Math. Soc., Zürich, 2008.

[4] E. Novak and H. Woźniakowski, Approximation of infinitely differentiable multivariate functions is intractable, to appear in *J. Complexity* .

[5] I. H. Sloan, H. Woźniakowski, An intractability result for multiple integration, *Math. Comp.* **66**, 1119–1124, 1997.

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

[7] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.

[8] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.

[9] A. G. Werschulz and H. Woźniakowski, Tractability of multivariate approximation over a weighted unanchored Sobolev space, to appear in Constructive Approximation.

[10] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, *J. of Computing and Information* **4**, 1–19, 1994.