Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings

Thomas Kühn* Sebastian Mayer† Tino Ullrich†‡

March 14, 2016

Abstract

In this paper, we reveal a new connection between approximation numbers of periodic Sobolev type spaces, where the smoothness weights on the Fourier coefficients are induced by a (quasi-)norm $\|\cdot\|$ on $\mathbb{R}^d$, and entropy numbers of the embedding $\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty$. This connection yields preasymptotic error bounds for approximation numbers of isotropic Sobolev spaces, spaces of analytic functions, and spaces of Gevrey type in L_2 and H^1, which find application in the context of Galerkin methods. Moreover, we observe that approximation numbers of certain Gevrey type spaces behave preasymptotically almost identical to approximation numbers of spaces of dominating mixed smoothness. This observation can be exploited, for instance, for Galerkin schemes for the electronic Schrödinger equation, where mixed regularity is present.

1 Introduction

Approximation numbers, also known as linear n-width, are one of the fundamental concepts in approximation theory. In Hilbert space settings, they describe the worst-case error that occurs when we approximate a class of functions by projecting them onto the optimal finite-dimensional subspace. Hence, approximation numbers are also of interest in the numerical analysis of partial differential equations (PDE) as they provide reliable a-priori error estimates for certain Galerkin methods. In this context, approximation numbers related to isotropic Sobolev spaces, Sobolev spaces of mixed regularity, and spaces of Gevrey type appear naturally. Subject of this paper are preasymptotic bounds for these approximation numbers, which substantially improve the known error bounds in high-dimensional settings.

*University of Leipzig, Augustusplatz 10, 04109 Leipzig, Germany.

†Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany.

‡Corresponding author, Email: tino.ullrich@hcm.uni-bonn.de, Phone: +492287362224, Fax: +492287362251

Bounds, which describe the decay in the rank $n \in \mathbb{N}$ of the optimal projection operator, have been known for decades for the aforementioned approximation numbers. In high-dimensional settings, where the functions to be approximated depend on a large number of variables $d \in \mathbb{N}$, these classical bounds become problematic. They only inadequately capture the effects of the approximation problem's dimensionality d. We state this issue precisely in the subsequent Sections 1.2, 1.3, and 1.4. For the moment, we only stress that this has consequences in at least two respects when the dimension d becomes large. For one, the classical bounds are trivial until n is exponentially large, say $n > 2^d$. This severely limits their applicability. Moreover, from the viewpoint of information-based complexity, it is impossible to determine the tractability of the approximation problem rigorously. To address any of the two issues, a first step is to uncover how the equivalence constants in the classical error bounds depend on the dimension d. As it turns out this is often not enough. In fact, one has to determine explicitly how the approximation numbers behave *preasymptotically*, that is, for small $n < 2^d$. This typically involves to find good estimates for complicated combinatorial problems that evolve from the structure of the smoothness spaces. As we will see, the preasymptotic behavior can be completely different from the asymptotic behavior.

1.1 An abstract characterization result: counting via entropy

The essence of this paper is that for certain relevant approximation numbers of periodic isotropic Sobolev spaces and periodic spaces of Gevrey type, it is not necessary to compute preasymptotics by hand. Instead we exploit that the approximation numbers can be determined by covering certain ℓ_p^d-unit balls with ℓ_∞^d-balls, a problem which is already well understood. In fact, this turns out to be just a special case of a general characterization result. This holds true for Sobolev type spaces $H^{\mathbf{w}}(\mathbb{T}^d)$ defined on the d-torus $\mathbb{T}^d = [0, 2\pi]^d$, where the smoothness weights $\mathbf{w} = (w(k))_{k \in \mathbb{Z}^d}$ take a special form. The spaces $H^{\mathbf{w}}(\mathbb{T}^d)$ are given by

$$H^{\mathbf{w}}(\mathbb{T}^d) = \{ f \in L_2(\mathbb{T}^d) : \sum_{k \in \mathbb{Z}^d} w(k)^2 |c_k(f)|^2 < \infty \}$$

where $c_k(f)$ denotes the kth Fourier coefficient. The weights in the sequence $\mathbf{w}$ are of the form

$$w(k) = \varphi(\|k\|), \quad k \in \mathbb{Z}^d, \tag{1}$$

where $\|\cdot\|$ is a (quasi-)norm on $\mathbb{R}^d$ and φ a univariate, monotonically increasing function φ with $\varphi(0) = 1$. The characterization results now states that the approximation numbers associated to the embedding $\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ are bounded for all $n \in \mathbb{N}$ from above and below as

$$1/\varphi(2/\varepsilon_n) \le a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le 1/\varphi(1/(4\varepsilon_n)). \tag{2}$$

Here, $\varepsilon_n = \varepsilon_n(\mathrm{id} : \ell_{\|\cdot\|}^d \to \ell_\infty^d)$ are *entropy numbers*. They give the smallest $\varepsilon > 0$ such that the $\ell_{\|\cdot\|}^d$-unit ball can be covered by n ℓ_∞^d-balls of radius ε. For the precise statement, see Theorem 4.3.

1.2 Preasymptotics for isotropic Sobolev spaces

The first concrete application of the abstract result (2) yields results for the isotropic Sobolev space $H^s(\mathbb{T}^d)$ with fractional smoothness $s > 0$. Isotropic Sobolev regularity is the natural notion of regularity for solutions of general elliptic PDEs, typically the solution will be contained in $H^1(\mathbb{T}^d)$ or $H^2(\mathbb{T}^d)$. The space $H^s(\mathbb{T}^d)$ can be defined as $H^s(\mathbb{T}^d) = H^{\mathbf{w}_{s,2}}(\mathbb{T}^d)$, where the weight sequence $\mathbf{w}_{s,2}$ is given by $w_{s,2}(k) = (1+\|k\|_2^2)^{s/2}$. Note that the norm $\|f|H^s(\mathbb{T}^d)\| = \sqrt{\sum_{k\in\mathbb{Z}^d} w_{s,2}(k)^2|c_k(f)|^2}$ is natural in the sense that if $s \in \mathbb{N}$, then this norm is equivalent up to a constant in s to the classical norm, which is defined in terms of the L_2-norms of the derivatives up to order s.

Concerning the approximation numbers $a_n(\mathrm{Id} : H^s(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$, the exact asymptotic decay in n has been known for decades. In 1967, J. W. Jerome [12] proved that

$$c_{s,d}\, n^{-s/d} \le a_n(\mathrm{Id} : H^s(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le C_{s,d}\, n^{-s/d}, \tag{3}$$

with constants $c_{s,d}$ and $C_{s,d}$ that were merely known to depend on the fractional smoothness s and the dimension d. For further references and historical remarks in this direction, see the monographs by Temlyakov [25] and Tikhomirov [26].

In order to obtain preasymptotics for (3) and clarify the d-dependence of the constants, we not only consider the weights $\mathbf{w}_{s,2}$ but the family of weights $\mathbf{w}_{s,p}$ given by

$$\begin{aligned} w_{s,p}(k) &= \quad (1+\textstyle\sum_{j=1}^d |k_j|^p)^{s/p} \quad && \text{if } 0 < p < \infty, \text{ and} \\ w_{s,p}(k) &= \max(1, |k_1|, \dots, |k_d|)^s && \text{if } p = \infty. \end{aligned} \tag{4}$$

For $0 < p < 1$, the weights $\mathbf{w}_{s,p}$ can be interpreted as imposing a *compressibility constraint* on the Fourier frequency vectors; the less $k \in \mathbb{Z}^d$ is aligned with one of the coordinate axes, the stronger the penalty through a large weight $w_{s,p}(k)$. The function spaces $H^{s,p}(\mathbb{T}^d) := H^{\mathbf{w}_{s,p}}(\mathbb{T}^d)$ coincide as sets with the classical isotropic Sobolev space. Applying the abstract result (2), we immediately obtain Theorem 1.1.

Theorem 1.1. *For $0 < p \le \infty$ and $s > 0$, we have*

$$a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \asymp_{s,p} \begin{cases} 1 & : 1 \le n \le d \\ \left[\frac{\log(1+d/\log n)}{\log n}\right]^{s/p} & : d \le n \le 2^d \\ d^{-s/p} n^{-s/d} & : n \ge 2^d. \end{cases}$$

It is clearly visible how a smaller compressibility parameter p makes the approximation problem easier by amplifying the preasymptotic logarithmic decay in n. The equivalent constants in Theorem 1.1 depend only on s and p and can be completely controlled. For $n \to \infty$, the constants in the lower and upper bound even converge,

$$\lim_{n\to\infty} n^{s/d} a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = (\mathrm{vol}(B_p^d))^{s/d} \asymp d^{-s/p},$$

see Corollary 5.2. A consequence of Theorem 1.1 is that we face the *curse of dimensionality* in the strict sense of information-based complexity if and only if $p = \infty$. Otherwise, the approximation problem is *weakly tractable*, despite the slow asymptotic decay $n^{-s/d}$. For details, see Section 7.

1.3 Spaces of Gevrey type and a connection to hyperbolic cross spaces

The classes of smooth functions that are nowadays called *Gevrey classes* were already introduced in 1918 by M. Gevrey [6], they occurred in a natural way in his research on partial differential equations. Since then they have played an important role in numerous applications, in particular in connection with Cauchy problems. The recent paper [14] introduces the periodic *spaces of Gevrey type* $G^{\alpha,\beta,p}(\mathbb{T}^d)$, $0 < \alpha, \beta, p < \infty$, which consist of all $f \in C^\infty(\mathbb{T}^d)$ such that the norm

$$\|f|G^{\alpha,\beta,p}(\mathbb{T}^d)\| := \left(\sum_{k\in\mathbb{Z}^d} \exp(2\beta\,\|k\|_p^\alpha)|c_k(f)|^2 \right)^{1/2}$$

is finite. Here $c_k(f)$ denotes the Fourier coefficient with respect to the frequency vector $k = (k_1, ..., k_d) \in \mathbb{Z}^d$, defined in Section 2 below. For $0 < \alpha < 1$, the spaces $G^{\alpha,\beta,p}(\mathbb{T}^d)$ coincide with the classical Gevrey classes and contain non-analytic functions, while for $\alpha \geq 1$ all functions in $G^{\alpha,\beta,p}(\mathbb{T}^d)$ are analytic. Some more background on Gevrey classes and references can be found in Section 6.

In Theorem 6.1 we prove lower and upper bounds for the approximation numbers a_n of the embedding $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ for all $n \in \mathbb{N}$ and arbitrary parameter values $\alpha, \beta > 0$, $0 < p \leq \infty$. Due to our proof technique we can determine the rate of convergence only up to a constant. However, for $0 < p \leq \infty$ and $\alpha < \min\{1, p\}$, we at least obtain an indication for the correct asymptotic behavior by the limit statement

$$\lim_{n\to\infty} a_n \cdot \exp(\lambda\beta n^{\alpha/d}) = 1,$$

where $\lambda := \mathrm{vol}(B_p^d)^{-\alpha/d}$, see Theorem 6.2.

What concerns preasymptotics the bounds turn out to be rather surprising in the particular situation $\alpha = p$. For $1 \leq n \leq 2^d$, we obtain the two-sided estimate

$$n^{-\frac{c_1(p)\beta}{\log(1+d/\log(n))}} \leq a_n(\mathrm{Id} : G^{p,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq n^{-\frac{c_2(p)\beta}{\log(1+d/\log(n))}} \leq n^{-\frac{c_2(p)\beta}{\log(1+d)}}. \quad (5)$$

This estimate is almost identical to the preasymptotic estimate which has been obtained in the recent paper [16] (see also (7) below) for approximation numbers of the embeddings

$$\mathrm{Id} : H^r_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d),$$

where $H^r_{\mathrm{mix}}(\mathbb{T}^d)$ is the Sobolev space of dominating mixed smoothness equipped with the norm

$$\|f|H^r_{\mathrm{mix}}(\mathbb{T}^d)\| := \left[\sum_{k\in\mathbb{Z}^d} |c_k(f)|^2 \prod_{j=1}^d (1 + |k_j|^2)^r \right]^{1/2}.$$

It is rather counterintuitive that the approximation numbers behave almost identically in the preasymptotic range. After all, the spaces of Gevrey type $G^{p,\beta,p}(\mathbb{T}^d)$ contain substantially smoother functions than the space $H^r_{\mathrm{mix}}(\mathbb{T}^d)$. We discuss this in more detail in Section 6.1 and give at least partial explanations for this odd phenomenon.

1.4 Preasymptotics for embeddings into $\mathbf{H}^{\mathbf{s}}$

Yserentant [29] proved that eigenfunctions of the positive spectrum of the electronic Schrödinger operator possess a dominating mixed regularity. To solve the electronic Schrödinger equation numerically, Galerkin methods combined with sparse grid techniques [7, 9] are widely used. The discussion in Subsection 1.5 below shows that one is particularly interested in measuring the error in the *energy space* H^1. From results in [8] it follows that

$$c_d n^{-(r-1)} \leq a_n(H^r_{\mathrm{mix}}(\mathbb{T}^d) \to H^1(\mathbb{T}^d)) \leq C_d n^{-(r-1)}, \tag{6}$$

with constants c_d, C_d depending on the dimension d. In [8] it has been claimed that $C_d = d^2 0.97515^d$. This result suggests that the truncation problem even gets easier with a growing number of electrons. However, as [4] shows, the constant C_d can only be chosen as above for exponentially large $n > (1+\gamma)^d$. This raises the question how the approximation numbers in (6) behave preasymptotically.

Unfortunately, the abstract result (2) cannot be applied to obtain preasymptotics for (6), since the space $H^r_{\mathrm{mix}}(\mathbb{T}^d)$ cannot be written as a space $H^{\mathbf{w}}(\mathbb{T}^d)$ with a weight sequence $\mathbf{w}$ of the form (1). However, the observations described in the previous subsection and results in [16] give an indication for the preasymptotic behavior. The results in [16, Thm. 4.9, 4.10, 4.17] provide the two-sided estimate

$$2^{-r}\left(\frac{1}{2n}\right)^{\frac{r}{2+\log(1/2+d/\log(n))}} \leq a_n(\mathrm{Id}: H^r_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq \left(\frac{e^2}{n}\right)^{\frac{r}{4+2\log_2 d}} \tag{7}$$

in the preasymptotic range $1 \leq n \leq 4^d$. With the coincidence

$$a_n(\mathrm{Id}: H^r_{\mathrm{mix}}(\mathbb{T}^d) \to H^s_{\mathrm{mix}}(\mathbb{T}^d)) = a_n(\mathrm{Id}: H^{r-s}_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)),$$

provided $r > s > 0$, we obtain from (7) by embedding

$$a_n(\mathrm{Id}: H^r_{\mathrm{mix}}(\mathbb{T}^d) \to H^s(\mathbb{T}^d)) \leq \left(\frac{e^2}{n}\right)^{\frac{r-s}{4+2\log_2 d}}. \tag{8}$$

The connection between spaces of Gevrey type and spaces of dominating mixed smoothness sketched in Subsection 1.3 (see also Subsection 6.1 below), might be useful to refine the result (8). Indeed, for spaces of Gevrey type we obtain the following result as a consequence of our abstract technique. For $1 \leq n \leq 2^d$, we have

$$c_1 \Lambda(n,d)^{s/2} n^{-\frac{c_1 \beta}{\gamma(n,d)}} \leq a_n(\mathrm{Id}: G^{2,r/2,2}(\mathbb{T}^d) \to H^s(\mathbb{T}^d)) \leq c_2 \Lambda(n,d)^{s/2} n^{-\frac{c_2 r}{\gamma(n,d)}}, \tag{9}$$

where $\Lambda(n,d) = \frac{\log(n)}{\gamma(n,d)}$ and $\gamma(n,d) = \log(1 + d/\log(n))$; compare with (5).

1.5 Approximation numbers and Galerkin methods

To conclude this introduction, let us outline the connection between approximation numbers and reliable a-priori error estimates for Galerkin methods. Consider a general elliptic variational problem in $H^s = H^s(\mathbb{T}^d)$, which is given by a bilinear symmetric form $a : H^s \times H^s \to \mathbb{R}$ and a right-hand side $f \in H^{-s}$. The bilinear symmetric form is assumed to satisfy, for any $u, v \in H^s$,

$$a(u,v) \leq \mu_1 \| u \,|H^s\| \| v \,|H^s\| \text{ and } a(u,u) \geq \mu_2 \| u \,|H^s\|^2.$$

Under this assumption, $a(\cdot,\cdot)$ generates the so called *energy norm equivalent* to the norm of H^s. The problem now is to find an element $u \in H^s$ such that

$$a(u,v) \;=\; (f,v) \text{ for all } v \in H^s. \tag{10}$$

In order to get an approximate numerical solution Galerkin methods solve the same problem on a finite dimensional subspace V_h in H^s,

$$a(u_h,v) \;=\; (f,v) \text{ for all } v \in V_h. \tag{11}$$

By the Lax-Milgram theorem [17], the problems (10) and (11) have unique solutions u^* and u_h^*, respectively, which by Céa's lemma [1], satisfy the inequality

$$\|u^* - u_h^* \,|H^s\| \;\leq (\mu_1/\mu_2) \inf_{v \in V_h} \|u^* - v \,|H^s\|. \tag{12}$$

The naturally arising question is how to choose the optimal n-dimensional subspace V_h and linear finite element approximation algorithms such that the right-hand side in (12) becomes as smalls as possible. Under the assumption that the solution u^* is contained in the unit ball of some smoothness space $U \subset H^s$, the minimal right-hand side in (12) is bounded from above by the approximation number $a_n(\mathrm{Id} : U \to H^s)$. Summarizing,

$$\|u^* - u_h^* \,|H^s\| \;\leq (\mu_1/\mu_2) a_n(\mathrm{Id} : U \to H^s)$$

gives a worst-case a-priori error estimate for the optimal n-dimensional subspace V_h.

2 Preliminaries

Notation As usual, the set $\mathbb{N}$ denotes the natural numbers, $\mathbb{Z}$ the integers and $\mathbb{R}$ the real numbers. By $\mathbb{T}$ we denote the torus represented by the interval $[0, 2\pi]$ where opposite points are identified. A function $f : \mathbb{T}^d \to \mathbb{C}$ is 2π-periodic in every component. If $f \in L_2(\mathbb{T}^d)$ then the Fourier coefficient $c_k(f)$ with respect to the frequency vector $k = (k_1, ..., k_d) \in \mathbb{Z}^d$ is given by

$$c_k(f) = (2\pi)^{-d/2} \int_{\mathbb{T}^d} f(x) e^{ik\cdot x} \, dx \, .$$

For a real number a we put $a_+ := \max\{a, 0\}$. The symbol d is always reserved for the dimension in $\mathbb{Z}^d$, $\mathbb{R}^d$, $\mathbb{N}^d$, and $\mathbb{T}^d$. For $0 < p \le \infty$ and $x \in \mathbb{R}$ we denote $\|x\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$ with the usual modification in the case $p = \infty$. We write ℓ_p^d for $\mathbb{R}^d$ equipped with the norm $\|\cdot\|_p$. By B_p^d we denote the closed unit ball of ℓ_p^d. When we write log, we always mean the logarithm to the base of 2. If X and Y are two Banach spaces, the norm of an element x in X will be denoted by $\|x|X\|$ and the norm of an operator $A : X \to Y$ is denoted by $\|A : X \to Y\|$. The symbol $X \hookrightarrow Y$ indicates that the embedding operator is continuous.

Approximation numbers Let X, Y be two Banach spaces. An operator of rank n is a mapping

$$A : X \to Y, \quad f \mapsto \sum_{j=1}^{n} \lambda_j(f) g_j,$$

where the $\lambda_1, \ldots, \lambda_n$ are linear functionals and $g_1, \ldots, g_n \in Y$. The n-th approximation number $a_n(T : X \to Y)$ of an operator $T : X \to Y$ is the worst-case error that occurs when approximating elements from the image of the unit ball of X in Y by the best possible operator of rank less than n. We equally could say that we approximate the operator $T : X \to Y$ by operators of finite rank in the operator norm, which motivates the notation. Formally, the n-th approximation numbers is defined by

$$\begin{aligned} a_n(T : X \to Y) &:= \inf_{\operatorname{rank} A < n} \sup_{\|f|X\| \le 1} \| Tf - Af|Y\| \\ &= \inf_{\operatorname{rank} A < n} \|T - A : X \to Y\|. \end{aligned} \tag{13}$$

Covering and entropy numbers Let $A \subset \mathbb{R}^d$. The *covering number* $N_\varepsilon(A)$ is the minimal natural number n for which there are points $x_1, \ldots, x_n$ in $\mathbb{R}^d$ such that

$$A \subseteq \bigcup_{i=1}^{n} (x_i + \varepsilon B_\infty^d).$$

Inverse to the covering numbers $N_\varepsilon(A)$ are the (non-dyadic) *entropy numbers*

$$\varepsilon_n(A, \ell_\infty^d) := \inf\{\varepsilon > 0 : N_\varepsilon(A) \le n\}.$$

If $A = B_{\|\cdot\|}^d = \{x \in \mathbb{R}^d : \|x\| \le 1\}$ is a unit ball, we also use the notation

$$\varepsilon_n(\mathrm{id} : \ell_{\|\cdot\|}^d \to \ell_\infty^d) := \varepsilon_n(B_{\|\cdot\|}^d, \ell_\infty^d).$$

In the applications which we have in mind $\|\cdot\|$ will be a classical (quasi-)norm $\|\cdot\| = \|\cdot\|_p$ for $0 < p \le \infty$. In this case, the behavior in n and d of the entropy numbers $\varepsilon_n(\mathrm{id} : \ell_p^d \to \ell_\infty^d)$ is completely understood [5, 13, 20, 22, 27]. For the reader's convenience, we restate the results.

Proposition 2.1. *For all $n \in \mathbb{N}$, we have*

$$n^{-1/d} \le \varepsilon_n(id : \ell_\infty^d \to \ell_\infty^d) \le 2n^{-1/d}.$$

However, $\varepsilon_n(id : \ell_\infty^d \to \ell_\infty^d) = 1$ as long as $n < 2^d$.

Proposition 2.2. *Let $0 < p < \infty$. Then,*

$$\varepsilon_n(id : \ell_p^d \to \ell_\infty^d) \asymp \begin{cases} 1 & : \ 1 \le n \le d, \\ \left(\frac{\log(1+d/\log n)}{\log n}\right)^{1/p} & : \ d \le n \le 2^d, \\ d^{-1/p} n^{-1/d} & : \ n \ge 2^d, \end{cases}$$

with constants independent of n and d.

The equivalence constants in Proposition 2.2 are not further specified in the literature. It is possible to calculate explicit, but rather lengthy expressions. We refrain from going more into detail at this point. In Section 3, we will comment on the behavior of the constants for $n \to \infty$.

Remark 2.3. The closely related entropy numbers $\varepsilon_n(\mathbb{S}_p^{d-1}, \ell_\infty^d)$, where $\mathbb{S}_p^{d-1} = \{x \in \mathbb{R}^d : \|x\|_p = 1\}$, have been understood only lately [10, 18]. It is no surprise that these behave identically to the entropy numbers $\varepsilon_n(id : \ell_p^d \to \ell_\infty^d)$, except that asymptotically they decay as $n^{-1/(d-1)}$. To prove the bounds on $\varepsilon_n(\mathbb{S}_p^{d-1}, \ell_\infty^d)$ one largely mimics the well-known proof for $\varepsilon_n(id : \ell_p^d \to \ell_\infty^d)$. Surprisingly, there is one case where this strategy fails. For $0 < p < 1$ and $n \ge 2^d$ the familiar volume arguments become inaccurate and it needs different techniques to obtain matching bounds, see [10].

Notions of tractability In the course of this paper we want to classify how the dimension d affects the hardness of the approximation problem $\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ depending on the weight sequence $\mathbf{w}$. The field of information-based complexity provides notions of tractability [19], which rate the difficulty of the approximation problem in terms of how its information complexity

$$n(\varepsilon, d) := \inf\{n \in \mathbb{N} : a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le \varepsilon\}$$

grows in $1/\varepsilon$ and d. Let us first note that for all weight sequences $\mathbf{w}$ considered in this paper the *normalized error criterion* is fulfilled, that is,

$$a_1(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = \|\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)\| = 1.$$

Now, the approximation problem is said to be *polynomially tractable* if $n(\varepsilon, d)$ is bounded polynomially in ε^{-1} and d, i.e., there exist numbers $C, r, q > 0$ such that

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

The approximation problem is called *quasi-polynomially tractable* if there exist two constants $C, t > 0$ such that

$$n(\varepsilon, d) \le C \exp(t(1 + \ln(1/\varepsilon))(1 + \ln d)) .$$

It is called *weakly tractable* if

$$\lim_{1/\varepsilon + d \to \infty} \frac{\log n(\varepsilon, d)}{1/\varepsilon + d} = 0 , \tag{14}$$

i.e., the information complexity $n(\varepsilon, d)$ neither depends exponentially on $1/\varepsilon$ nor on d. We say that the approximation problem is *intractable*, if (14) does not hold. If for some fixed $0 < \varepsilon < 1$ the information complexity $n(\varepsilon, d)$ is an exponential function in d then we say that the problem suffers from *the curse of dimensionality.* To make it precise, we face the curse if there exist positive numbers c, ε_0, γ such that

$$n(\varepsilon, d) \ge c(1 + \gamma)^d , \quad \text{for all } 0 < \varepsilon \le \varepsilon_0 \text{ and infinitely many } d \in \mathbb{N} .$$

3 Counting via entropy

The *grid number* $G(A)$ of a set $A \subseteq \mathbb{R}^d$ is the number of points in A that lie on the grid $\mathbb{Z}^d$. Formally,

$$G(A) = \sharp(A \cap \mathbb{Z}^d).$$

The grid numbers $G(rB^d_{\|\cdot\|})$, $r \in \mathbb{R}$ play a central role in the study of approximation numbers $a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ when the weights $\mathbf{w}$ are induced by some (quasi-)norm $\|\cdot\|$, see Section 4 below. In this section, we show that the combinatorics for grid numbers can be reduced to covering arguments, at least if the studied set is solid. We call a set $A \subseteq \mathbb{R}^d$ *solid* if for all $x \in A$ every vector $y \in \mathbb{R}^d$ which component-wise fulfills $|y_i| \le |x_i|$ is contained in A. For instance, the unit ball B^d_p is solid for any $0 < p \le \infty$.

Lemma 3.1. *For $A \subseteq R^d$ a solid set, we have*

$$N_1(A) \le G(A) \le N_\rho(A)$$

for any $\rho < 1/2$.

Proof. For $x \in A$, we define $\lfloor x \rfloor$ component-wise by $\lfloor x \rfloor_j = \operatorname{sign} x_i \ \lfloor |x_j| \rfloor$. Clearly, $\| \lfloor x \rfloor - x \|_\infty < 1$ for any $x \in A$. Since the set A is solid, the intersection $A \cap \mathbb{Z}^d$ forms a 1-net of A in ℓ^d_∞. Consequently, we have $N_1(A) \le G(A)$. The upper bound is a direct consequence of the fact that it needs at least $G(A)$ many balls of radius $\rho < 1/2$ to cover $A \cap \mathbb{Z}^d$. $\square$

A function $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ is called a p-norm for some $0 < p \le 1$ if $\|\cdot\|$ fulfills the norm axioms of absolute homogeneity and point separation and furthermore the p-triangle inequality, which is

$$\|x + y\|^p \le \|x\|^p + \|y\|^p$$

for $x, y \in \mathbb{R}^d$. The typical example for a p-norm with $0 < p < 1$ is $\|\cdot\| = \|\cdot\|_p$. If $A \subset \mathbb{R}^d$ is the unit ball of a p-norm $\|\cdot\|$, there is another relation between covering and grid numbers. In this relation the quantity

$$\lambda_{\|\cdot\|}(d) := \|\sum_{i=1}^{d} e_i\|$$

appears, where $e_1, \ldots, e_d$ denote the canonical basis vectors in $\mathbb{R}^d$. Note that if $\|\cdot\| = \|\cdot\|_p$ for $0 < p \leq \infty$ we have $\lambda_{\|\cdot\|}(d) = d^{1/p}$.

Lemma 3.2. *Let $\|\cdot\|$ be a p-norm in $\mathbb{R}^d$ for some $0 < p \leq 1$. For $r > \lambda_{\|\cdot\|}(d)/2$, put*

$$l(r, p, d) := \left(r^p - \lambda_{\|\cdot\|}(d)^p/2^p\right)^{1/p},$$
$$L(r, p, d) := \left(r^p + \lambda_{\|\cdot\|}(d)^p/2^p\right)^{1/p}.$$

With $B^d_{\|\cdot\|}$ denoting the unit ball, we have the relation

$$N_{1/2}(l(r, p, d)B^d_{\|\cdot\|}) \leq G(rB^d_{\|\cdot\|}) \leq N_{1/2}(L(r, p, d)B^d_{\|\cdot\|}).$$

Proof. Let $Q_k = k + [-1/2, 1/2]^d$. The p-triangle inequality for $\|\cdot\|$ yields

$$l(r, p, d)B^d_{\|\cdot\|} \subseteq \bigcup_{\substack{k \in \mathbb{Z}^d, \\ \|k\| \leq r}} Q_k \subseteq L(r, p, d)B^d_{\|\cdot\|}. \tag{15}$$

The left-hand side inclusion shows that the set $rB^d_{\|\cdot\|} \cap \mathbb{Z}^d$ is a 1/2-net of $l(r, p, d)B^d_{\|\cdot\|}$ in ℓ^d_∞. This shows the left-hand side inequality of the statement. The second inequality follows from the right-hand side inclusion by a simple volume argument. □

Lemma 3.2 yields the following bounds for entropy numbers. Note that the upper bound is a refinement for large n of the usual upper bound found in the literature, compare also with Propositions 2.1, 2.2.

Lemma 3.3. *Let $\|\cdot\|$ be a p-norm for some $0 < p \leq 1$. For $n > (d^{1/p}/2)^d \operatorname{vol}(B^d_{\|\cdot\|})$, we have*

$$\frac{1}{2}(n/\operatorname{vol}(B^d_{\|\cdot\|}))^{1/d} \leq \varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty) \leq \frac{1}{2}\left((n/\operatorname{vol}(B^d_{\|\cdot\|}))^{p/d} - 2^{1-p}d^{1/p})\right)^{-1/p}.$$

Proof. The lower bound is the standard lower bound, which follows from simple volume arguments and in fact holds true for all $n \in \mathbb{N}$. To see the upper bound, choose $r = \left((n/\operatorname{vol}(B^d_{\|\cdot\|}))^{p/d} - d^{p/p}/2^p)\right)^{1/p}$. Then it follows from the right-hand side inclusion of (15) that $G(rB^d_p) \leq n$ and further from the left-hand side inclusion of (15) that $\varepsilon_n(\mathrm{id} : \ell^d_p \to \ell^d_\infty) \leq 1/(2l(r, p, d))$, where $l(r, p, d)$ is defined in Lemma 3.2. It remains to plug in the formula for r. □

Let us briefly come back to the discussion on the equivalence constants in Proposition 2.1 and Proposition 2.2. An interesting question is whether the equivalence constants in the lower and upper bounds necessarily have to be different or whether this is just an artifact of the used proof techniques. Lemma 3.3 allows a partial answer. In the limit $n \to \infty$ we have

$$\lim_{n\to\infty} n^{1/d} \varepsilon_n(id : \ell^d_{\|\cdot\|} \to \ell^d_\infty) = 1/2 \cdot \mathrm{vol}(B^d_{\|\cdot\|})^{1/d}. \tag{16}$$

If $\|\cdot\| = \|\cdot\|_p$ for $0 < p \le \infty$, then $\mathrm{vol}(B^d_{\|\cdot\|})^{1/d} = \mathrm{vol}(B^d_p)^{1/d} \asymp d^{-1/p}$, see [28].

4 Characterization of approximation numbers

In this section, we prove a number of characterization results for approximation numbers $a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ when the weight sequence $\mathbf{w}$ is derived from some (quasi-)norm in $\mathbb{R}^d$. To begin with, let us recapitulate some well-known facts about approximation numbers of weighted spaces. Let $\mathbf{w} = (w(k))_{k\in\mathbb{Z}^d}$ be an arbitrary weight sequence such that $1/\mathbf{w} := (1/w(k))_{k\in\mathbb{Z}^d} \in \ell_\infty(\mathbb{Z}^d)$. It is well-known that the approximation numbers are given by the *non-increasing rearrangement* $(\sigma_n)_{n\in\mathbb{N}}$ of the inverse weight sequence $1/\mathbf{w}$, that is,

$$a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = \sigma_n \tag{17}$$

for all $n \in \mathbb{N}$. We briefly sketch the proof of this fact, details and further references can be found in [15, Section 2.2]. Consider the isometries

$$A^{\mathbf{w}} : H^{\mathbf{w}}(\mathbb{T}^d) \to \ell_2(\mathbb{Z}^d),\ f \mapsto \big(w(k)c_k(f)\big)_{k\in\mathbb{Z}^d} \tag{18}$$

and

$$F : \ell_2(\mathbb{Z}^d) \to L_2(\mathbb{T}^d),\ (\xi_k)_{k\in\mathbb{Z}^d} \mapsto (2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \xi_k e^{ikx}, \tag{19}$$

as well as the diagonal operator

$$D^{\mathbf{w}} : \ell_2(\mathbb{Z}^d) \to \ell_2(\mathbb{Z}^d),\ (\xi_k)_{k\in\mathbb{Z}^d} \mapsto (\xi_k/w(k))_{k\in\mathbb{Z}^d}. \tag{20}$$

Obviously, we have $\mathrm{Id} = F \circ D^{\mathbf{w}} \circ A^{\mathbf{w}}$, which is illustrated by the commutative diagram below.
It is known that the approximation numbers of the diagonal operator are given by $(\sigma_n)_{n\in\mathbb{N}}$, and from $\|A^{\mathbf{w}}\| = \|F\| = 1$ we conclude

$$a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = a_n(D^{\mathbf{w}} : \ell_2(\mathbb{Z}^d) \to \ell_2(\mathbb{Z}^d)) = \sigma_n.$$

We come to our first characterization result. For weight sequences $\mathbf{w}$ given by a (quasi-)norm $\|\cdot\|$ on $\mathbb{R}^d$, we show that the non-increasing rearrangement $(\sigma_n)_{n\in\mathbb{N}}$ is in fact equivalent up to constants to the entropy numbers $\varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty)$.

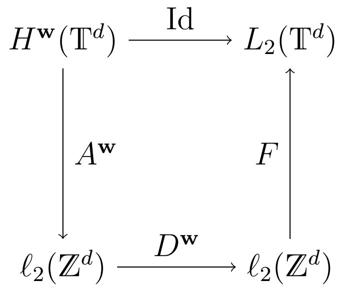

Figure 1: Commutative diagram for the embedding $\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$.

Theorem 4.1. *Let $\|\cdot\|$ be some (quasi-)norm on $\mathbb{R}^d$ such that $\min_{i=1,\ldots,d} \|e_i\| = 1$, where $e_1, \ldots, e_d$ denotes the canonical basis in $\mathbb{R}^d$. Consider the weight sequence $\mathbf{w} = (w(k))_{k\in\mathbb{Z}^d}$ given by $w(k) := \max\{1, \|k\|\}$. For every $n \in \mathbb{N}$, we have*

$$1/2\, \varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty) \le a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le 4\, \varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty).$$

Proof. Let $(\sigma_n)_{n\in\mathbb{N}}$ denote the non-increasing rearrangement of $(1/w(k))_{k\in\mathbb{Z}^d}$. By (17) we know that $a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = \sigma_n$. Since $G(mB^d_{\|\cdot\|}) = \sharp\{k \in \mathbb{Z}^d : \|k\| \le m\} = \sharp\{k \in \mathbb{Z}^d : w(k) \le m\}$ and $w(me_{i^*}) = m$, where $i^* = \arg\min_{i=1,\ldots,d} \|e_i\|$, we have $\sigma_{G(mB^d_{\|\cdot\|})} = 1/m$.

Let us first prove the upper bound. For brevity, we write $\varepsilon_n = \varepsilon_n(id : \ell^d_{\|\cdot\|} \to \ell^d_\infty)$ in the following. For given $n \in \mathbb{N}$, let $\varepsilon > \varepsilon_n$ and put $m := \lfloor 1/((2+\delta)\varepsilon)\rfloor$ for some $\delta > 0$. By virtue of Lemma 3.1, we obtain

$$n \ge N_\varepsilon(B^d_{\|\cdot\|}) = N_{m\varepsilon}(mB^d_{\|\cdot\|}) \ge N_{1/(2+\delta)}(mB^d_{\|\cdot\|}) \ge G(mB^d_{\|\cdot\|}).$$

The monotonicity of approximation numbers yields

$$\sigma_n \le \sigma_{G(mB^d_{\|\cdot\|})} = 1/m \le 2(2+\delta)\varepsilon. \tag{21}$$

Since δ can be chosen arbitrarily close to 0 and ε arbitrarily close to ε_n, we reach at $\sigma_n \le 4\varepsilon_n$.

To prove the lower bound, assume $\varepsilon < \varepsilon_n$ for some $n \in \mathbb{N}$ and put $m = \lceil 1/\varepsilon\rceil$. We have $n \le N_\varepsilon(B^d_{\|\cdot\|}) \le N_1(mB^d_{\|\cdot\|}) \le G(mB^d_{\|\cdot\|})$, where the last estimate is due to Lemma 3.1. Thus, $\sigma_n \ge \sigma_{G(mB^d_{\|\cdot\|})} = 1/m \ge 1/2\varepsilon$. Again, as ε may be chosen arbitrarily close to ε_n, we have $\sigma_n \ge 1/2\varepsilon_n$. □

Remark 4.2. The assumption $\min_{i=1,\ldots,d} \|e_i\| = 1$ in Theorem 4.1 has only been made to keep the formulation of the statement and the proof as simple as possible. In particular, the normalized error criterion is always fulfilled. If $\min_{i=1,\ldots,d} \|e_i\| = c \ne 1$, then the statement still holds true, provided we define $w(k) := \max\{c, \|k\|\}$. Otherwise, i.e. when keeping the definition $w(k) := \max\{1, \|k\|\}$, the statement holds true for sufficiently large $n > n_0(c,d)$, where $n_0(c,d)$ can depend on c and the dimension d.

The statement of Theorem 4.1 can be easily generalized.

Theorem 4.3. *Let $\|\cdot\|$ be some (quasi-)norm on $\mathbb{R}^d$ as in Theorem 4.1 and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing function satisfying $\varphi(0) = 1$. Consider the weight sequence $\mathbf{w} = \varphi(\|\cdot\|)$ given by $w(k) := \varphi(\|k\|)$. Writing $\varepsilon_n = \varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty)$, we have, for all $n \in \mathbb{N} \setminus \{1\}$, the estimate*

$$\frac{1}{\varphi(2/\varepsilon_n)} \le a_n(\mathrm{Id} : H^{\varphi(\|\cdot\|)}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le \frac{1}{\varphi(1/(4\varepsilon_n))}.$$

Proof. Let $\widetilde{\mathbf{w}}$ be the weight sequence given by $\widetilde{w}(k) = \max\{1, \|k\|\}$ for $k \in \mathbb{Z}^d$. Further, let $(\sigma_n)_{n\in\mathbb{N}}$ be the non-increasing rearrangement of $1/\widetilde{\mathbf{w}}$. Note that $\varphi(\|k\|) = \varphi(\widetilde{w}(k))$ for $k \neq 0$ since $\min_{i=1,\dots,d} \|e_i\| = 1$. Put $\gamma_1 = 1$ and

$$\gamma_n = \frac{1}{\varphi(1/\sigma_n)} \tag{22}$$

for natural $n > 1$. Since φ is monotonically increasing the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is non-increasing and thus the non-increasing rearrangement of $(1/w(k))_{k\in\mathbb{Z}^d}$. It remains to combine (22) with the finding of Theorem 4.1. □

The constants in the lower and upper bound of Theorem 4.1 do not match. In general and for arbitrary $n \in \mathbb{N}$, we indeed cannot expect to see constants any better in the bounds. However, we have

$$\lim_{n\to\infty} \frac{a_n(\mathrm{Id} : H^{\max\{1,\|\cdot\|\}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))}{\varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty)} = 2,$$

which is a consequence of the following limit result.

Theorem 4.4. *Let $\|\cdot\|$ be a p-norm in $\mathbb{R}^d$ for $0 < p \le 1$. Further, let φ be given by $\varphi(t) = \exp(\beta g(t))$ with $\beta > 0$ and monotonically increasing, differentiable g satisfying $g(0) = 0$ and $\lim_{t\to\infty} g'(t)t^{1-p} = 0$. Recall the weight sequence $\mathbf{w} = \varphi(\|\cdot\|)$ defined in Theorem 4.3. Using the shorthands $a_n = a_n(\mathrm{Id} : H^{\varphi(\|\cdot\|)}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ and $\varepsilon_n = \varepsilon_n(\mathrm{id} : \ell^d_{\|\cdot\|} \to \ell^d_\infty)$, it holds true that*

$$\lim_{n\to\infty} a_n \varphi(1/(2\varepsilon_n)) = 1.$$

Proof. Let $(\sigma_n)_{n\in\mathbb{N}}$ be the non-increasing rearrangement of $(1/\max\{1, \|k\|\})_{k\in\mathbb{Z}^d}$ and $(\gamma_n)_{n\in\mathbb{N}}$ be the non-increasing rearrangement of $(1/\varphi(\|k\|))_{k\in\mathbb{Z}^d}$. Let us first refine the upper bound (21). Consider $n \in \mathbb{N}$ sufficiently large such that $\varepsilon_n < 1/2$. With $\delta > 0$ arbitrary, $\varepsilon > 0$ such that $\varepsilon_n < \varepsilon < 1/(2+\delta)$, and $m := \lfloor 1/((2+\delta)\varepsilon) \rfloor$, we obtain

$$\sigma_n \le \sigma_{G(mB_p^d)} = 1/m \le \frac{(2+\delta)\varepsilon}{1-(2+\delta)\varepsilon}.$$

Since we may choose δ arbitrarily close to 0 and ε arbitrarily close to ε_n, we obtain $\sigma_n \le 1/h_1(1/(2\varepsilon_n))$, where $h_1(t) = t - 1$.

To obtain a refinement of the lower bound, choose for $\varepsilon < \varepsilon_n$ the natural number $m = \lceil 1/(2\varepsilon) \rceil$. Using Lemma 3.2 we obtain $n \leq N_\varepsilon(B^d_{\|\cdot\|}) \leq N_{1/2}(mB^d_{\|\cdot\|}) \leq G(\widetilde{m}B^d_{\|\cdot\|})$, where

$$\widetilde{m} := \left(m^p + \lambda_{\|\cdot\|}(d)^p/2^p\right)^{1/p}.$$

Hence,

$$\sigma_n \geq 1/\widetilde{m} \geq 1/h_2(1/(2\varepsilon)),$$

where $h_2(t) = ((t+1)^p + (\lambda_{\|\cdot\|}(d)^p/2^p)^{1/p}$. Since we may choose ε arbitrarily close to ε_n, we obtain $1/h_2(1/(2\varepsilon_n)) \leq \sigma_n$.

Combining the refined estimates with equation (22), we obtain from multiplying by $\varphi(1/(2\varepsilon_n))$ the two-sided estimate

$$\frac{\varphi(1/(2\varepsilon_n))}{\varphi(h_2(1/(2\varepsilon_n)))} \leq \gamma_n \varphi(1/(2\varepsilon_n)) \leq \frac{\varphi(1/(2\varepsilon_n))}{\varphi(h_1(1/(2\varepsilon_n)))}. \tag{23}$$

We have $\ln\left(\frac{\varphi(x)}{\varphi(h_1(x))}\right) = \beta(g(x) - g(x-1)) = \beta g'(\xi_x)$ for some $x - 1 \leq \xi_x \leq x$. From the assumptions on g, it obviously follows that $\lim_{x\to\infty} g'(\xi_x) = 0$. Hence, we have $\lim_{n\to\infty} \gamma_n \varphi(1/(2\varepsilon_n)) \leq 1$. For the estimate from below we have to show that $g(x) - g(h_2(x)) \to 0$ for $x \to \infty$. By the mean value theorem, it follows that

$$\begin{aligned}|x - h_2(x)| &\leq 1 + |x + 1 - h_2(x)| \\ &\leq 1 + \frac{\lambda_{\|\cdot\|}(d)^p}{p2^p}[(x+1) + \mu]^{1/p-1}\end{aligned}$$

for some $\mu \in [0, \lambda_{\|\cdot\|}(d)^p/2^p]$. Due to $p \leq 1$ we further may estimate $|x - h_2(x)| \leq C_d x^{1-p}$ for some $C_d > 0$. Combined with another application of the mean value theorem, this yields

$$|g(x) - g(h_2(x))| \leq |g'(\xi)||x - h_2(x)| \leq C_d|g'(\xi)|x^{1-p} \leq C_d|g'(\xi)|\xi^{1-p},$$

where $\xi \in [x, h_2(x)]$. Since we have assumed $\lim_{x\to\infty} g'(x)x^{1-p}$ it follows that $1 \leq \lim_{n\to\infty} \gamma_n \varphi(1/(2\varepsilon_n))$. □

5 Isotropic Sobolev spaces

In this section, we give further details and additional remarks to the results presented in Subsection 1.2 of the introduction.

Proof of Theorem 1.1. For $0 < p < \infty$, Theorem 4.3 with $\|\cdot\| = \|\cdot\|_p$ and $\varphi(t) = (1 + t^p)^{s/p}$ yields

$$2^{-(1+s/p)}\varepsilon_n(\mathrm{id} : \ell_p^d \to \ell_\infty^d)^s \leq a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq 4^s \varepsilon_n(\mathrm{id} : \ell_p^d \to \ell_\infty^d)^s.$$

It remains to apply Proposition 2.2. In case $p = \infty$, the argumentation is analogous with $\varphi(t) = \max\{1, t\}^s$. □

In the special case $p = \infty$, let us restate Theorem 1.1 with explicit expressions for the equivalence constants.

Theorem 5.1. *For $p = \infty$ and $s > 0$, we have*

$$2^{-(1+1/p)s} n^{-s/d} \leq a_n(\mathrm{Id} : H^{s,\infty}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq 8^s n^{-s/d}.$$

for all $n \in \mathbb{N}$. However, $a_n(\mathrm{Id} : H^{s,\infty}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = 1$ as long as $n \leq 2^d$.

Proof. Combine Theorem 4.3 with $\varphi(t) = \max\{1, t\}^s$ and Proposition 2.1. □

Theorem 4.4 applied to the approximation numbers $a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ yields the following corollary.

Corollary 5.2. *Let $0 < p \leq \infty$ and $s > 0$. Then*

$$\lim_{n\to\infty} n^{s/d} a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) = (\mathrm{vol}(B_p^d))^{s/d} \asymp d^{-s/p}.$$

Remark 5.3. This work continues the considerations made in [15]. Theorems 1.1, 4.1, and 5.1 extend [15, Thm. 4.3, 4.11, and 4.14], which covered only the cases $p = 1$, $p = 2$, and $p = 2s$. Moreover, by Theorem 1.1 we close the logarithmic gap in [15, Thm. 4.6] and confirm [15, Rem. 4.7].

Remark 5.4 (Approximation in L_∞)**.** In our results the approximation error is measured in $L_2(\mathbb{T}^d)$. It would be highly interesting to have analogs to Theorem 1.1, Theorem 4.1, and Theorem 5.1 for the approximation in $L_\infty(\mathbb{T}^d)$. The recent work [2] provides the formula

$$a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_\infty(\mathbb{T}^d)) = \Big(\sum_{j\geq n} \sigma_j^2\Big)^{1/2},$$

where $(\sigma_n)_{n\in\mathbb{N}}$ is again the non-increasing rearrangement of the inverse weight sequence $(1/w_{s,p}(k))_{k\in\mathbb{Z}^d}$. In principle, this allows to prove an analog to Theorem 4.1. However, the constants in the known bounds for the entropy numbers of the embbeding $\mathrm{id} : \ell_p^d \to \ell_\infty^d$ are not good enough to obtain meaningful preasymptotics.

Remark 5.5 (Entropy numbers and ridge functions)**.** Bounds as in Theorems 1.1 and 5.1 appear also in a recent work by two of the authors on a totally different subject, namely the reconstruction of ridge functions defined on the Euclidean unit ball from a limited number of function values [18]. A ridge function is a highly structured function taking the form $f(x) = g(a \cdot x)$, where g is univariate and $a \in \mathbb{R}^d$. It is no coincidence that the bounds in Theorems 1.1 and 5.1 and [18, Prop. 4.1, 4.2, Thm. 4.3] show a similar behavior in n and d. In both situations, entropy numbers of the embeddings $\mathrm{id} : \ell_p^d \to \ell_q^d$ raised to a smoothness parameter characterize the worst-case error.

Though compressibility constraints play a role in each of the problems, the used algorithms exploit the compressibility totally differently. While in the present paper algorithms benefit from compressibility by collecting information predominantly along coordinate axes, a higher degree of compressibility in the ridge function recovery problem allows algorithms to spread sampling points less dense but still in a uniform fashion across the domain.

6 Spaces of Gevrey type

In this section, we study approximation numbers of spaces $G^{\alpha,\beta,p}(\mathbb{T}^d) = H^{\mathbf{w}^G_{\alpha,\beta,p}}(\mathbb{T}^d)$ with exponential weights given by

$$w^G_{\alpha,\beta,p}(k) = \exp(\beta\|k\|_p^\alpha), \quad k \in \mathbb{Z}^d. \tag{24}$$

As already indicated in the introduction, the study of spaces $G^{\alpha,\beta,p}(\mathbb{T}^d)$ is motivated by classical Gevrey classes. Let us elaborate a bit more on this before we discuss our results in detail. For the interested reader we note that a standard reference on Gevrey spaces and its applications is Rodino's book [21].

The classical Gevrey class $\mathbf{G}^\sigma(\mathbb{R}^d)$, $\sigma > 1$, consists of all $f \in C^\infty(\mathbb{R}^d)$ with the following property:

> For every compact subset $K \subset \mathbb{R}^d$ there are constants $C, R > 0$ such that for all $x \in K$ and all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ the inequality
>
> $$|D^\alpha f(x)| \le C R^{\alpha_1+\ldots+\alpha_d}(\alpha_1! \cdots \alpha_d!)^s$$
>
> holds.

If f is 2π-periodic in each coordinate, i.e. if $f \in C^\infty(\mathbb{T}^d)$, the growth conditions on the derivatives can be rephrased in terms of Fourier coefficients: f belongs to $\mathbf{G}^\sigma(\mathbb{R}^d)$ if and only if there exists a constant $\beta > 0$ such that

$$\sum_{k\in\mathbb{Z}^d} \exp(2\beta\,\|k\|_1^{1/\sigma})|c_k(f)|^2 < \infty\,.$$

Here one can replace $\|k\|_1$ by any other (quasi-)norm on $\mathbb{R}^d$. This gives only a different constant β, but the exponent $1/\sigma$ does not change. This was the motivation in [14] to introduce the periodic Gevrey spaces $G^{\alpha,\beta,p}(\mathbb{T}^d)$, $0 < \alpha < 1$, $0 < \beta, p < \infty$, which consist of all $f \in C^\infty(\mathbb{T}^d)$ such that the norm

$$\|f|G^{\alpha,\beta,p}(\mathbb{T}^d)\| := \Big(\sum_{k\in\mathbb{Z}^d} \exp(2\beta\,\|k\|_p^\alpha)|c_k(f)|^2\Big)^{1/2}$$

is finite. For convenience of notation we changed the exponent, setting $\alpha := 1/\sigma$. Clearly, all these spaces are Hilbert spaces.

In the definition of $G^{\alpha,\beta,p}(\mathbb{T}^d)$ one can extend the range of parameters to $\alpha > 0$. The decisive difference is that the periodic Gevrey spaces, i.e. those with $0 < \alpha < 1$, contain non-analytic functions, while for $\alpha \ge 1$ all functions in $G^{\alpha,\beta,p}(\mathbb{T}^d)$ are analytic.

We come to the first result of this section. As an immediate consequence of Propositions 2.1, 2.2 and Theorem 4.3 we obtain

Theorem 6.1. *Let $\alpha, \beta > 0$ and $0 < p \le \infty$. Consider the approximation numbers*

$$a_n := a_n(\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)).$$

(i) *For* $1 \le n \le d$, *we have* $a_n \asymp_{\alpha,\beta,p} 1$.

(ii) *For* $d \le n \le 2^d$, *we have*

$$-\ln(a_n) \asymp_{\alpha,p} \beta \left[\frac{\log(n)}{\log(1+d/\log(n))}\right]^{\alpha/p}.$$

(iii) *For* $n \ge 2^d$, *we have*

$$-\ln(a_n) \asymp_{\alpha,p} \beta d^{\alpha/p} n^{\alpha/d}.$$

The limit result in Theorem 4.4 can be specialized as follows. Unfortunately, our proof technique does not work for classes of analytic functions.

Theorem 6.2. *Let* $0 < p \le \infty$, $0 < \alpha < \min\{1, p\}$, *and* $\beta > 0$. *For*

$$a_n := a_n(\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)),$$

we have

$$\lim_{n\to\infty} a_n \cdot \exp(\beta \operatorname{vol}(B_p^d)^{-\alpha/d} n^{\alpha/d}) = 1.$$

Proof. Let $\tilde{p} := \min\{1, p\}$. Further, let $h_2(x) = ((x+1)^{\tilde{p}} + d^{\tilde{p}/p}/2^{\tilde{p}})^{1/\tilde{p}}$ and $h_3(x) = (x^{\tilde{p}} - d^{\tilde{p}/p}/2^{1-\tilde{p}})^{1/\tilde{p}} - 1$. If we put $x_n = n^{1/d} \operatorname{vol}(B_p^d)^{-1/d}$, then the two-sided estimate (23) in combination with Lemma 3.3 can be reformulated as

$$\exp(\beta(x_n^\alpha - h_2(x_n)^\alpha)) \le a_n \exp(\beta \operatorname{vol}(B_p^d)^{-\alpha/d} n^{\alpha/d}) \le \exp(\beta(x_n^\alpha - h_3(x_n)^\alpha)).$$

Copying the arguments given below of (23) we conclude that $\lim_{n\to\infty}(x_n^\alpha - h_2(x_n)^\alpha) = 0$ if $\alpha < \tilde{p}$. Using similar arguments, we also get that $\lim_{n\to\infty}(x_n^\alpha - h_3(x_n)^\alpha) = 0$ if $\alpha < \tilde{p}$. □

6.1 A connection with spaces of dominating mixed smoothness

In the special situation $\alpha = p$, the estimate in Theorem 6.1 (ii) can be written more transparently. Namely, for $d \le n \le 2^d$, we have constants $c_1(p)$ and $c_2(p)$ such that

$$n^{-\frac{c_1(p)\beta}{\log(1+d/\log(n))}} \le a_n(\mathrm{Id} : G^{p,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le n^{-\frac{c_2\beta}{\log(1+d/\log(n))}} \le n^{-\frac{c_2(p)\beta}{\log(1+d)}}. \tag{25}$$

We see that the dimension d affects the polynomial decay in n only logarithmically. In information-based complexity, such a decay behavior is called *quasi-polynomial*. This observation is highly remarkable for the following reason. The preasymptotic characteristics in (25) closely resemble the preasymptotics observed in [16] for embeddings of Sobolev spaces with dominating mixed smoothness. Concretely, the recent paper [16], involving two of the present authors, studies approximation numbers of the embedding $\mathrm{Id} : H^s_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$, where the Sobolev space with dominating mixed smoothness $H^s_{\mathrm{mix}}(\mathbb{T}^d)$ is equipped with one of the—in the classical sense equivalent—norms

$$\|f|H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d)\| := \left(\sum_{k\in\mathbb{Z}^d} |c_k(f)|^2 \prod_{j=1}^d (1+|k_j|^p)^{2s/p}\right)^{1/2}, \quad p \in \{1,2\}.$$

Now, in the preasymptotic range $1 \leq n \leq 4^d$, the authors of [16] observe

$$2^{-s}\left(\frac{1}{2n}\right)^{\frac{s}{c_1(p)+\log(1/2+d/\log(n))}} \leq a_n(\mathrm{Id}: H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq \left(\frac{e^2}{n}\right)^{\frac{c_2(p)s}{2+\log_2 d}}, \qquad (26)$$

where $c_1(1) = 0$, $c_1(2) = 2$, and $c_2(p) = 1/p$ for $p \in \{1, 2\}$, see [16, Thm. 4.9, 4.10, 4.17].

The close resemblance of (25) and (26) is rather counterintuitive. After all, the space $G^{p,s,p}(\mathbb{T}^d)$ contains much smoother functions than the Sobolev space with dominating mixed regularity $H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d)$, which is clearly visible in the asymptotic decay, see Remark 6.5. But apparently, the stronger notion of smoothness does not pay off in the preasymptotic range. Let us try to gain a deeper understanding of this unexpected relationship between spaces of Gevrey type and Sobolev spaces of dominating mixed smoothness. From the simple estimate $\prod_{j=1}^d (1 + |k_j|^p)^{s/p} \leq \exp(s/p\|k\|_p^p)$ we conclude that we have the norm-one embedding

$$G^{p,s/p,p}(\mathbb{T}^d) \hookrightarrow H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d). \qquad (27)$$

Hence, the lower bound in (25), with $\beta = s/p$, yields a lower bound for the approximation numbers $a_n(\mathrm{Id}: H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$, which is only slightly worse than (26) with regard to the polynomial decay in n. Note that (26) has been obtained by doing the combinatorics explicitly for this special situation, whereas (25) followed immediately from the characterization provided by Theorem 4.3 and the known behavior of the entropy numbers $\varepsilon_n(\mathrm{id}: \ell_p^d \to \ell_\infty^d)$, see Proposition 2.2.

In view of the norm-one embedding (27), the surprising part in fact is that the upper bound in (26) is not substantially worse than (25). For the simplest case $p = 1$ and $s = 1$, there is a good explanation in terms of grid numbers. The interested reader will find it easy to generalize this to $s > 1$. Consider the grid numbers of the ℓ_1^d-ball

$$\ln(r)B_1^d = \{x \in \mathbb{R}^d : \exp(\|x\|_1) \leq r\}$$

and the hyperbolic cross

$$\mathcal{H}_r^d := \{x \in \mathbb{R}^d : \prod_{j=1}^d (1 + |x_j|) \leq r\}.$$

The first determine the behavior of the approximation numbers $a_n(\mathrm{Id}: G^{1,1,1}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$, since $G(\ln(r)B_1^d) = \sharp\{k \in \mathbb{Z}^d : w_{1,1,1}(k) \leq r\}$ (recall the considerations made in Section 4). The latter determine the approximation numbers $a_n(\mathrm{Id}: H^{s,p}_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$. We will show now that these grid numbers behave sufficiently similar for $1 \leq r \leq 2^d$. Essential ingredient of the proof is the observation that, for $l \leq \log_2(r)$, the projections $P_l \ln(r)B_1^d$ and $P_l\mathcal{H}_r^d$ have similar volumes, where $P_l: \mathbb{R}^d \to \mathbb{R}^l$, $(x_1, \ldots, x_d) \mapsto (x_1, \ldots, x_l)$.

Lemma 6.3. *Let $1 \leq r \leq 2^d$. Then, we have*

$$G(\ln(r)B_1^d) \leq G(\mathcal{H}_r^d) \leq rG(c\ln(r)B_1^d),$$

where $c = 1 + 1/\ln(2)$.

Proof. The left-hand side follows trivially from $\prod_{j=1}^{d}(1+|x_j|) \leq \exp(\|x\|_1)$. For the right-hand side estimate, we first note that

$$G(\mathcal{H}_r^d) = 1 + \sum_{l=1}^{\log_2(r)} 2^l \binom{d}{l} A(r,l),$$

where $A(r,l) := \sharp\{k \in \mathbb{N}^l : \prod_{j=1}^{d}(1+k_j) \leq r\}$, see [16, Lem. 3.1]. Further, for $2^l \leq r$, we have $A(r,l) \leq v_l(r)$, where $v_l(r) = \mathrm{vol}(\mathcal{H}_r^l \cap \{x \in \mathbb{R}^l : x_j \geq 1\})$, and $v_l(r) \leq r\frac{(\ln(r))^{l-1}}{(l-1)!}$, see [16, Lem. 3.2]. Consider now

$$\begin{aligned} G(r,l) &:= \{k \in \mathbb{N}^l : \exp(\|k\|_1) \leq r\}, \\ w_l(r) &:= \mathrm{vol}(\ln(r)B_1^l \cap \{x \in \mathbb{R}^l : x_j \geq 0\}) = \frac{(\ln(r))^l}{l!}. \end{aligned}$$

For $k \in \mathbb{Z}^l$ let $Q_k^l = k + [0,1]^l$. From the two-sided set inclusion

$$\{x \in \mathbb{R}^l : x_j \geq 1,\ \exp(\|x\|_1) \leq r\} \subset \bigcup_{k \in G(r,l)} Q_k^l \subset \{x \in \mathbb{R}^l : x_j \geq 0,\ \exp(\|x\|_1) \leq r\}$$

we obtain, by taking volumes and a change of variables, the two-sided estimate

$$w_l(r/e^l) \leq \sharp G(r,l) \leq w_l(r).$$

Now, using $l \leq \ln(r)/\ln(2)$, we observe

$$A(r,l) \leq v_l(r) = r w_{l-1}(r) \leq r w_l(r) \leq r \sharp G(e^l r, l) \leq r \sharp G(r^c, l).$$

It remains to note that $G(c\ln(r)B_1^d) = 1 + \sum_{l=1}^{\log_2(r)} 2^l \binom{d}{l} \sharp G(r^c, l)$, which can be seen easily by adopting the proof of [16, Lem. 3.1]. □

Remark 6.4. The direct preasymptotic calculations made in [16, Thm. 4.9] for the Sobolev space of dominating mixed smoothness can be adopted for the space of Gevrey type $G^{1,s,1}(\mathbb{T}^d)$ using the elements introduced in the proof of Lemma 6.3. This yields, for $1 \leq n \leq 2^d$, the estimate from above

$$a_n(\mathrm{Id} : G^{1,s,1}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq \left(\frac{e^2}{n}\right)^{\frac{s}{1+\log_2(d)}}.$$

Compare with (25).

Remark 6.5. In contrast to the preasymptotic range, the approximation numbers $a_n(\mathrm{Id} : G^{p,s/p,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ and $a_n(\mathrm{Id} : H^s_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d))$ behave asymptotically completely different. On one side, we have the well-known result

$$a_n(\mathrm{Id} : H^s_{\mathrm{mix}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \asymp_d n^{-s}(\ln n)^{(d-1)s}$$

for the Sobolev space with dominating mixed smoothness, see [16] and the references therein. On the other side, we learn from Theorem 6.1 (iii) that

$$q_1^{-sdn^{p/d}} \leq a_n(\mathrm{Id} : G^{p,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \leq q_2^{-sdn^{p/d}},$$

where $q_1 = \exp(c_1/p), q_2 = \exp(c_2/p)$.

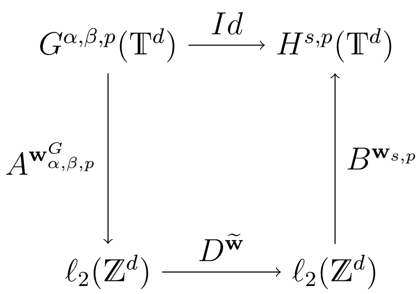

Figure 2: Commutative diagram for the embedding $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)$.

6.2 Preasymptotics for embeddings into $\mathbf{H}^{\mathbf{s}}$

In this section, we consider approximation numbers of the embedding

$$\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d),$$

assuming $s \leq \beta\alpha$. Here, as before, $G^{\alpha,\beta,p}(\mathbb{T}^d)$ is the periodic space of Gevrey type defined by the weight sequence $\mathbf{w}^G_{\alpha,\beta,p}$, see (24), and $H^{s,p}(\mathbb{T}^d)$ is the isotropic periodic Sobolev space defined by the weight sequence $\mathbf{w}_{s,p}$, see (4).

For $\mathbf{w}$ an arbitrary weight sequence, recall the operators $A^{\mathbf{w}}$, $D^{\mathbf{w}}$, and F defined in (18), (20) and (19) in Section 4. Further, let

$$B^{\mathbf{w}} : \ell_2(\mathbb{Z}^d) \to H^{\mathbf{w}}(\mathbb{T}^d),\ (\xi_k)_{n\in\mathbb{Z}^d} \mapsto (2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \xi_k/w(k)e^{ikx}.$$

We can write the embedding $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)$ as $\mathrm{Id} = B^{\mathbf{w}_{s,p}} \circ D^{\widetilde{\mathbf{w}}} \circ A^{\mathbf{w}^G_{\alpha,\beta,p}}$, where $\widetilde{\mathbf{w}} = \mathbf{w}^G_{\alpha,\beta,p}/\mathbf{w}_{s,p}$, see Figure 2 for an illustration. At the same time, we also have $\mathrm{Id} = F \circ D^{\widetilde{\mathbf{w}}} \circ A^{\widetilde{\mathbf{w}}}$. Hence, recalling the considerations made at the beginning of Section 4, it is clear that

$$a_n(\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)) = a_n(\mathrm{Id} : H^{\widetilde{\mathbf{w}}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)).$$

Note that $\widetilde{w}(k) = \widetilde{\varphi}(\|k\|_p)$, where

$$\widetilde{\varphi}(t) = \exp(\beta t^\alpha)/t^s. \tag{28}$$

Since we have assumed $s \leq \beta\alpha$, the function $\widetilde{\varphi}$ is monotonically increasing for all $t \geq 1$. Consequently, we may apply Theorem 4.3 and obtain the following worst-case error estimates.

Theorem 6.6. *Let $\alpha, \beta, s > 0$, such that $s \leq \beta\alpha$, and $0 < p \leq \infty$. Consider the approximation numbers*

$$a_n := a_n(\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)).$$

(i) For $1 \leq n \leq d$, we have $a_n \asymp_{\alpha,\beta,s,p} 1$.

(ii) For $d \leq n \leq 2^d$, we have

$$-\ln(a_n) + s\ln\left(\frac{\log(n)}{\log(1+d/\log(n))}\right) \asymp_{\alpha,s,p} -1 + \beta\left(\frac{\log(n)}{\log(1+d/\log(n))}\right)^{\alpha/p}.$$

(iii) For $n \geq 2^d$, we have

$$-\ln(a_n) + s\ln(d) + s/d\ln(n) \asymp_{\alpha,s,p} -1 + \beta d^{\alpha/p} n^{\alpha/d}.$$

In case that $\alpha = p$, $\beta = r/p$, which is particularly interesting in view of a comparison with spaces of dominating mixed smoothness and the discussion in Subsection 1.4, we can rewrite Theorem 6.6 (ii) as follows.

Corollary 6.7. *Consider the approximation numbers*

$$a_n := a_n(\mathrm{Id} : G^{p,r/p,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)).$$

For $1 \leq n \leq 2^d$, we have

$$\begin{aligned} c_1(p,s)\left(\frac{\log(n)}{\log(1+d/\log(n))}\right)^{s/p} n^{-\frac{c_1(p)r}{p\log(1+d/\log(n))}} \\ \leq a_n \leq c_2(p,s)\left(\frac{\log(n)}{\log(1+d/\log(n))}\right)^{s/p} n^{-\frac{c_2(p)r}{p\log(1+d/\log(n))}}. \end{aligned} \tag{29}$$

Proof. Let $\alpha = p$, $\beta = r/p$. Then the asserted follows by Theorem 6.6 (ii). □

As a last point in this section we provide the following limit result.

Theorem 6.8. *Let $0 < p \leq \infty$, $0 < \alpha < \min\{1, p\}$, and $\beta > 0$. For*

$$a_n := a_n(\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to H^{s,p}(\mathbb{T}^d)),$$

we have

$$\lim_{n\to\infty} a_n \cdot \frac{\exp(\beta \operatorname{vol}(B_p^d)^{-\alpha/d} n^{\alpha/d}) \operatorname{vol}(B_p^d)^{s/d}}{n^{s/d}} = 1.$$

Proof. Let $\varepsilon_n := \varepsilon_n(\mathrm{id} : \ell_p^d \to \ell_\infty^d)$. The general estimate (23) now takes the form

$$\frac{1}{\widetilde{\varphi}(h_2(1/(2\varepsilon_n)))} \leq a_n \leq \frac{1}{\widetilde{\varphi}(h_1(1/(2\varepsilon_n)))},$$

where $\widetilde{\varphi}$ is defined in (28) and h_1, h_2 are defined in the proof of Theorem 4.4. This can be further estimated to

$$\frac{1}{(2\varepsilon_n)^s \exp(\beta h_2(1/(2\varepsilon_n))^\alpha)} \leq a_n \leq \frac{1}{(2\varepsilon_n)^s \exp(\beta h_1(1/(2\varepsilon_n))^\alpha)}.$$

Writing $x_n = n^{1/d} \operatorname{vol}(B_p^d)^{-1/d}$, it is easy to see that plugging in the estimates of Lemma 3.3 leads to

$$\left(1 - \frac{2^{1-p} d^{1/p}}{x_n}\right)^{s/p} \exp(\beta(x_n^\alpha - h_2(x_n)^\alpha) \le a_n \frac{\exp(\beta \operatorname{vol}(B_p^d)^{-\alpha/d} n^{\alpha/d}) \operatorname{vol}(B_p^d)^{s/d}}{n^{s/d}}$$

and

$$a_n \frac{\exp(\beta \operatorname{vol}(B_p^d)^{-\alpha/d} n^{\alpha/d}) \operatorname{vol}(B_p^d)^{s/d}}{n^{s/d}} \le \exp(\beta(x_n^\alpha - h_3(x_n)^\alpha)),$$

where h_3 is defined in the proof of Theorem 6.2. It remains to apply the arguments which we already used in the proof of Theorem 6.2. □

7 Tractability analysis

We conclude this paper with a tractability discussion. The tractability results follow more or less immediately from the worst-case error bounds which we have derived in the preceding sections.

Theorem 7.1. *Let $s > 0$ and $0 < p \le \infty$. Then the approximation problem*

$$\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$$

(i) suffers from the curse of dimensionality iff $p = \infty$ (for all $s > 0$),

(ii) does not suffer from the curse of dimensionality iff $p < \infty$ and $s > 0$,

(iii) is intractable iff $p < \infty$ and $s \le p$,

(iv) is weakly tractable iff $p < \infty$ and $s > p$.

Theorem 7.2. *Let $\alpha, \beta > 0$ and $0 < p \le \infty$. Then the approximation problem*

$$\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$$

is quasi-polynomial tractable if and only if $\alpha \ge p$.

Before we turn to the proofs of Theorems 7.1, 7.2, let us stress at this point that for the situations discussed here, the decay of approximation numbers in the preasymptotic range determines the tractability. This is a particularly interesting observation regarding the isotropic Sobolev space. As we have already pointed out in Section 1, the asymptotic decay

$$a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \asymp_{s,p,d} n^{-s/d}$$

is often considered a typical indicator for the curse of dimensionality. However, as Theorem 7.1 shows, the approximation problem suffers only from the curse of dimensionality in the strict sense of information-based complexity when we equip the isotropic Sobolev space with the norm $\|\cdot \mid H^{s,\infty}(\mathbb{T}^d)\|$. Otherwise, the approximation problem is weakly tractable, despite the bad asymptotic decay $n^{-s/d}$. For $p = 1, 2, 2s$ this has already been observed in [15], see Remark 7.5. Concerning spaces of Gevrey type, it is no surprise in light of Section 6.1 that we obtain a similar tractability as has been observed for Sobolev spaces with dominating mixed regularity in [16]. For some further remarks on the tractability of Gevrey embeddings, see Remark 7.7.

For the proof of Theorem 7.1 we have to translate the bounds of Theorem 1.1 into bounds for the information complexity. These bounds are given in Lemma 7.3. We omit the proof, which is technical and lengthy but requires only standard arguments.

Lemma 7.3. *For $s > 0$ and $0 < p < \infty$, consider the information complexity*

$$n(\varepsilon, d) = \min\{n \in \mathbb{N} : a_n(\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le \varepsilon\}.$$

(i) From above, we have the bounds

$$\log n(\varepsilon, d) \lesssim_{s,p} \begin{cases} \log(d) & : \varepsilon_1^U \le \varepsilon \le 1 \\ \log(d)\ (1/\varepsilon)^{p/s} & : \varepsilon_2^U \le \varepsilon \le \varepsilon_1^U \\ \log(1/\varepsilon)\ (1/\varepsilon)^{p/s} & : \varepsilon_3^U(\gamma) \le \varepsilon \le \varepsilon_2^U \\ \log(1/\varepsilon)\ (1/\varepsilon)^{\frac{p\gamma}{s(p+\gamma)}} & : \varepsilon \le \varepsilon_3^U(\gamma) \end{cases}$$

where $\gamma \ge 0$ and

$$\varepsilon_1^U := C_{s,p}\left[\frac{\log(1 + d/\log d)}{\log d}\right]^{s/p}, \quad \varepsilon_2^U := C_{s,p} d^{-s/p}, \quad \varepsilon_3^U := C_{s,p} 2^{-s} d^{-s(1/p+1/\gamma)},$$

The constant $C_{s,p}$ is the same as in the upper bound of Theorem 1.1.

(ii) From below, we have the bound

$$\log n(\varepsilon, d) \gtrsim_{s,p} (1/\varepsilon)^{p/s} \quad \text{for } \varepsilon_2^L \le \varepsilon \le \varepsilon_1^L,$$

where

$$\varepsilon_1^L := c_{s,p}\left[\frac{\log(1 + d/\log(d))}{\log(d)}\right]^{s/p}, \quad \varepsilon_2^L := c_{s,p} 2^{-s} (1/d)^{s/p}.$$

The constant $c_{s,p}$ is identical to the one in the lower bound of Theorem 1.1.

Proof of Theorem 7.1.

(i) For $n \le 2^d$, Theorem 5.1 states that $a_n = 1$. Hence, we have $n(\varepsilon, d) \ge 2^d$ for all $\varepsilon < 1$ and the problem suffers from the curse of dimensionality.

(ii) We show that for there is an $\varepsilon > 0$ such that $n(\varepsilon, d)$ is polynomial in d. Fix some $\varepsilon > \varepsilon_2^U$. By Lemma 7.3 (i), there is $\tilde{C}_{s,p} > 0$ such that

$$n(\varepsilon, d) \le d^{\tilde{C}_{s,p}(1/\varepsilon)^{p/s}}.$$

Since $\varepsilon > \varepsilon_2^U$ the above estimate holds true for all $d > C_{s,p}^{p/s}(1/\varepsilon)^{p/s}$. Consequently, the problem cannot suffer from the curse of dimensionality.

(iii) To prove intractability is suffices that there is a sequence $(\varepsilon_i, d_i)_{i\in\mathbb{N}}$ such that the limit in (14) does not exist. Let $(d_i)_{i\in\mathbb{N}}$ with $d_i \in \mathbb{N}$ and $d_i \to \infty$ for $i \to \infty$. Let $\varepsilon_2^L \le \varepsilon_i \le c_{s,p}(1/d_i)^{s/p}$. Then,

$$c_{s,p}^{p/s} 2^{-p} (1/\varepsilon_i)^{p/s} \le d_i \le c_{s,p}^{p/s} (1/\varepsilon_i)^{p/s},$$

and thus

$$\frac{\log n(\varepsilon_{i,i} d)}{d_i + 1/\varepsilon_i} \ge c_{s,p} \frac{(1/\varepsilon_i)^{p/s}}{d + 1/\varepsilon_i} \ge c_{s,p} \frac{(1/\varepsilon_i)^{p/s}}{c_{s,p}^{p/s}(1/\varepsilon_i)^{p/s} + 1/\varepsilon_i}.$$

Finally, since $p/s \ge 1$, we have $1/\varepsilon_i \le (1/\varepsilon_i)^{p/s}$ and thus

$$\frac{\log n(\varepsilon_i, d_i)}{d_i + 1/\varepsilon_i} \ge \frac{c_{s,p}}{c_{s,p}^{p/s} + 1} > 0 \quad \text{for all } i \in \mathbb{N}.$$

In consequence, the problem is not weakly tractable and must be intractable.

(iv) We have to show that the information complexity grows slower than both $2^{1/\varepsilon}$ and 2^d. Put $x = 1/\varepsilon + d$. Since both $1/\varepsilon \le x$ and $d \le x$, we have for all ε and all d that

$$\log n(\varepsilon, d) \le \tilde{C}_{s,p} \log(x) x^{p/s}.$$

Hence, $\lim_{x\to\infty} \log n(\varepsilon, d)/x = 0$ as $p < s$.

□

The tractability analysis for the approximation problem $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ can be reduced to the tractability analysis for the problem $\mathrm{Id} : H^{\mathbf{w}_p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$. Basis is the following general observation.

Lemma 7.4. *Let* $\mathbf{w}$ *be an arbitrary weight sequence and let*

$$n^{\mathbf{w}}(\varepsilon, d) := \min\{n \in \mathbb{N} : a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le \varepsilon\}.$$

For $\varphi : [0, \infty) \to [0, \infty)$ *monotonically increasing, consider the weight sequence* $\varphi(\mathbf{w})$ *given by* $\varphi(\mathbf{w})(0) := 1$ *and* $\varphi(\mathbf{w})(k) := \varphi(w(k))$ *for* $k \in \mathbb{Z}^d \setminus \{0\}$. *We have*

$$n^{\varphi(\mathbf{w})}(\varepsilon, d) = n^{\mathbf{w}}(1/\varphi(1/\varepsilon), d).$$

Proof. Let $(\sigma_n)_{n\in\mathbb{N}}$ and $(\gamma_n)_{n\in\mathbb{N}}$ be the non-increasing rearrangements of $1/\mathbf{w}$ and $1/\varphi(\mathbf{w})$, respectively. Then, using (22), we obtain

$$\begin{aligned} n^{\varphi(\mathbf{w})}(\varepsilon, d) &= \min\{n \in \mathbb{N} : \gamma_n \le \varepsilon\} = \min\{n \in \mathbb{N} : 1/\varphi(1/\sigma_n) \le \varepsilon\} \\ &= n^{\mathbf{w}}(1/\varphi^{-1}(1/\varepsilon), d). \end{aligned}$$

□

Proof of Theorem 7.2. With $\varphi(t) = \exp(\beta t^\alpha)$ and $\gamma = p$, Lemma 7.4 in combination with Lemma 7.3 (i) yields

$$\ln n(\varepsilon, d) \lesssim_{\alpha,\beta,p} \begin{cases} \ln(d) & : \tilde{\varepsilon}_1^U \le \varepsilon \le 1 \\ \ln(d)\ln(1/\varepsilon)^{p/\alpha} & : \tilde{\varepsilon}_2^U \le \varepsilon \le \tilde{\varepsilon}_1^U \\ \ln(\ln(1/\varepsilon))\ln(1/\varepsilon)^{p/\alpha} & : \tilde{\varepsilon}_3^U \le \varepsilon \le \tilde{\varepsilon}_2^U \\ \ln(\ln(1/\varepsilon))\ln(1/\varepsilon)^{p/(2\alpha)} & : \varepsilon \le \tilde{\varepsilon}_3^U \end{cases}$$

where $\tilde{\varepsilon}_i^U = 1/\varphi(1/\varepsilon_i^U)$. Since in the third case we may estimate $\ln(\ln(1/\varepsilon)) \lesssim_{\alpha,\beta,p} \ln(d)$ due to $\tilde{\varepsilon}_3^U \le \varepsilon$ and in the forth case we may estimate $\ln(\ln(1/\varepsilon)) \lesssim_{\alpha,\beta,p} \ln(1/\varepsilon)^{p/(2\alpha)}$, we obtain

$$\ln n(\varepsilon, d) \lesssim_{\alpha,\beta,p} \ln(d)\ln(1/\varepsilon)^{p/\alpha}$$

for all $0 < \varepsilon \le 1$ and $d \in \mathbb{N}$, which leads to quasi-polynomial tractability if $\alpha \ge p$. That $\alpha \ge p$ is also a necessary condition for quasi-polynomial tractability follows immediately by Lemma 7.4 and Lemma 7.3 (ii). □

Remark 7.5. The tractability of approximating the identity $\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ by finite-rank operators has already been studied in [15] for $p = 1$, $p = 2$, and $p = 2s$. In the case $p = 2$, however, the authors could not show whether the problem is intractable or weakly tractable when $1 < s \le 2$, see [15, Thm. 5.5, Cor. 5.7]. The results from Section 4 allow to close this gap and furthermore to reproduce all tractability results obtained in [15]. For a different proof that allows to close the gap, we refer to the recent paper [24].

Remark 7.6. Concerning the standard notions of tractability, asking for compressibility of frequency vectors ($0 < p \le 1$) only has the effect that we need less smoothness to obtain weak tractability, see Theorem 7.1, (iv). To get a comprehensive understanding of the effect of compressibility, we need two additional notions of tractability introduced only recently. For $\alpha, \beta > 0$ a problem is called *(α, β)-weakly tractable* [24] if

$$\lim_{1/\varepsilon + d \to \infty} \frac{\log n(\varepsilon, d)}{1/\varepsilon^\alpha + d^\beta} = 0\,.$$

A problem is called *uniformly weakly tractable* [23] if it is (α,β)-weakly tractable for all α,$\beta > 0$. From Lemma 7.3 we can conclude that the approximation problem $\mathrm{Id} : H^{s,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ is *(α, β)-weakly tractable* for $\alpha > p/s$ and all $\beta > 0$ (which has also been observed in [24, Thm 4.1]). Hence, if we impose a very strong compressibility constraint—which means that p gets small—then we have almost *uniform weak tractability.*

Remark 7.7. The recent paper [3] studies the tractability of approximating embeddings $\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ by operators of finite-rank for weight sequences $\mathbf{w}$ of the form $w(k) = \omega^{\sum_{j=1}^d a_j |k_j|^{b_j}}$, $k \in \mathbb{Z}^d$, where $\omega > 1$, $0 < a_1 \le a_2 \le a_3 \le \dots$ and $\inf b_j > 0$. Let

$$n^{\mathbf{w}}(\varepsilon, d) := \min\{n \in \mathbb{N} : a_n(\mathrm{Id} : H^{\mathbf{w}}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \le \varepsilon\}$$

be the information complexity of the approximation problem. In [3] it is studied under which conditions a modified, stronger notion of weak tractability is satisfied, namely

$$\lim_{\ln(1/\varepsilon)+d\to\infty} \frac{\ln n^{\mathbf{w}}(\varepsilon,d)}{\ln(1/\varepsilon)+d} = 0. \tag{30}$$

The Gevrey weights $\mathbf{w}^G_{\alpha,\beta,p}$, defined in (24), fit into the setting of [3] if $\alpha = p$ (by choosing $a_1 = a_2 = \cdots = \beta$ and $b_1 = b_2 = \cdots = p$). From [3, Thm. 1] it is immediately clear that the approximation problem $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ is not weakly tractable in the above sense if $\alpha = p$. What more can be said? From Lemma 7.4 we get

$$\lim_{1/\varepsilon+d} \frac{\ln n^{\mathbf{w}_{\alpha,p}}(\varepsilon,d)}{d+1/\varepsilon} = \lim_{1/\varepsilon+d} \frac{\ln n^{\mathbf{w}^G_{\alpha,\beta,p}}(\varepsilon,d)}{d+\ln(1/\varepsilon)}.$$

Hence the approximation problem $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ is weakly tractable in the modified sense (30) if and only if the approximation problem $\mathrm{Id} : H^{\alpha,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$ is weakly tractable in the classical sense, which is the case if and only if $\alpha > p$. We conclude that weak tractability in the modified sense (30) is almost equivalent to quasi-polynomial tractability for the approximation problem $\mathrm{Id} : G^{\alpha,\beta,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)$, cf. Theorem 7.2.

As a final remark let us point out that other than claimed in [3] the space $H^{\mathbf{w}}(\mathbb{T}^d)$ consists of analytic functions if and only if $\inf b_j \geq 1$. The proof provided in [3, Section 10] is wrong, and even under the additional assumption $\inf b_j \geq 1$ incomplete as it only shows convergence of the Taylor expansion. For a correct proof, see [11].

Acknowledgments

Tino Ullrich and Sebastian Mayer gratefully acknowledge support by the German Research Foundation (DFG) and the Emmy-Noether programme, Ul-403/1-1. Thomas Kühn was supported in part by the Spanish Ministerio de Economía y Competitividad (MTM2013-42220-P).

References

[1] Jean Céa. Approximation variationnelle des problèmes aux limites. In *Annales de l'institut Fourier*, volume 14, pages 345–444, 1964.

[2] Fernando Cobos, Thomas Kühn, and Winfried Sickel. Optimal approximation of multivariate periodic Sobolev functions in the sup-norm. *arXiv preprint arXiv:1505.02636*, 2015.

[3] Josef Dick, Peter Kritzer, Friedrich Pillichshammer, and Henryk Woźniakowski. Approximation of analytic functions in Korobov spaces. *J. Complexity*, 30(2):2–28, 2014.

[4] Dinh Dũng and Tino Ullrich. N-widths and ε-dimensions for high-dimensional approximations. *Foundations of Computational Mathematics*, 13(6):965–1003, 2013.

[5] David Eric Edmunds and Hans Triebel. *Function spaces, entropy numbers, differential operators*, volume 120. Cambridge University Press, 2008.

[6] Maurice Gevrey. Sur la nature analytique des solutions des équations aux dérivées partielles. premier mémoire. In *Annales Scientifiques de l'École Normale Supérieure*, volume 35, pages 129–190. Société mathématique de France, 1918.

[7] Michael Griebel and Jan Hamaekers. Sparse grids for the Schrödinger equation. *M2AN Math. Model. Numer. Anal.*, 41(2):215–247, 2007.

[8] Michael Griebel and Stephan Knapek. Optimized general sparse grid approximation spaces for operator equations. *Mathematics of Computation*, 78(268):2223–2257, 2009.

[9] Michael Griebel, Stephan Knapek, and Gerhard Zumbusch. Numerical simulation in molecular dynamics: Numerics, algorithms, parallelization, applications. 5:xii+470, 2007.

[10] Aicke Hinrichs and Sebastian Mayer. Entropy numbers of spheres in Banach and quasi-Banach spaces. *J. Approx. Theory*, 200:144–152, 2015.

[11] Christian Irrgeher, Peter Kritzer, Gunther Leobacher, and Friedrich Pillichshammer. Integration in Hermite spaces of analytic functions. *J. Complexity*, 31(3):380–404, 2015.

[12] Joseph W Jerome. On the l2n-width of certain classes of functions of several variables. *Journal of Mathematical Analysis and Applications*, 20(1):110–123, 1967.

[13] Thomas Kühn. A lower estimate for entropy numbers. *Journal of Approximation Theory*, 110(1):120–124, 2001.

[14] Thomas Kühn and Michael Petersen. Approximation in periodic Gevrey spaces. *Preprint*, 2015.

[15] Thomas Kühn, Winfried Sickel, and Tino Ullrich. Approximation numbers of Sobolev embeddings—sharp constants and tractability. *J. Complexity*, 30(2):95–116, 2014.

[16] Thomas Kühn, Winfried Sickel, and Tino Ullrich. Approximation of mixed order Sobolev functions on the d-torus: Asymptotics, preasymptotics, and d-dependence. *Constructive Approximation*, 42(3):353–398, 2015.

[17] PD Lax and AN Milgram. Parabolic equations: Contributions to the theory of partial differential equations. 1954. *Annals of mathematical studies*, 1954.

[18] Sebastian Mayer, Tino Ullrich, and Jan Vybíral. Entropy and sampling numbers of classes of ridge functions. *Constructive Approximation*, 42(2):231–264, 2015.

[19] Erich Novak and Henryk Woźniakowski. Tractability of multivariate problems. Vol. 1: Linear information, 2008.

[20] Christian Richter and Michael Stehling. Entropy numbers and lattice arrangements in $l_\infty(\Gamma)$. *Math. Nachr.*, 284(7):818–830, 2011.

[21] Luigi Rodino. *Linear partial differential operators in Gevrey spaces.* World Scientific, 1993.

[22] Carsten Schütt. Entropy numbers of diagonal operators between symmetric Banach spaces. *Journal of approximation theory*, 40(2):121–128, 1984.

[23] Paweł Siedlecki. Uniform weak tractability. *Journal of Complexity*, 29(6):438–453, 2013.

[24] Paweł Siedlecki and Markus Weimar. Notes on (s,t)-weak tractability: a refined classification of problems with (sub)exponential information complexity. *J. Approx. Theory*, 200:227–258, 2015.

[25] V. N. Temlyakov. *Approximation of periodic functions.* Computational Mathematics and Analysis Series. Nova Science Publishers, Inc., Commack, NY, 1993.

[26] VM Tikkomirov. Approximation theory. In *Analysis II*, pages 93–243. Springer, 1990.

[27] Hans Triebel. Fractals and spectra: Related to Fourier analysis and function spaces. pages viii+271, 2011.

[28] Xianfu Wang. Volumes of generalized unit balls. *Mathematics Magazine*, 78(5):390–395, 2005.

[29] Harry Yserentant. *Regularity and approximability of electronic wave functions.* Springer, 2010.