Lipschitz widths

Guergana Petrova, Przemysław Wojtaszczyk *

October 26, 2022

To Ron DeVore, with the utmost respect and admiration

Abstract

This paper introduces a measure, called Lipschitz widths, of the optimal performance possible of certain nonlinear methods of approximation. Notably, those Lipschitz widths provide a theoretical benchmark for the approximation quality achieved via deep neural networks. The paper also discusses basic properties of the Lipschitz widths and their relation to entropy numbers and other well known widths such as the Kolmogorov and the stable manifold widths. We show that Lipschitz widths with fixed Lipschitz constant and entropy numbers decay very similar, while when the Lipschitz constant grows with n, the Lipschitz width could be much smaller than the entropy numbers.
AMS subject classification: 41A46, 41A65, 82C32
Key words: widths, entropy numbers, neural networks

1 Introduction

Nonlinear methods of approximation provide reliable and efficient ways of investigating the underlying phenomena in many application areas. Despite of their extensive usage however, there is still a lack of comprehensive understanding of the intrinsic limitations of these nonlinear methods, even on a purely theoretical level. Several mathematical concepts, called widths, have been established to access numerous aspects of the quality of linear and nonlinear approximations. As such, we mention the classical by now Kolmogorov, linear, manifold, Gelfand widths, which give a theoretical benchmark on what is the best possible performance of particular methods of approximation. We refer the reader to [8], where a summary of different nonlinear widths and their relations to one another is discussed.

Recently, Deep Neural Networks (DNN) have been used extensively as a method of choice for variety of machine learning problems and as a computational platform in many other areas. Despite of their empirical successes, the explanation of the reasons behind their stellar performance is still in its infancy. On mathematical level, DNN can be viewed as a method of nonlinear approximation of an underlying function f, where the approximant $\Phi(y) \approx f$ is a continuous function, generated by a DNN with parameters y. It can be shown that the mapping which to every choice of parameters y of the DNN assigns $\Phi(y)$ is in fact a Lipschitz mapping. Thus, DNN approximation is a particular case of a nonlinear approximation of a function f, or a compact class $\mathcal{K}$, by the images of

*G.P. was supported by the NSF Grant DMS 2134077, Tripods Grant CCF-1934904, and ONR Contract N00014-20-1-278. P. W. was supported by National Science Centre, Polish grant UMO-2016/21/B/ST1/00241.

Lipschitz mappings. Then, the question of DNN optimal performance is intimately related to the quantification of the optimal performance of such nonlinear methods and to the introduction and study of corresponding ways to measure it. A width, called *stable manifold* width, was presented in [5], with the sole purpose to determine the optimal performance of such nonlinear methods in the context of numerical computation, where the stability plays an essential role. In this paper, we take a slightly different point of view and introduce the concept of *Lipschitz* widths, where we are not so concerned about the numerical stability of the method, but , like in the case of DNN, rather about the best possible performance of these nonlinear methods of approximation.

Our setting is a Banach space X equipped with a norm $\|\cdot\|_X$, where we wish to approximate the elements f of a compact subset $\mathcal{K} \subset X$ of X with error measured in this norm. For every fixed $n \in \mathbb{N}$ and every $\gamma \geq 0$, the approximants to $\mathcal{K}$ will come from the images $\Phi(y) \in X$ of all possible γ-Lipschitz maps $\Phi : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$, where B_{Y_n} is the unit ball in $\mathbb{R}^n$ with respect to some norm $\|\cdot\|_{Y_n}$ in $\mathbb{R}^n$. The quality of this approximation is a critical element in the design and analysis of various numerical methods, among which are DNNs. Note that any numerical method based on Lipschitz mappings will have performance no better than the optimal performance of this approximation method. On the other hand, it may not be easy to actually design a numerical method for a particular application that achieves this optimal performance.

In our analysis, we examine model classes $\mathcal{K} \subset X$, i.e., compact subsets $\mathcal{K}$ of X, that summarize what we know about the target function f. Classical model classes $\mathcal{K}$ are finite balls in smoothness spaces like the Lipschitz, Sobolev, or Besov spaces. The *Lipschitz* widths $d_n^\gamma(\mathcal{K})_X$ then quantify the best possible performance of the above approximation methods on a given model class $\mathcal{K}$.

The paper is organized as follows. Some of the basic properties of Lipschitz widths are discussed in §2, where we present many elementary but very useful properties of $d_n^\gamma(\mathcal{K})_X$, such as, for example, the fact that for a fixed $n \in \mathbb{N}$, $d_n^\gamma(\mathcal{K})_X$ is a continuous function of $\gamma \geq 0$, see Theorem 2.7. We also prove the statements

$$\lim_{n\to\infty} d_n^\gamma(\mathcal{K})_X = 0, \quad \text{and} \quad \lim_{\gamma\to\infty} d_n^\gamma(\mathcal{K})_X = 0,$$

each of which characterizes the set $\mathcal{K}$ as a totally bounded set, see Lemma 2.8 and Lemma 2.11. We also show that in the definition of Lipschitz width the infimum over all norms is achieved for some special norm, see Theorem 2.14. While this observation may not be very useful in practical applications, it can be helpful in theoretical considerations.

The relation between Lipschitz widths and entropy numbers $\varepsilon_n(\mathcal{K})_X$ is investigated in §3, where among other things we prove that, see Theorem 3.3, for any compact subset $\mathcal{K} \subset X$ of a Banach space X we have

$$d_n^{2\,\mathrm{rad}(\mathcal{K})}(\mathcal{K})_X \leq \varepsilon_n(\mathcal{K})_X, \quad n = 1, 2, \ldots.$$

Examples are given to show that this inequality is almost optimal. We also discuss in this section estimates from below and above for the Lipschitz width $d_n^\gamma(\mathcal{K})_X$, provided bounds for the entropy numbers $\varepsilon_n(\mathcal{K})_X$ are available. These are stated in Theorem 3.1, where we prove that when the Lipschitz constant γ is fixed, both the entropy numbers and the Lipschitz width have a similar decay. Some of our estimates are optimal, as demonstrated in Theorem 3.12, where we show that the Lipschitz widths could be smaller than the entropy numbers for certain compact classes $\mathcal{K}$.

Since the Lipschitz width is a new concept of width, we compare it with some of the well known classical widths. We show that for appropriate values of the parameter γ, Lipschitz widths are smaller than the Kolmogorov widths, see §4, Theorem 4.1. They are also smaller than the stable manifold widths, see §5, Theorem 5.1. However, as demonstrated by the provided Examples, in both cases, their actual behavior may be very different.

At last, in §6, we discuss the Lipschitz widths $d_n^{\gamma_n}(\mathcal{K})_X$, where the parameter $\gamma_n = C'\lambda^n$, with $C' > 0$ and $\lambda > 2$ being fixed constants, and show that they provide a theoretical benchmark for the performance of certain DNN approximation, see Theorem 6.1. The analysis of these widths is performed in Theorem 6.3 and Corollary 6.4, where it is demonstrated that there is indeed a gain in the performance of the Lipschitz width $d_n^{\gamma_n}(\mathcal{K})_X$ when compared to the entropy numbers $\varepsilon_n(\mathcal{K})_X$ in the following sense

$$\text{if} \quad \varepsilon_n(\mathcal{K})_X \asymp \frac{[\log_2 n]^\beta}{n^\alpha} \quad \Rightarrow \quad d_n^{\gamma_n}(\mathcal{K})_X \asymp \frac{[\log_2 n]^\beta}{n^{2\alpha}}.$$

This estimate, when applied in the case of $\mathcal{K}$ being the unit ball of certain Besov spaces, extends some results from [6] to the case when error is measured in L_p, $p \neq \infty$.

2 Definition and basic properties

We are mainly interested in compact sets, however we define the basic concepts for bounded sets. We consider a bounded subset $\mathcal{K} \subset X$ of a Banach space $(X, \|\cdot\|_X)$ with norm $\|\cdot\|_X$ and denote by $(\mathbb{R}^n, \|.\|_{Y_n})$, $n \geq 1$ the n-dimensional Banach space with a fixed norm $\|\cdot\|_{Y_n}$. For $\gamma \geq 0$, we define the *fixed Lipschitz* width

$$d^\gamma(\mathcal{K}, Y_n)_X := \inf_{\Phi_n} \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \Phi_n(y)\|_X, \tag{2.1}$$

where the infimum is taken over all Lipschitz mappings

$$\Phi_n : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X, \quad B_{Y_n} := \{y \in \mathbb{R}^n : \ \|y\|_{Y_n} \leq 1\},$$

that satisfy the Lipschitz condition

$$\sup_{y, y' \in B_{Y_n}} \frac{\|\Phi_n(y) - \Phi_n(y')\|_X}{\|y - y'\|_{Y_n}} \leq \gamma, \tag{2.2}$$

with constant γ. Next, we define the *Lipschitz* width

$$d_n^\gamma(\mathcal{K})_X := \inf_{\|\cdot\|_{Y_n}} d^\gamma(\mathcal{K}, Y_n)_X, \tag{2.3}$$

where the infimum is taken over all norms $\|\cdot\|_{Y_n}$ in $\mathbb{R}^n$. Clearly, we have that for every norm $\|\cdot\|_{Y_n}$ on $\mathbb{R}^n$,

$$d_n^\gamma(\mathcal{K})_X \leq d^\gamma(\mathcal{K}, Y_n)_X, \quad n \geq 1. \tag{2.4}$$

Before going further, let us recall the definition of a diameter and radius of a bounded set $\mathcal{M} \subset X$,

$$\operatorname{diam}(\mathcal{M}) := \sup_{f, g \in \mathcal{M}} \|f - g\|_X \leq 2 \inf_{g \in X} \sup_{f \in \mathcal{M}} \|f - g\|_X =: 2 \operatorname{rad}(\mathcal{M}).$$

From (2.2) we see that a 0-Lipschitz function is simply the constant function. Thus, for any Y_n we get

$$\operatorname{rad} \mathcal{K} = d_n^0(\mathcal{K}, Y_n)_X = d_n^0(\mathcal{K})_X. \tag{2.5}$$

We next list some elementary properties of the Lipschitz widths $d_n^\gamma(\mathcal{K})_X$ that we gather in the following remark.

Remark 2.1. *For any bounded subset $\mathcal{K} \subset X$ of a Banach space X, any $n \in \mathbb{N}$, and any $\gamma > 0$, we have*

(i) The Lipschitz width $d_n^\gamma(\mathcal{K})_X$ satisfies the relation

$$d_n^\gamma(\mathcal{K})_X = \inf_{k \leq n} \inf_{\|\cdot\|_{Y_k}} d^\gamma(\mathcal{K}, Y_k)_X. \tag{2.6}$$

(ii) The space $(\mathbb{R}^n, \|\cdot\|_{Y_n})$ in (2.1) and (2.3) can be replaced by any normed space $(X_n, \|\cdot\|_{X_n})$ of dimension n, that is

$$d_n^\gamma(\mathcal{K})_X = \inf_{\|\cdot\|_{X_n}} d^\gamma(\mathcal{K}, X_n)_X, \quad \text{where} \quad d^\gamma(\mathcal{K}, X_n)_X = \inf_{\Phi_n} \sup_{f \in \mathcal{K}} \inf_{x \in B_{X_n}} \|f - \Phi_n(x)\|_X, \tag{2.7}$$

with $B_{X_n} := \{x \in X_n : \ \|x\|_{X_n} \leq 1\}$.

(iii) $d_n^\gamma(\mathcal{K})_X$ is a monotone decreasing function of γ and n. More precisely,

- *If $\gamma_1 \leq \gamma_2$ then $d_n^{\gamma_2}(\mathcal{K})_X \leq d_n^{\gamma_1}(\mathcal{K})_X$;*
- *If $n_1 \leq n_2$ then $d_{n_2}^\gamma(\mathcal{K})_X \leq d_{n_1}^\gamma(\mathcal{K})_X$.*

(iv) For every fixed $n \in \mathbb{N}$ and $\gamma \geq 0$, we have $d_n^\gamma(\mathcal{K})_X \leq \mathrm{rad}(\mathcal{K}) < \infty$.

(v) For every fixed $n \in \mathbb{N}$ and $\gamma \geq 0$, we have $d_n^\gamma(\mathcal{K})_X = d_n^\gamma(\bar{\mathcal{K}})_X$ where $\bar{\mathcal{K}}$ denotes the closure of $\mathcal{K}$.

Proof: We discuss only (i) and leave the proof of the rest of the remark to the reader. Since

$$\inf_{k \leq n} \inf_{\|\cdot\|_{Y_k}} d^\gamma(\mathcal{K}, Y_k)_X \leq d_n^\gamma(\mathcal{K})_X,$$

to show (i), it suffices to show that for every norm $\|\cdot\|_{Y_k}$ with $1 \leq k < n$, there exists a norm $\|\cdot\|_{Y_n}$ on $\mathbb{R}^n$ such that

$$d^\gamma(\mathcal{K}, Y_n)_X \leq d^\gamma(\mathcal{K}, Y_k)_X. \tag{2.8}$$

Indeed, let us fix a γ-Lipschitz map $\Phi_k : (B_{Y_k}, \|\cdot\|_{Y_k}) \to X$ which achieves $d^\gamma(\mathcal{K}, Y_k)_X$ (if such map does not exist, we can use limiting arguments). We then define the norm $\|\cdot\|_{Y_n}$ as

$$\|(y, y')\|_{Y_n} := \|y\|_{Y_k} + \|y'\|_{\mathbb{R}^{n-k}},$$

where $\|.\|_{\mathbb{R}^{n-k}}$ is any norm in $\mathbb{R}^{n-k}$, and the γ-Lipschitz mapping $\Phi_n : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$ as

$$\Phi_n((y, y')) := \Phi_k(y),$$

which proves (2.8). □

The requirement that Φ_n is Lipschitz on the unit ball B_{Y_n} rather than on the whole $\mathbb{R}^n$ is necessary for a meaningful definition in the case of a separable Banach space X. If we assume for a moment that the Lipschitz width were defined via Lipschitz mappings on the whole $\mathbb{R}^n$, we can prove an analogue of Remark 2.1 (iii) and therefore focus on the one dimensional case which would provide an upper bound for $d_n^\gamma(\mathcal{K})_X$. For the separable Banach space X, we consider a countable dense set $\{\varphi_1, \varphi_2, \dots, \} \subset X$. For any fixed $\gamma > 0$, we define the points $\{t_j\}$ on the real line by

$$t_0 = 0, \quad t_1 = \gamma^{-1}\|\varphi_1\|_X, \quad t_{j+1} = t_j + \gamma^{-1}\|\varphi_{j+1} - \varphi_j\|_X, \quad j = 1, \dots,$$

and the mapping $\Phi_1 : \mathbb{R} \to X$ as the continuous piecewise linear function of t with values in X with the property

$$\Phi_1(t) = \begin{cases} 0, & t \le t_0, \\ \varphi_j, & t = t_j. \end{cases}$$

Clearly, $\Phi_1(\mathbb{R})$ is dense in X and Φ_1 is a γ-Lipschitz mapping. Thus, for every set $\mathcal{K} \subset X$ the one dimensional Lipschitz width $d_1^\gamma(\mathcal{K})_X$ would be zero.

Lemma 2.2. *For any bounded subset $\mathcal{K} \subset X$ of a Banach space X, any $n \in \mathbb{N}$, and any $\gamma > 0$, we can restrict the infimum in* (2.3) *only to normed spaces $(\mathbb{R}^n, \|.\|_{\mathcal{Y}_n})$ with the additional property that the norm $\|\cdot\|_{\mathcal{Y}_n}$ satisfies the condition*

$$\|y\|_{\ell_\infty^n} := \max_j |y_j| \le \|y\|_{\mathcal{Y}_n} \le \sum_{j=1}^n |y_j| =: \|y\|_{\ell_1^n}, \quad y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n. \tag{2.9}$$

Proof: Let $(\mathbb{R}^n, \|\cdot\|_{Y_n})$ be any normed space. It follows from the Auerbach lemma (see e.g. [4, p.43] or [15, II.E.11]), that we can find vectors $(\bar{v}_j)_{j=1}^n \subset R^n$ and linear functionals $(f_j)_{j=1}^n$ on the space $(\mathbb{R}^n, \|\ \|_{Y_n})$ such that

$$\|\bar{v}_j\|_{Y_n} = \|f_j\|_{Y_n^*} = 1, \quad j = 1, \dots, n, \tag{2.10}$$

and

$$f_i(\bar{v}_j) = \delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \ne j. \end{cases} \tag{2.11}$$

We define a new norm $\|.\|_{\mathcal{Y}_n}$ on $\mathbb{R}^n$ as

$$\|y\|_{\mathcal{Y}_n} := \|\sum_{j=1}^n y_j \bar{v}_j\|_{Y_n}, \quad y = (y_1, \dots, y_n) \in \mathbb{R}^n,$$

which, using the triangle inequality and (2.10), satisfies the inequality

$$\|y\|_{\mathcal{Y}_n} \le \sum_{j=1}^n |y_j| \|\bar{v}_j\|_{Y_n} = \sum_{j=1}^n |y_j|. \tag{2.12}$$

On the other hand, using (2.10) and (2.11), we have

$$\|y\|_{\mathcal{Y}_n} = \|\sum_{j=1}^n y_j \bar{v}_j\|_{Y_n} = \sup_{f:\ \|f\|_{Y_n^*}=1} |f(\sum_{j=1}^n y_j \bar{v}_j)| \ge |f_i(\sum_{j=1}^n y_j \bar{v}_j)| = |y_i|, \quad i = 1, \dots, n. \tag{2.13}$$

Therefore, it follows from (2.12) and (2.13) that the newly defined norm satisfies (2.9). If we consider the mapping ϕ_0 defined as

$$\phi_0(y) := \sum_{j=1}^n y_j \bar{v}_j, \quad y = (y_1, \dots, y_n) \in R^n,$$

one can show that $\phi_0 : (B_{\mathcal{Y}_n}, \|\cdot\|_{\mathcal{Y}_n}) \to (B_{Y_n}, \|\cdot\|_{Y_n})$ and that $\phi_0(B_{\mathcal{Y}_n}) = B_{Y_n}$. Now, for any γ-Lipschitz mapping $\Phi_n : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$, we define the map $\tilde{\Phi}_n : (B_{\mathcal{Y}_n}, \|\cdot\|_{\mathcal{Y}_n}) \to X$ as

$$\tilde{\Phi}_n := \Phi_n \circ \phi_0.$$

Note that $\tilde{\Phi}_n$ is γ-Lipschitz since

$$\begin{aligned}
\|\tilde{\Phi}_n(y') - \tilde{\Phi}_n(y)\|_X &= \|\Phi_n \circ \phi_0(y') - \Phi_n \circ \phi_0(y)\|_X \leq \gamma \|\phi_0(y') - \phi_0(y)\|_{Y_n} \\
&= \gamma \|\sum_{j=1}^{n} (y'_j - y_j)\bar{v}_j\|_{Y_n} = \gamma \|y' - y\|_{\mathcal{Y}_n}.
\end{aligned}$$

In addition, $\tilde{\Phi}_n(B_{\mathcal{Y}_n}) = \Phi_n(B_{Y_n})$, and the proof is completed. □

2.1 Packing, covering and entropy numbers

Before going further, we recall in this section the well known concepts of packing, covering, and entropy numbers for compact sets $\mathcal{M}$, which we will use in our study of Lipschitz widths. The reader may find a more detailed exposition of those concepts in many books, see, for example, [4, 14, 12].

Minimal ε-covering number $N_\varepsilon(\mathcal{M})$ of a compact set $\mathcal{M} \subset X$:
A collection $\{g_1, \dots, g_m\} \subset X$ of elements of X is called an ε-covering of $\mathcal{M}$ if

$$\mathcal{M} \subset \bigcup_{j=1}^{m} B(g_j, \varepsilon), \quad \text{where} \quad B(g_j, \varepsilon) := \{f \in X : \|f - g_j\|_X \leq \varepsilon\}.$$

An ε-covering of $\mathcal{M}$ whose cardinality is minimal is called *minimal ε-covering* of $\mathcal{M}$. We denote by $N_\varepsilon(\mathcal{M})$ the cardinality of the minimal ε-covering of $\mathcal{M}$.

Minimal inner ε-covering number $\tilde{N}_\varepsilon(\mathcal{M})$ of a compact set $\mathcal{M} \subset X$:
It is defined exactly as $N_\varepsilon(\mathcal{M})$ but we additionally require that the centers $\{g_1, \dots, g_m\}$ of the covering are elements from $\mathcal{M}$.

Entropy numbers $\varepsilon_n(\mathcal{M})_X$ of a compact set $\mathcal{M} \subset X$:
For every fixed $n \geq 0$, the *entropy number* $\varepsilon_n(\mathcal{M})_X$ is the infimum of all $\varepsilon > 0$ for which 2^n balls with centers from X and radius ε cover $\mathcal{M}$. If we put the additional restriction that the centers of these balls are from $\mathcal{M}$, then we define the so called *inner entropy number* $\tilde{\varepsilon}_n(\mathcal{M})_X$. Formally, we write

$$\varepsilon_n(\mathcal{M})_X = \inf\{\varepsilon > 0 \;:\; \mathcal{M} \subset \bigcup_{j=1}^{2^n} B(g_j, \varepsilon),\; g_j \in X,\; j = 1, \dots, 2^n\},$$

$$\tilde{\varepsilon}_n(\mathcal{M})_X = \inf\{\varepsilon > 0 \;:\; \mathcal{M} \subset \bigcup_{j=1}^{2^n} B(h_j, \varepsilon),\; h_j \in \mathcal{M},\; j = 1, \dots, 2^n\}.$$

Maximal ε-packing number $P_\varepsilon(\mathcal{M})$ of a compact set $\mathcal{M} \subset X$:
A collection $\{f_1, \dots, f_\ell\} \subset \mathcal{M}$ of elements from $\mathcal{M}$ is called an ε-packing of $\mathcal{M}$ if

$$\min_{i \neq j} \|f_i - f_j\|_X > \varepsilon.$$

An ε-packing of $\mathcal{M}$ whose size is maximal is called *maximal ε-packing* of $\mathcal{M}$. We denote by $P_\varepsilon(\mathcal{M})$ the cardinality of the maximal ε-packing of $\mathcal{M}$.

We have the following inequalities for every $\varepsilon > 0$ and every compact set $\mathcal{M}$

$$P_\varepsilon(\mathcal{M}) \geq \tilde{N}_\varepsilon(\mathcal{M}) \geq P_{2\varepsilon}(\mathcal{M}), \tag{2.14}$$

$$\varepsilon_n(\mathcal{M})_X \leq \tilde{\varepsilon}_n(\mathcal{M})_X \leq 2\varepsilon_n(\mathcal{M})_X. \tag{2.15}$$

Remark 2.3. *Let us recall the classical relations between those concepts and compactness. We call the set $\mathcal{M}$ totally bounded if for every $\varepsilon > 0$ we have $N_\varepsilon(\mathcal{M}) < \infty$. This is equivalent to the fact that $\lim_{n\to\infty} \varepsilon_n(\mathcal{M})_X = 0$. Each compact set is totally bounded. Actually, a subset $\mathcal{M}$ of a Banach space is compact if and only if it is totally bounded and closed. The interested reader will find a detailed study on the topic in many books on functional analysis or metric topology.*

Remark 2.4. *In what follows later, we will use the fact that the Lipschitz widths and the entropy numbers are invariant with respect to translation, that is, for any $n \in \mathbb{N}, \gamma \geq 0$, and any $f \in X$ we have*

$$d_n^\gamma(\mathcal{K})_X = d_n^\gamma(\mathcal{K} - f)_X, \quad \varepsilon_n(\mathcal{K})_X = \varepsilon_n(\mathcal{K} - f)_X.$$

We want to mention the following remark that shows the behavior of the entropy numbers of an image of a γ-Lipschitz mapping.

Remark 2.5. *For the two normed spaces $(X_0, \|\cdot\|_{X_0})$ and $(X_1, \|\cdot\|_{X_1})$, and the γ-Lipschitz map $\Phi : (\mathcal{K}_0, \|\cdot\|_{X_0}) \to (\mathcal{K}_1, \|\cdot\|_{X_1})$, where $\mathcal{K}_0 \subset X_0$, $\mathcal{K}_1 \subset X_1$, we have that:*

- *If $\Phi(\mathcal{K}_0) = \mathcal{K}_1$, then the inner entropy numbers $\tilde{\varepsilon}_k(\mathcal{K}_1)_{X_1} \leq \gamma\tilde{\varepsilon}_k(\mathcal{K}_0)_{X_0}$ for $k = 1, 2, \ldots$. To see this, we take a minimal inner ε-covering number of $\mathcal{K}_0$ and look at its image by Φ.*
- *If $\Phi : (B_{Y_n}, \|\cdot\|_{Y_n}) \to (B_{Z_m}, \|\cdot\|_{Z_m})$ is a γ-Lipschitz map from the unit ball B_{Y_n} onto the unit ball B_{Z_m}, then $n \geq m$. This follows from the inequality above and the fact that (see e.g. (1.1.10) in [4]) for any unit ball B of any Banach space $(X_\ell, \|\cdot\|_{X_\ell})$ of dimension ℓ we have $4 \cdot 2^{-k/\ell} \geq \varepsilon_k(B)_{X_\ell} \geq 2^{-k/\ell}$, $k = 1, 2, \ldots$.*
- *If $\Phi(\mathcal{K}_0)$ approximates $\mathcal{K}_1$ with accuracy ε_2 and $A \subset \mathcal{K}_0$ approximates $\mathcal{K}_0$ with accuracy ε_1, then $\Phi(A)$ approximates $\mathcal{K}_1$ with accuracy $\gamma\varepsilon_1 + \varepsilon_2$. To see this, we take $f \in \mathcal{K}_1$, the corresponding $g \in \mathcal{K}_0$ such that $\|\Phi(g) - f\|_{X_1} \leq \varepsilon_2$ and $g_0 \in A$ such that $\|g_0 - g\|_{X_0} \leq \varepsilon_1$.*

2.2 Dependence of $d_n^\gamma(\mathcal{K})_X$ on γ

We start this section by proving the fact that the Lipschitz width $d_n^\gamma(\mathcal{K})_X$ is a continuous function of γ. To do that, we first prove the following lemma.

Lemma 2.6. *For every $n \geq 1$, every $\gamma > 0$, and every norm $\|\cdot\|_{Y_n}$ in $\mathbb{R}^n$, the fixed Lipschitz width $d^\gamma(\mathcal{K}, Y_n)_X$ satisfies the inequality*

$$\mathrm{rad}(\mathcal{K}) - \gamma \leq d^\gamma(\mathcal{K}, Y_n)_X \leq \mathrm{rad}(\mathcal{K}). \tag{2.16}$$

Proof: The right hand-side follows from (2.5) and Remark 2.1, (iii). To show the left hand-side inequality in (2.16), we notice that for any γ-Lipschitz map Φ, every $f \in \mathcal{K}$ and $y \in B_{Y_n}$ we have

$$\|f - \Phi(y)\|_X \geq \|f - \Phi(0)\|_X - \|\Phi(0) - \Phi(y)\|_X \geq \|f - \Phi(0)\|_X - \gamma,$$

since $\|\Phi(0) - \Phi(y)\|_X \leq \gamma \|y\|_{Y_n} \leq \gamma$. Therefore we obtain the inequality

$$\sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \Phi(y)\| \geq \sup_{f \in \mathcal{K}} \|f - \Phi(0)\| - \gamma,$$

which gives

$$d^\gamma(\mathcal{K}, Y_n)_X \geq \inf_{\Phi} \sup_{f \in \mathcal{K}} \|f - \Phi(0)\|_X - \gamma. \tag{2.17}$$

Note now that for every Φ,

$$\sup_{f \in \mathcal{K}} \|f - \Phi(0)\|_X \geq \inf_{g \in X} \sup_{f \in \mathcal{K}} \|f - g\|_X = \mathrm{rad}(\mathcal{K}),$$

and thus it follows from (2.17) that

$$d^\gamma(\mathcal{K}, Y_n)_X \geq \mathrm{rad}(\mathcal{K}) - \gamma,$$

and the proof is completed. □

Theorem 2.7. *For every compact subset $\mathcal{K} \subset X$ of a Banach space X and any $n \in \mathbb{N}$, the Lipschitz width $d_n^\gamma(\mathcal{K})_X$ is a continuous function of $\gamma \geq 0$.*

Proof: We first show the continuity of the Lipschitz width at $\gamma = 0$. It follows from Lemma 2.6 that

$$\mathrm{rad}(\mathcal{K}) - \gamma \leq d_n^\gamma(\mathcal{K})_X \leq \mathrm{rad}(\mathcal{K}).$$

We let $\gamma \to 0$ and obtain

$$\lim_{\gamma \to 0} d_n^\gamma(\mathcal{K})_X = \mathrm{rad}(\mathcal{K}) = d_n^0(\mathcal{K})_X,$$

which proves the continuity at $\gamma = 0$, see (2.5).

To show that the Lipschitz width is continuous for $\gamma > 0$, we fix $n \in \mathbb{N}$ and denote by

$$h(\gamma) := d_n^\gamma(\mathcal{K})_X.$$

According to Remark 2.1 (iv), $h(\gamma) < \infty$ for every $\gamma > 0$. Let us assume that h is not a continuous function. Then, there exist $\gamma_0 > 0$, $\delta > 0$, and a sequence of positive numbers $\varepsilon_k \to 0$, such that

$$h(\gamma_0 + \varepsilon_k) + \delta \leq h(\gamma_0 - \varepsilon_k), \quad \text{for every } k. \tag{2.18}$$

We fix $\varepsilon := \varepsilon_k < \gamma_0$. From the definition of Lipschitz widths, there exists a $(\gamma_0 + \varepsilon)$-Lipschitz map $\Phi_n : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$ such that

$$h(\gamma_0 + \varepsilon) \leq \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \Phi_n(y)\|_X \leq h(\gamma_0 + \varepsilon) + \varepsilon. \tag{2.19}$$

Now we define the mapping

$$\tilde{\Phi}_n := \xi \Phi_n, \quad \text{where} \quad \xi := \frac{\gamma_0 - \varepsilon}{\gamma_0 + \varepsilon} \quad \text{and} \quad 0 < \xi < 1.$$

Clearly, $\tilde{\Phi}_n$ is a $(\gamma_0 - \varepsilon)$-Lipschitz mapping, and therefore

$$\begin{aligned} h(\gamma_0 - \varepsilon) &\leq \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \tilde{\Phi}_n(y)\|_X = \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|\xi(f - \Phi_n(y)) + (1-\xi) f\|_X \\ &\leq \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} (\xi \|f - \Phi(y)\|_X + (1-\xi)\|f\|_X) \\ &\leq \xi \sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \Phi_n(y)\|_X + (1-\xi) \sup_{f \in \mathcal{K}} \|f\|_X \\ &\leq \xi(h(\gamma_0 + \varepsilon) + \varepsilon) + (1-\xi) C, \end{aligned}$$

where we have used (2.19) and the fact that $\sup_{f \in \mathcal{K}} \|f\|_X = C < \infty$ (since $\mathcal{K}$ is compact). The latter inequality and (2.18) give

$$h(\gamma_0 + \varepsilon) + \delta \leq h(\gamma_0 - \varepsilon) \leq \xi(h(\gamma_0 + \varepsilon) + \varepsilon) + (1-\xi) C$$

which is a contradiction for a sufficiently small $\varepsilon = \varepsilon_k$ since $\xi \to 1$ as $\varepsilon_k \to 0$. □

We finish the investigation of the behavior of the Lipschitz width with respect to γ with the following lemma.

Lemma 2.8. *For any $\mathcal{K} \subset X$, the set $\mathcal{K}$ is totally bounded iff for every $n \geq 1$*

$$\lim_{\gamma \to \infty} d_n^{\gamma}(\mathcal{K})_X = 0.$$

Proof: Assume that $\mathcal{K}$ is totally bounded. From the monotonicity of the Lipschitz width with respect to n, see Remark 2.1 (iii), it suffices to consider only the case $n = 1$. For $\delta > 0$, we fix a minimal delta covering $(f_j)_{j=1}^{\mathcal{N}_\delta(\mathcal{K})}$ of $\mathcal{K}$ and choose γ such that

$$2\gamma \geq \operatorname{diam} \mathcal{K} \cdot (\mathcal{N}_\delta(\mathcal{K}) - 1).$$

We consider the points

$$t_j := -1 + 2\frac{j-1}{N_\delta(\mathcal{K}) - 1}, \quad j = 1, \dots, \mathcal{N}_\delta(\mathcal{K}),$$

in the unit ball of $(\mathbb{R}, |.|)$, that is $([-1,1], |\cdot|)$, and define $\Phi : [-1,1] \to X$ as the continuous piecewise linear function such that

$$\Phi(t_j) = f_j, \quad j = 1, \dots, \mathcal{N}_\delta(\mathcal{K}).$$

Its Lipschitz constant is no more than

$$\max_{j=1,\dots,\mathcal{N}_\delta(\mathcal{K})-1} \frac{\|f_{j+1} - f_j\|_X}{|t_{j+1} - t_j|} \leq \frac{\operatorname{diam} \mathcal{K} \cdot (\mathcal{N}_\delta(\mathcal{K}) - 1)}{2} \leq \gamma.$$

and we have

$$\sup_{f\in\mathcal{K}}\inf_{y\in[-1,1]}\|f-\Phi(y)\|_X\leq\delta.$$

This gives

$$d_1^\gamma(\mathcal{K})_X\leq\delta,$$

and therefore $\lim_{\gamma\to\infty}d_1^\gamma(\mathcal{K})_X=0$.

To prove the converse, we take arbitrary $\epsilon>0$. Since $\lim_{\gamma\to\infty}d_n^\gamma(\mathcal{K})_X=0$, there exist a norm $\|\cdot\|_{Y_n}$ and a γ-Lipschitz map Φ_n such that

$$\sup_{f\in\mathcal{K}}\inf_{g\in B_{Y_n}}\|f-\Phi_n(g)\|_X<\epsilon/2. \tag{2.20}$$

We fix $\{g_1,\dots,g_N\}\subset B_{Y_n}$ such that $\bigcup_{j=1}^N B(g_j,\epsilon/2\gamma)\supset B_{Y_n}$ From (2.20) we infer that for every $f\in\mathcal{K}$ we can find $g_f\in B_{Y_n}$ such that $\|f-\Phi_n(g_f)\|_X\leq\epsilon/2$. Therefore, there exists j_0 such that $g_f\in B(g_{j_0},\epsilon/2\gamma)$, and then

$$\|f-\Phi_n(g_{j_0})\|_X\leq\|f-\Phi_n(g_f)\|_X+\|\Phi_n(g_f)-\Phi_n(g_{j_0})\|_X\leq\epsilon,$$

which gives that $N_\epsilon(\mathcal{K})\leq N<\infty$. Since ϵ is arbitrary, $\mathcal{K}$ is totally bounded. □

Remark 2.9. *It follows from Lemma* 2.8*, using Remark 2.1 (iii), that if $\mathcal{K}$ is not totally bounded then there exists $\delta>0$ such that*

$$\inf_{\gamma>0,n>0}d_n^\gamma(\mathcal{K})_X\geq\delta.$$

Remark 2.10. *Note that all statements in this paper are valid for sets $\mathcal{K}$ whose closures are compact rather than sets $\mathcal{K}$ that are compact. Therefore, since we work in Banach spaces, all statements are valid for $\mathcal{K}$ being only a totally bounded set rather than a compact set.*

2.3 Dependence of the Lipschitz width $d_n^\gamma(\mathcal{K})_X$ on n.

In this section, we discuss the behavior of the Lipschitz width with respect to n. The following Lemma holds.

Lemma 2.11. *Let $\mathcal{K}\subset X$ be a subset of a Banach space X. If for every $\epsilon>0$ there exist $n\in\mathbb{N}$ and $\gamma>0$ such that $d_n^\gamma(\mathcal{K})_X<\epsilon$, then $\mathcal{K}$ is totally bounded (i.e. its closure is compact).*

Proof: To prove the lemma, we fix $\eta>0$ and show that $\mathcal{K}$ is contained in the union of a finite collection of balls with radius η. It follows from the conditions of the lemma that we can find an integer n_0 and a parameter $\gamma>0$ such that

$$d_{n_0}^\gamma(\mathcal{K})_X<\eta/2.$$

Therefore there exists a norm $\|\cdot\|_{Y_{n_0}}$ in $\mathbb{R}^{n_0}$ and a γ-Lipschitz map $\Phi:(B_{Y_{n_0}},\|\cdot\|_{Y_{n_0}})\to X$ such that

$$\sup_{f\in\mathcal{K}}\inf_{y\in B_{Y_{n_0}}}\|f-\Phi(y)\|<\eta/2.$$

More precisely, for every $f\in\mathcal{K}$, we can find $y\in B_{Y_{n_0}}$ such that

$$\|f-\Phi(y)\|_X<\eta/2. \tag{2.21}$$

Let $\{y_j\}_{j=1}^N \subset B_{Y_{n_0}}$ be an $\eta/(2\gamma)$-covering for the compact set $B_{Y_{n_0}}$, that is

$$B_{Y_{n_0}} \subset \bigcup_{n=1}^N B(y_j, \eta/(2\gamma)),$$

and therefore for every $y \in B_{Y_{n_0}}$ we can find y_j, $j \in \{1, 2, \dots, N\}$, such that $\|y - y_j\|_{Y_{n_0}} \le \eta/(2\gamma)$. Thus we have

$$\|\Phi(y) - \Phi(y_j)\|_X \le \gamma \|y - y_j\|_{Y_{n_0}} \le \eta/2,$$

and

$$\Phi(B_{Y_{n_0}}) \subset \bigcup_{n=1}^N B(\Phi(y_j), \eta/2).$$

From the latter result and (2.21) it follows that $\mathcal{K} \subset \bigcup_{n=1}^N B(\phi(y_j), \eta)$, and the proof is completed. $\square$

Remark 2.12. *Note that Lemma 2.11 states in particular that a subset $\mathcal{K} \subset X$ of a Banach space X is totally bounded if there exists $\gamma > 0$ such that $\lim_{n\to\infty} d_n^\gamma(\mathcal{K})_X = 0$. The converse statement is also true, see Corollary 3.4 and Corollary 4.3.*

2.4 A single norm defines the Lipschitz width

In this section, we extend Lemma 2.2 and show that in the definition of Lipschitz width the infimum over all norms $\|\cdot\|_{Y_n}$ is achieved for some norm that satisfies (2.9). While this fact may not be very useful in practical applications, it has a certain theoretical merit. In our argument, we use the following version of Ascoli's theorem, whose proof can be found in [5], and which we state below.

Lemma 2.13. *Let (X, d) be a separable metric space and (Y, ρ) be a metric space for which every closed ball is compact. Let $F_j : X \to Y$ be a sequence of γ-Lipschitz maps for which there exists $a \in X$ and $b \in Y$ such that $F_j(a) = b$ for $j = 1, 2, \dots$. Then, there exists a subsequence F_{j_k}, $k \ge 1$, which is point-wise convergent to a function $F : X \to Y$ and F is γ-Lipschitz. If (X, d) is also compact, then the convergence is uniform.*

Now we are ready to state and prove the following fact.

Theorem 2.14. *For any $n \in \mathbb{N}$, any compact set $\mathcal{K} \subset X$, and any constant $\gamma > 0$ there is a norm $\|\cdot\|_Y$ on $\mathbb{R}^n$ satisfying (2.9) such that*

$$d_n^\gamma(\mathcal{K})_X = d_n^\gamma(\mathcal{K}, Y)_X.$$

Proof: It follows from Lemma 2.2 that we can find a sequence $(\Psi^j)_{j=1}^\infty$ of γ-Lipschitz maps $\Psi^j : (B_{\mathcal{Y}_j}, \|\cdot\|_{\mathcal{Y}_j}) \to X$, where the norms $\|.\|_{\mathcal{Y}_j}$ on $\mathbb{R}^n$ satisfy (2.9), such that

$$d_j := \sup_{f\in\mathcal{K}} \inf_{y\in B_{\mathcal{Y}_j}} \|f - \Psi^j(y)\|_X \to d_n^\gamma(\mathcal{K}) \quad \text{as} \quad j \to \infty.$$

There is a subsequence $\|.\|_{\mathcal{Y}_{j_k}}$ of the sequence of norms $\|.\|_{\mathcal{Y}_j}$ that converges pointwise on $\mathbb{R}^n$ and uniformly on $B_{\ell_\infty^n}$ to a norm $\|.\|_Y$ on $\mathbb{R}^n$ satisfying (2.9). Indeed, one can check that the functions $F_j : (\mathbb{R}^n, \|\cdot\|_{\ell_1^n}) \to \mathbb{R}$, defined as $F_j(y) := \|y\|_{\mathcal{Y}_j}$ satisfy $F_j(0) = 0$, and

$$|F_j(y') - F_j(y)| = |\|y'\|_{\mathcal{Y}_j} - \|y\|_{\mathcal{Y}_j}| \leq \|y' - y\|_{\mathcal{Y}_j} \leq \|y' - y\|_{\ell_1^n},$$

where we have used (2.9). Thus, the sequence $(F_j)_{j=1}^\infty$ satisfies the conditions of Lemma 2.13 with $\gamma = 1$, $a = b = 0$, and so we can find a subsequence F_{j_k} that converges point-wise on $\mathbb{R}^n$ and uniformly on $B_{\ell_\infty^n}$. In fact, the limit function F of this subsequence is a norm, which we denote by $\|\cdot\|_Y$. Clearly, this norm satisfies inequalities (2.9).

Now, passing to a subsequence, we will assume that $\|.\|_{\mathcal{Y}_j}$ converge uniformly on $B_{\ell_\infty^n}$ to the function $\|.\|_Y$. Thus, there is $j_0 \in \mathbb{N}$ such that for any $j \geq j_0$ there is ε_j with the properties $0 < \varepsilon_j < 1$, $\lim_{j\to\infty} \varepsilon_j = 0$ and

$$\|y\|_{\mathcal{Y}_j} - \varepsilon_j \leq \|y\|_Y \leq \|y\|_{\mathcal{Y}_j} + \varepsilon_j, \quad \text{for all} \quad \|y\|_{\ell_\infty^n} \leq 1.$$

For example, we can take

$$\varepsilon_j := \sup_{y:\|y\|_{\ell_\infty^n} \leq 1} |\|y\|_{\mathcal{Y}_j} - \|y\|_Y|,$$

and j_0 big enough. Since $B_{\mathcal{Y}_j} \subset B_{\ell_\infty^n}$, $j = 1, 2, \ldots$, and $B_Y \subset B_{\ell_\infty^n}$, we have for all $y \in B_{\mathcal{Y}_j} \cup B_Y$

$$\|y\|_{\mathcal{Y}_j} - \varepsilon_j \leq \|y\|_Y \leq \|y\|_{\mathcal{Y}_j} + \varepsilon_j. \tag{2.22}$$

The latter inequality gives that for $y \in B_Y$ we have

$$\|y\|_{\mathcal{Y}_j} \leq 1 + \varepsilon_j \quad \Rightarrow \quad y \in (1 + \varepsilon_j) B_{\mathcal{Y}_j},$$

and so

$$B_Y \subset (1 + \varepsilon_j) B_{\mathcal{Y}_j}. \tag{2.23}$$

Next, let $j \geq j_0$. For any y with $\|y\|_{\mathcal{Y}_j} \leq 1 - \varepsilon_j < 1$, we have from (2.22) that

$$\|y\|_Y \leq 1 \quad \Rightarrow \quad y \in B_Y,$$

and therefore

$$(1 - \varepsilon_j)^{-1} B_{\mathcal{Y}_j} \subset B_Y. \tag{2.24}$$

It follows from (2.23) and (2.24) that

$$(1 - \varepsilon_j)^{-1} B_{\mathcal{Y}_j} \subset B_Y \subset (1 + \varepsilon_j) B_{\mathcal{Y}_j}, \quad j \geq j_0. \tag{2.25}$$

Let us now define the mapping $\tilde{\Psi}^j : (1 + \varepsilon_j) B_{\mathcal{Y}_j} \to X$, as

$$\tilde{\Psi}^j(y) := \Psi^j((1 + \varepsilon_j)^{-1} y).$$

Note that

$$\|\tilde{\Psi}^j(y') - \tilde{\Psi}^j(y)\|_X \leq \frac{\gamma}{1 + \varepsilon_j} \|y' - y\|_{\mathcal{Y}_j} < \gamma \|y' - y\|_{\mathcal{Y}_j}, \quad y', y \in (1 + \varepsilon_j) B_{\mathcal{Y}_j},$$

where we have used that Ψ^j is γ-Lipschitz. We denote by $\bar{\Psi}^j$ the restriction of $\tilde{\Psi}^j$ on B_Y, see (2.25).

Now we fix $f \in \mathcal{K}$ and $j \geq j_0$. For every $\varepsilon > 0$, we can find $y = y(f,j,\varepsilon) \in B_{\mathcal{Y}_j}$ such that $\|f - \Psi^j(y)\|_X < d_j + \varepsilon$. We set

$$z := y/(1-\varepsilon_j) \in (1-\varepsilon_j)^{-1} B_{\mathcal{Y}_j} \subset B_Y,$$

and observe that

$$\begin{aligned}
\inf_{x \in B_Y} \|f - \bar{\Psi}^j(x)\|_X &\leq \|f - \bar{\Psi}^j(z)\|_X = \|f - \tilde{\Psi}^j(z)\|_X = \|f - \Psi^j((1+\varepsilon_j)^{-1}z)\|_X \\
&= \|f - \Psi^j((1-\varepsilon_j^2)^{-1}y)\|_X \leq \|f - \Psi^j(y)\|_X + \|\Psi^j(y) - \Psi^j((1-\varepsilon_j^2)^{-1}y)\|_X \\
&< d_j + \varepsilon + \gamma \frac{\varepsilon_j^2}{1-\varepsilon_j^2}\|y\|_{\mathcal{Y}_j} \leq d_j + \varepsilon + \gamma \frac{\varepsilon_j^2}{1-\varepsilon_j^2}.
\end{aligned}$$

By letting $\varepsilon \to 0$ and taking supremum over $f \in \mathcal{K}$, we obtain

$$d^\gamma(\mathcal{K}, Y)_X \leq \sup_{f \in \mathcal{K}} \inf_{x \in B_Y} \|f - \bar{\Psi}^j(x)\|_X \leq d_j + \gamma \frac{\varepsilon_j^2}{1-\varepsilon_j^2}.$$

Since $d_j \to d_n^\gamma(\mathcal{K})_X$ and $\varepsilon_j \to 0$ as $j \to \infty$, we derive that $d^\gamma(\mathcal{K}, Y)_X \leq d_n^\gamma(\mathcal{K})_X$, and the proof is completed. □

3 Lipschitz widths and entropy numbers

In this section, we study the relation between the Lipschitz widths $d_n^\gamma(\mathcal{K})_X$ and the entropy numbers $\varepsilon_n(\mathcal{K})_X$ of a compact set $\mathcal{K}$. We look at widely used in the literature assumptions on the asymptotic decay of $\varepsilon_n(\mathcal{K})_X$ and show how it relates to $d_n^\gamma(\mathcal{K})_X$. This is rather technical, so we state here the main result in the section, which is in fact a corollary of Theorem 3.3, see (3.2) with $k = 1$, Theorem 3.9. and Remark 2.1 (iii).

Theorem 3.1. *Let $\mathcal{K} \subset X$ be a compact subset of a Banach space X, $n \in \mathbb{N}$, and $d_n^\gamma(\mathcal{K})_X$ be the Lipschitz width for $\mathcal{K}$ with Lipschitz constant $\gamma \geq 2\,\mathrm{rad}(\mathcal{K})$. Then the following holds:*

(i) For $\alpha > 0$, $\beta \in \mathbb{R}$, we have

$$\varepsilon_n(\mathcal{K})_X \leq C \frac{[\log_2 n]^\beta}{n^\alpha}, \quad n = 1, 2, \dots, \quad \Rightarrow \quad d_n^\gamma(\mathcal{K})_X \leq C' \frac{[\log_2 n]^\beta}{n^\alpha}, \quad n = 1, 2, \dots,$$

$$\varepsilon_n(\mathcal{K})_X \geq C \frac{[\log_2 n]^\beta}{n^\alpha}, \quad n = 1, 2, \dots, \quad \Rightarrow \quad d_n^\gamma(\mathcal{K})_X \geq C' \frac{[\log_2 n]^\beta}{n^\alpha [\log_2 n]^\alpha}, \quad , n = 1, 2, \dots.$$

(ii) For $\alpha > 0$, we have

$$\varepsilon_n(\mathcal{K})_X \asymp \frac{1}{[\log_2 n]^\alpha}, \quad n = 1, 2, \dots, \quad \Rightarrow \quad d_n^\gamma(\mathcal{K})_X \asymp \frac{1}{[\log_2 n]^\alpha}, \quad n = 1, 2, \dots \quad . \tag{3.1}$$

(iii) For $0 < \alpha < 1$, we have

$$\varepsilon_n(\mathcal{K})_X \leq C 2^{-cn^\alpha}, \quad n = 1, 2, \dots, \quad \Rightarrow \quad d_n^\gamma(\mathcal{K})_X \leq C' 2^{-cn^\alpha}, \quad n = 1, 2, \dots,$$

$$\varepsilon_n(\mathcal{K})_X \geq C 2^{-cn^\alpha}, \quad n = 1, 2, \dots, \quad \Rightarrow \quad d_n^\gamma(\mathcal{K})_X \geq C' 2^{-c' n^{\alpha/(1-\alpha)}}, \quad n = 1, 2, \dots.$$

We conclude our study with various examples showing the sharpness of our estimates.

3.1 Lipschitz widths are smaller than entropy numbers

We start with the construction of a particular Lipschitz function that can be viewed as a sum of 'bumps', each one supported on a closed ball from a Banach space Y. We use this function to show that the Lipschitz widths of a compact set $\mathcal{K} \subset X$ are smaller than the entropy numbers of that set.

Lemma 3.2. *If $(B^j) := (B(y_j, \rho_j))$ is a family of disjoint open balls in a Banach space Y, then for every sequence (φ_j) of γ_j-Lipschitz mappings $\varphi_j : Y \to X$, $j = 1, 2, \ldots$, with the property that $\varphi_j \equiv 0$ on the complement of B^j, the mapping $\Phi : Y \to X$, defined as $\Phi_0 = \sum_j \varphi_j$ is a Lipschitz map with Lipschitz constant $\sup_j \gamma_j$. In particular, for any sequence (f_j) of elements $f_j \in X$ with $\|f_j\|_X = 1$ and any sequence (σ_j) of real numbers, the mappings $\phi_j : Y \to X$, $j = 1, 2, \ldots$, defined as*

$$\phi_j(y) = \sigma_j \left(1 - \frac{\|y_j - y\|_Y}{\rho_j}\right)_+ \cdot f_j, \quad \text{where} \quad (t)_+ := \max\{0, t\}, \ t \in \mathbb{R},$$

are $|\sigma_j|/\rho_j$-Lipschitz mappings. Their sum, the mapping $\Phi := \sum_j \phi_j$, is a Lipschitz mapping with Lipschitz constant $\sup_j |\sigma_j|/\rho_j$ and $\Phi(y_j) = \sigma_j f_j$, $j = 1, 2, \ldots$.

Proof: The proof follows from the observation that the sum of Lipschitz mappings with disjoint supports is again a Lipschitz mapping with a Lipschitz constant bounded by the supremum of the Lipschitz constants of these mappings and we omit the details. □

We use the Lipschitz function Φ, constructed in Lemma 3.2 to prove the following theorem.

Theorem 3.3. *For any compact subset $\mathcal{K} \subset X$ of a Banach space X and any $n \geq 1$ we have that*

$$d_n^{2^k \operatorname{rad}(\mathcal{K})}(\mathcal{K})_X \leq \varepsilon_{kn}(\mathcal{K})_X, \quad k = 1, 2, \ldots. \tag{3.2}$$

In particular, when $k = n$, we have

$$d_n^{2^n \operatorname{rad}(\mathcal{K})}(\mathcal{K})_X \leq \varepsilon_{n^2}(\mathcal{K})_X, \quad n = 1, 2, \ldots. \tag{3.3}$$

Proof: We fix $k \in N$. Let $\mathcal{K} \subset X$ be a compact set in a Banach space X, let $\eta > 0$, and let

$$\mathcal{X}_{kn} := \{f'_1, \ldots, f'_{2^{kn}}\} \subset \mathcal{K}$$

be the set such that for every $f \in \mathcal{K}$ we can find $f'_j \in \mathcal{X}_{kn}$ such that

$$\|f - f'_j\|_X \leq \varepsilon_{kn}(\mathcal{K})_X + \eta. \tag{3.4}$$

Since $\mathcal{K}$ is bounded, we can assume that $\mathcal{K} \subset B_X(0, r)$ for some $r > 0$. Let us divide the unit ball $(B_n, \|\cdot\|_{\ell_\infty^n}) = [-1, 1]^n \subset \mathbb{R}^n$ into 2^{kn} non-overlapping open balls B^j, each of side length 2^{1-k}. Let us denote by y_j the center of B^j and define a map $\phi_j : \mathbb{R}^n \to X$ as

$$\phi_j(y) = \left(1 - 2^k \|y_j - y\|_{\ell_\infty^n}\right)_+ \cdot f'_j \in X, \quad j = 1, \ldots, 2^{kn},$$

and

$$\Phi := \sum_{j=1}^{2^{kn}} \phi_j.$$

We apply Lemma 3.2 with $\sigma_j = \|f'_j\|_X$, $f_j = \frac{1}{\|f'_j\|_X} f'_j$, $\rho_j = 2^{-k}$, $Y = (\mathbb{R}^n, \|\cdot\|_{\ell_\infty^n})$ and conclude that $\Phi : Y \to X$ is a map with Lipschitz constant $\gamma := 2^k \max_j \|f'_j\|_X \leq 2^k r$, and $\Phi(y_j) = f'_j$. Therefore, we have $d_n^{2^k r}(\mathcal{K})_X \leq \varepsilon_{kn}(\mathcal{K})_X + \eta$, and taking $\eta \to 0$, we obtain

$$d_n^{2^k r}(\mathcal{K})_X \leq \varepsilon_{kn}(\mathcal{K})_X, \quad n \geq 1. \tag{3.5}$$

Now, for any $\varepsilon > 0$ we can find $g = g(\varepsilon) \in X$ such that

$$\sup_{f \in \mathcal{K}} \|f - g\|_X < \operatorname{rad}(\mathcal{K}) + \varepsilon 2^{-k},$$

We apply (3.5) for the set $(\mathcal{K} - g)$ with $r = \operatorname{rad}(\mathcal{K}) + \varepsilon 2^{-k}$ and using Remark 2.4, we arrive at

$$d_n^{2^k \operatorname{rad}(\mathcal{K}) + \varepsilon}(\mathcal{K})_X \leq \varepsilon_{kn}(\mathcal{K})_X.$$

The statement (3.2) of the theorem is obtained from the latter inequality using the continuity of the Lipschitz width $d_n^\gamma(\mathcal{K})_X$ with respect to γ, see Theorem 2.7. □

Corollary 3.4. *For every compact subset $\mathcal{K} \subset X$ of a Banach space X and every $\gamma \geq 2\operatorname{rad}(\mathcal{K})$ we have*

$$\lim_{n \to \infty} d_n^\gamma(\mathcal{K})_X = 0.$$

Proof: This follows from Theorem 3.3, Remark 2.1 (iii) and the fact that $\lim_{n\to\infty} \varepsilon_n(\mathcal{K})_X = 0$ for compact sets $\mathcal{K}$, see Remark 2.3. □

Example 3.5. We want to point out that for some cases of $\mathcal{K}$ and values of n the estimate (3.2) in Theorem 3.3 cannot be improved. We consider the Hilbert space H which we identify with the sequence space

$$\ell_2 := \{x = (x_1, \dots, x_j, \dots) : \ \|x\|_{\ell_2}^2 = \|x\|_H^2 = \sum_{j=1}^\infty x_j^2 < \infty, \ x_j \in \mathbb{R}\}.$$

For each $n = 1, 2, \dots,$ we construct the compact set $\mathcal{K}_n$,

$$\mathcal{K}_n := \{e_1, e_2, \dots, e_{2^n}, e_{2^n+1}\} \subset \ell_2,$$

where (e_j) is the standard basis in ℓ_2, that is, all coordinate components of e_j are 0's, except the j-th, which is 1. Then we have

$$\tilde{\varepsilon}_n(\mathcal{K}_n)_H \leq 3 d_{n/7}^{2\operatorname{diam}(\mathcal{K}_n)}(\mathcal{K}_n)_H.$$

Indeed, since $\|e_i - e_j\|_H = \sqrt{2}$, $i \neq j$, it follows that $\tilde{\varepsilon}_k(\mathcal{K}_n)_H = \sqrt{2}$, $k \leq n$. Now suppose that we have $d_s^\gamma(\mathcal{K}_n)_H < \sqrt{2}/3$ for some s and γ. This means that there exists a norm $\|\cdot\|_{Y_s}$ on $\mathbb{R}^s$ and a γ-Lipschitz map ϕ, $\phi : (B_{Y_s}, \|\cdot\|_{Y_s}) \to H$, defined on the unit ball B_{Y_s}, with the property that $\|\phi(y^j) - e_j\|_H < \sqrt{2}/3$, $j = 1, \dots, 2^n + 1$, $y^j \in B_{Y_s}$, $j = 1, \dots, 2^n + 1$. Since for $i \neq j$,

$$\gamma \|y^j - y^i\|_{Y_s} \geq \|\phi(y^j) - \phi(y^i)\|_H = \|(\phi(y^j) - e_j) + (e_j - e_i) + (e_i - \phi(y^i))\|_H > \sqrt{2} - 2\sqrt{2}/3 = \sqrt{2}/3,$$

we have that $\{y^j\}_{j=1}^{2^n+1}$ is $\sqrt{2}/(3\gamma)$ packing of B_{Y_s}. Using (2.14), we obtain that

$$\widetilde{\mathcal{N}}_{\sqrt{2}/(6\gamma)}(B_{Y_s}) \geq 2^n + 1 \quad \Rightarrow \quad \tilde{\varepsilon}_n(B_{Y_s})_H \geq \sqrt{2}/(6\gamma).$$

On the other hand, it follows from [4] that $4 \cdot 2^{-n/s} \geq \varepsilon_n(B_{Y_s})_H \geq 2^{-1}\tilde{\varepsilon}_n(B_{Y_s})_H$, and therefore,

$$\sqrt{2}/(12\gamma) \leq 4 \cdot 2^{-n/s}.$$

Thus, for any pair (γ, s) such that $\sqrt{2}/(12\gamma) > 4 \cdot 2^{-n/s}$ we get $3d_s^\gamma(\mathcal{K}_n)_H \geq \sqrt{2} = \tilde{\varepsilon}_n(\mathcal{K}_n)_H$. This holds, for example, when $\gamma = 2\,\mathrm{diam}(\mathcal{K}_n) = 2\sqrt{2}$ and $s = n/7$.

3.2 Estimates for Lipschitz widths from below

We start this section with a lower bound on the Lipschitz constant γ in $d_n^\gamma(\mathcal{K})_X$. The following proposition holds.

Proposition 3.6. *If $d_n^\gamma(\mathcal{K})_X < \varepsilon$ for a compact subset $\mathcal{K} \subset X$ of a Banach space X, then*

$$\gamma \geq \frac{1}{3}\varepsilon N_{2\varepsilon}^{1/n}(\mathcal{K}), \tag{3.6}$$

where $N_\varepsilon(\mathcal{K})$ is the ε-covering number of $\mathcal{K}$. In particular, if $d_n^\gamma(B_{Z_m})_X < \varepsilon$, then

$$\gamma \geq \frac{1}{3}2^{-m/n}\varepsilon^{1-m/n}. \tag{3.7}$$

Proof: If $d_n^\gamma(\mathcal{K}) < \varepsilon$, then there is a γ-Lipschitz map Φ and a norm $\|\cdot\|_{Y_n}$, $\Phi : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$ such that $\Phi(B_{Y_n})$ approximates $\mathcal{K}$ up to accuracy ε. Let us consider $\Phi(B_{Y_n})$ and let $\{y_j\}_{j=1}^N \subset B_{Y_n}$ be such that $\{\Phi(y_j)\}_{j=1}^N$ is a maximal ε-packing of $\Phi(B_{Y_n})$. Then, we have

$$\varepsilon < \|\Phi(y_j) - \Phi(y_{j'})\|_X \leq \gamma\|y_j - y_{j'}\|_{Y_n},$$

and thus

$$\|y_j - y_{j'}\|_{Y_n} > \varepsilon\gamma^{-1}, \quad j \neq j', \quad j, j' = 1, \ldots, N.$$

Therefore, see e.g. [12, Chp. 15 Prop. 1.3],

$$N \leq P_{\varepsilon\gamma^{-1}}(B_{Y_n}) \leq 3^n(\varepsilon\gamma^{-1})^{-n} = \left(\frac{3}{\varepsilon}\right)^n \gamma^n. \tag{3.8}$$

For every $z \in \mathcal{K}$, we can find $\Phi(y)$, $y \in B_{Y_n}$ such that $\|z - \Phi(y)\|_X < \varepsilon$ since $\Phi(B_{Y_n})$ approximates $\mathcal{K}$ up to accuracy ε. Since the set $\{\Phi(y), \Phi(y_1), \ldots, \Phi(y_N)\}$ is not an ε-packing for $\Phi(B_{Y_n})$, there is index j_0, $1 \leq j_0 \leq N$, such that $\|\Phi(y) - \Phi(y_{j_0})\|_X \leq \varepsilon$. Then,

$$\|z - \Phi(y_{j_0})\|_X \leq \|z - \Phi(y)\|_X + \|\Phi(y) - \Phi(y_{j_0})\|_X < 2\varepsilon,$$

and thus $\{\Phi(y_j)\}_{j=1}^N$ is a 2ε-covering of $\mathcal{K}$, which gives

$$N \geq N_{2\varepsilon}(\mathcal{K}).$$

Combining the latter estimate with (3.8) gives (3.6). In particular, when $\mathcal{K} = B_{Z_m}$, we know that

$$N_{2\varepsilon}(B_{Z_m}) \geq (2\varepsilon)^{-m},$$

and therefore we obtain (3.7). The proof is completed. □

Lemma 3.7. *Let $\mathcal{K} \subset X$ be a compact set and $\gamma > 0$ be a fixed constant. If there is $n > n_0$, $n_0 = n_0(c_0, \alpha, \beta)$ such that*

$$d_n^\gamma(\mathcal{K})_X < c_0 \frac{[\log_2 n]^\beta}{n^\alpha}, \quad \textit{with} \quad \alpha > 0, \quad \textit{and} \quad \beta \in \mathbb{R},$$

then

$$\varepsilon_m(\mathcal{K})_X < C \frac{[\log_2 m]^{\alpha+\beta}}{m^\alpha}, \quad \textit{with } m = cn \log_2 n, \tag{3.9}$$

where C, c are fixed constants, depending only on γ, c_0, α and β.

Proof: We use Proposition 3.6 with $\varepsilon = c_0[\log_2 n]^\beta n^{-\alpha}$ to obtain that

$$N_{2\varepsilon}(\mathcal{K}) \le \left(\frac{3\gamma}{\varepsilon}\right)^n = (3\gamma c_0^{-1}[\log_2 n]^{-\beta} n^\alpha)^n < 2^{n(\log_2(3\gamma c_0^{-1}) + \alpha \log_2 n - \beta \log_2(\log_2 n))} < 2^{cn \log_2 n},$$

and therefore

$$\varepsilon_{cn \log_2 n}(\mathcal{K})_X \le 2c_0[\log_2 n]^\beta n^{-\alpha}.$$

If we set $m = cn \log_2 n > cn$, then $n = m/c \log_2 n$ and we get

$$\varepsilon_m(\mathcal{K})_X \le 2c_0[\log_2 n]^\beta [m/c \log_2 n]^{-\alpha} = 2c_0 c^\alpha m^{-\alpha} [\log_2 n]^{\beta+\alpha}. \tag{3.10}$$

Since

$$\log_2 m = \log_2 c + \log_2 n + \log_2 \log_2 n,$$

for n sufficiently big we have

$$2^{-1} \log_2 n < \log_2 m < 3 \log_2 n,$$

and the statement follows from (3.10). □

Lemma 3.7 is similar to the classical Carl's inequalities [3], traditionally used to provide lower bounds. However, there is an important difference. Note that Lemma 3.7 works for each n separately, whenever the Carl's inequality requires an assumption for all $j \le n$. On the other hand the Carl's inequality gives the upper bound for ε_n not ε_m.

Next, we continue with a series of results presenting lower bounds for the Lipschitz widths of compact sets, provided we have information about the entropy numbers of these sets. We start with a natural consequence of Proposition 3.6.

Proposition 3.8. *Let $\mathcal{K} \subset X$ be a compact set and let*

$$\varepsilon_n(\mathcal{K})_X > \eta_n, \quad n = 1, 2, \ldots,$$

where $(\eta_n)_{n=1}^\infty$ is a sequences of real numbers decreasing to zero. Let for some $m \in \mathbb{N}$ and some $\delta > 0$

$$d_m^\gamma(\mathcal{K})_X < \delta.$$

Then we have

$$\eta_{m \log_2(3\gamma\delta^{-1})} < 2\delta. \tag{3.11}$$

Proof: We apply Proposition 3.6 with $\varepsilon = \delta$ and obtain

$$N_{2\delta}(\mathcal{K}) \leq \left(\frac{3\gamma}{\delta}\right)^m = 2^{m\log_2(3\gamma\delta^{-1})}.$$

Using our assumptions and the definition of entropy numbers, we derive

$$2\delta \geq \varepsilon_{m\log_2(3\gamma\delta^{-1})}(\mathcal{K})_X > \eta_{m\log_2(3\gamma\delta^{-1})}.$$

□

The next theorem discusses lower bounds of the Lipschitz widths $d_n^\gamma(\mathcal{K})_X$ in the case when $\gamma > 0$ is a fixed constant.

Theorem 3.9. *For any compact set $\mathcal{K} \subset X$ the following holds:*

(i) If for some constants $c_1 > 0, \alpha > 0$ and $\beta \in \mathbb{R}$ we have

$$\varepsilon_n(\mathcal{K})_X > c_1 \frac{(\log_2 n)^\beta}{n^\alpha}, \quad n = 1, 2, \ldots,$$

then for each $\gamma > 0$ there exists a constant $C > 0$ such that

$$d_n^\gamma(\mathcal{K})_X \geq C \frac{(\log_2 n)^{\beta-\alpha}}{n^\alpha}, \quad n = 1, 2, \ldots. \tag{3.12}$$

(ii) If for some constants $c_1 > 0, \alpha > 0$ we have

$$\varepsilon_n(\mathcal{K})_X > c_1(\log_2 n)^{-\alpha}, \quad n = 1, 2, \ldots,$$

then for each $\gamma > 0$ there exists a constant C such that

$$d_n^\gamma(\mathcal{K})_X \geq C(\log_2 n)^{-\alpha}, \quad n = 1, 2, \ldots. \tag{3.13}$$

(iii) If for some constants $c_1, c > 0$ and $1 > \alpha > 0$ we have

$$\varepsilon_n(\mathcal{K})_X > c_1 2^{-cn^\alpha}, \quad n = 1, 2, \ldots,$$

then for each $\gamma \geq 2\,\mathrm{rad}(\mathcal{K})$ we have

$$d_n^\gamma(\mathcal{K})_X \geq C 2^{-c_2 n^{\alpha/(1-\alpha)}}, \quad n = 1, 2, \ldots, \tag{3.14}$$

where $C, c_2 > 0$ are constants depending on γ, c, and α.

Proof: We prove (i) by contradiction. If (3.12) does not hold for some constant C, then there exists a strictly increasing sequence of integers $(n_k)_{k=1}^\infty$, such that

$$a_k := \frac{d_{n_k}^\gamma(\mathcal{K})_X n_k^\alpha}{(\log_2 n_k)^{\beta-\alpha}} \to 0 \quad \text{as} \quad k \to \infty.$$

Thus, we can write

$$d_{n_k}^\gamma(\mathcal{K})_X = \frac{a_k\,[\log_2 n_k]^{\beta-\alpha}}{n_k^\alpha} < \frac{2a_k\,[\log_2 n_k]^{\beta-\alpha}}{n_k^\alpha} =: \delta_k \text{ for } k = 1, 2, \ldots. \tag{3.15}$$

Now we apply Proposition 3.8 with $\eta_n = c_1 \frac{(\log_2 n)^\beta}{n^\alpha}$ and obtain

$$c_1 \left[\log_2(n_k \log_2(3\gamma\delta_k^{-1}))\right]^\beta {n_k}^{-\alpha} \left[\log_2(3\gamma\delta_k^{-1})\right]^{-\alpha} \leq 4 \frac{a_k \left[\log_2 n_k\right]^{\beta-\alpha}}{{n_k}^\alpha},$$

which we rewrite as

$$\left[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))\right]^\beta \left[\log_2(3\gamma\delta_k^{-1})\right]^{-\alpha} \leq C_1 a_k \left[\log_2 n_k\right]^{\beta-\alpha}, \quad \text{where } C_1 = 4/c_1. \tag{3.16}$$

Observe that

$$\log_2(3\gamma\delta_k^{-1}) = \log_2(1.5\gamma) + \log_2 a_k^{-1} + \alpha \log_2 n_k + (\alpha - \beta)\log_2(\log_2 n_k),$$

and therefore for k big enough we obtain

$$\log_2(3\gamma\delta_k^{-1}) \leq 2\left[\log_2(a_k^{-1}) + \alpha \log_2 n_k\right]. \tag{3.17}$$

The latter inequality and (3.16) give

$$2^{-\alpha}\left[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))\right]^\beta \left[\log_2(a_k^{-1}) + \alpha \log_2 n_k\right]^{-\alpha} \leq C_1 a_k \left[\log_2 n_k\right]^{\beta-\alpha},$$

which is equivalent to

$$a_k^{-1}\left[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))\right]^\beta \leq 2^\alpha C_1 \left[\frac{\log_2(a_k^{-1})}{\log_2 n_k} + \alpha\right]^\alpha \left[\log_2 n_k\right]^\beta. \tag{3.18}$$

Note that since $\delta_k \to 0$ as $k \to 0$, we have that for k big enough $\log_2(\log_2(3\gamma\delta_k^{-1})) > 0$. Now we consider several cases.
Case 1: $\beta \geq 0$. In this case we have for k big enough

$$\left[\log_2 n_k\right]^\beta \leq \left[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))\right]^\beta,$$

and therefore it follows from (3.18) that

$$a_k^{-1} \leq 2^\alpha C_1 \left[\frac{\log_2(a_k^{-1})}{\log_2 n_k} + \alpha\right]^\alpha < C[\log_2(a_k^{-1})]^\alpha,$$

which contradicts the fact that $a_k \to 0$ (and thus $a_k^{-1} \to \infty$).
Case 2: $\beta < 0$. In this case we have

$$\left[\log_2 n_k + \log_2(3\gamma\delta_k^{-1})\right]^\beta < \left[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))\right]^\beta, \tag{3.19}$$

and therefore it follows from (3.18) that

$$a_k^{-1}\left[\log_2 n_k + \log_2(3\gamma\delta_k^{-1})\right]^\beta \leq 2^\alpha C_1 \left[\frac{\log_2(a_k^{-1})}{\log_2 n_k} + \alpha\right]^\alpha \left[\log_2 n_k\right]^\beta.$$

This gives, using (3.17)

$$\begin{aligned} a_k^{-1} &\leq 2^\alpha C_1 \left[\frac{\log_2(a_k^{-1})}{\log_2 n_k} + \alpha\right]^\alpha \left[1 + \frac{\log_2(3\gamma\delta_k^{-1})}{\log_2 n_k}\right]^{-\beta} \\ &\leq 2^\alpha C_1 \left[\frac{\log_2(a_k^{-1})}{\log_2 n_k} + \alpha\right]^\alpha \left[1 + 2\alpha + 2\frac{\log_2(a_k^{-1})}{\log_2 n_k}\right]^{-\beta} < C[\log_2(a_k^{-1})]^{\alpha-\beta}, \end{aligned}$$

which also contradicts the fact that $a_k \to 0$ (and thus $a_k^{-1} \to \infty$).

To prove (ii), we repeat the argument for (i), namely, we assume that (ii) does not hold. Therefore there exists a strictly increasing sequence of integers $(n_k)_{k=1}^\infty$, such that

$$b_k := d_{n_k}^\gamma(\mathcal{K})_X[\log_2 n_k]^\alpha \to 0 \quad \text{as} \quad k \to \infty.$$

We write

$$d_{n_k}^\gamma(\mathcal{K})_X = b_k[\log_2 n_k]^{-\alpha} < 2b_k[\log_2 n_k]^{-\alpha} =: \delta_k \text{ for } k = 1, 2, \dots, \tag{3.20}$$

and use Proposition 3.8 with $\eta_n = c_1(\log_2 n)^{-\alpha}$ to derive

$$c_1 \left[\log_2(n_k \log_2(3\gamma\delta_k^{-1}))\right]^{-\alpha} \leq 4b_k[\log_2 n_k]^{-\alpha}.$$

The latter inequality is equivalent to

$$[\log_2 n_k + \log_2(\log_2(3\gamma\delta_k^{-1}))]^{-\alpha} \leq C_1 b_k(\log_2 n_k)^{-\alpha}, \; C_1 := 4/c_1, \tag{3.21}$$

which, after using (3.19) with $\beta = -\alpha$ gives

$$[\log_2 n_k + \log_2(3\gamma\delta_k^{-1})]^{-\alpha} \leq C_1 b_k \left[\log_2 n_k\right]^{-\alpha}.$$

We continue by writing the above inequality as

$$b_k^{-1} \leq C_1 \left[1 + \frac{\log_2(3\gamma\delta_k^{-1})}{\log_2 n_k}\right]^\alpha \leq \left[1 + 2\alpha + 2\frac{\log_2(b_k^{-1})}{\log_2 n_k}\right]^\alpha \leq C[\log_2(b_k^{-1})]^\alpha,$$

where we have used (3.17). The latter inequality contradicts the fact that b_k tends to zero, and the proof of (ii) is completed.

We now prove (iii). To simplify the notation, we denote by $d_n := d_n^\gamma(\mathcal{K})_X$ and observe that, according to Corollary 3.4, $d_n \to 0$ for $n \to \infty$ when $\gamma \geq \text{rad}(\mathcal{K})$. We use Proposition 3.8 with $\delta = 2d_n$ and $\eta_n = c_1 2^{-cn^\alpha}$ to obtain the inequality

$$4d_n \geq c_1 2^{-c[n \log_2(3\gamma d_n^{-1})]^\alpha}, \tag{3.22}$$

which can be rewritten as

$$2^{c[n \log_2(3\gamma d_n^{-1})]^\alpha} \geq \frac{c_1}{4} d_n^{-1} \quad \Leftrightarrow \quad 2^{c[n(\log_2 \xi_n)]^\alpha} \geq \frac{c_1}{12\gamma}\xi_n =: A\xi_n, \text{ where } \xi_n := 3\gamma d_n^{-1} \to \infty \text{ as } n \to \infty.$$

Taking logarithm on both sides of the inequality and using the fact that $\xi_n \to \infty$ we obtain for n big enough

$$\log_2 \xi_n \leq cn^\alpha(\log_2 \xi_n)^\alpha - \log_2 A \leq 2cn^\alpha(\log_2 \xi_n)^\alpha, \tag{3.23}$$

and therefore $\log_2 \xi_n \leq (2c)^{1/(1-\alpha)} n^{\alpha/(1-\alpha)}$. Returning back to the notation for the Lipschitz width, we obtain

$$3\gamma 2^{-(2c)^{1/(1-\alpha)} n^{\alpha/(1-\alpha)}} \leq d_n^\gamma(\mathcal{K})_X$$

for n big enough. This completes the proof of (iii) by choosing the constants appropriately so that the above inequality holds for all n. □

3.2.1 Sharpness of the results

In this section, we provide examples which show that some of the results in Theorem 3.9 cannot be improved. We start with the following remark.

Remark 3.10. *The requirement $\alpha < 1$ in Theorem 3.9 (iii), and therefore in Theorem 3.1 (iii), is necessary. The simplest example is $\mathcal{K} = [0,1] \subset \mathbb{R}$, where we have $\epsilon_n(\mathcal{K})_{\mathbb{R}} = 2^{-(n+1)}$ and $d_n^\gamma(\mathcal{K})_{\mathbb{R}} = 0$ when $\gamma \geq \frac{1}{2}$ and $n \geq 1$. The Lipschitz width is zero because $d_n^\gamma(\mathcal{K})_{\mathbb{R}} \leq d_1^{1/2}(\mathcal{K})_{\mathbb{R}}$ for the discussed range of n and γ, and $d_1^{1/2}(\mathcal{K})_{\mathbb{R}} = 0$. The latter holds since the mapping $\Phi_1 : [-1,1] \to \mathbb{R}$, defined as $\Phi_1(t) = \frac{1}{2}(t+1)$ is a 1/2-Lipschitz mapping for which $\Phi_1([-1,1]) = \mathcal{K}$. This shows that an estimate for $d_n^\gamma(\mathcal{K})_X$ in terms of n is not possible.*

Next, we provide an example of a compact set $\mathcal{K}$ for which the the entropy numbers behave like n^{-1}, while the Lipschitz width behaves as $[n \log_2(n+1)]^{-1}$.

We consider the Banach space $X = \mathbf{c_0}$ of all sequences that converge to 0, equipped with the ℓ_∞ norm and its compact subset

$$\mathcal{K}(\sigma) := \{\sigma_j e_j\}_{j=1}^\infty \cup \{0\} \subset \mathbf{c_0}, \tag{3.24}$$

determined by the strictly decreasing converging to 0 sequence $\sigma := (\sigma_j)_{j=1}^\infty$, where $(e_j)_{j=1}^\infty$ are the standard basis in c_0. Since

$$\|\sigma_j e_j - \sigma_{j'} e_{j'}\|_{\ell_\infty} = \sigma_j, \quad \text{for all} \quad j' > j,$$

it follows that the ball with center $\sigma_j e_j$ and radius σ_j contains all points $\sigma_{j'} e_{j'}$ with $j' > j$ and none with $j' < j$. Thus, if we look for 2^n balls with centers in $\mathcal{K}(\sigma)$ covering $\mathcal{K}(\sigma)$ with smallest radius, we take the balls $B(\sigma_j e_j, \sigma_{2^n})$, $j = 1, 2, \ldots, 2^n$, with centers $\sigma_j e_j$ and radius σ_{2^n}. Each of the first $2^n - 1$ balls contain only one point from $\mathcal{K}(\sigma)$, while the last ball $B(\sigma_{2^n} e_{2^n}, \sigma_{2^n})$ contains the rest of the points $\{\sigma_j e_j\}_{j=2^n}^\infty \cup \{0\}$, which gives

$$\tilde{\varepsilon}_n(\mathcal{K}(\sigma))_X = \sigma_{2^n}. \tag{3.25}$$

We investigate the behavior of $d_n^\gamma(\mathcal{K}(\sigma))_X$. We shall use the following lemma which gives upper bounds for the Lipschitz widths for the sets $\mathcal{K}(\sigma)$.

Lemma 3.11. *Consider the strictly decreasing sequence $\sigma := (\sigma_j)_{j=1}^\infty$, $\sigma_j \to 0$ as $j \to \infty$, and the set $\mathcal{K}(\sigma)$, defined in* (3.24). *If $\sigma_1 \leq \gamma/2$ and we can find N (finite or infinite) such that*

$$\sum_{j=1}^N \sigma_j^n \leq (\gamma/2)^n, \tag{3.26}$$

then $d_n^\gamma(\mathcal{K}(\sigma))_X \leq \sigma_N$.

Proof: We consider the case when N is finite. Similar arguments hold in the infinite case. For every σ_j, $j = 1, \ldots, N$, we define $\ell_j \in \mathbb{N} \cup \{0\}$ as

$$2^{-\ell_j - 1} < 2\frac{\sigma_j}{\gamma} \leq 2^{-\ell_j}. \tag{3.27}$$

Then it follows from (3.26) that

$$\sum_{j=1}^{N} 2^{-n\ell_j} \leq \sum_{j=1}^{N} (4\sigma_j/\gamma)^n \leq 2^n. \tag{3.28}$$

Since $(\sigma_j)_{j=1}^{\infty}$ is a decreasing sequence, we have that $2^{-\ell_1} \geq 2^{-\ell_2} \geq 2^{-\ell_3} \geq \cdots \geq 2^{-\ell_N}$. Note that some of the ℓ_j's can be equal to each other. Let $k_1, k_2, \ldots, k_s = N$, be the indices such that

$$\ell_1 = \ldots = \ell_{k_1} < \ell_{k_1+1} = \ldots = \ell_{k_2} < \ell_{k_2+1} = \ldots = \ell_{k_s} = \ell_N.$$

We set $k_0 = 0$ and rewrite inequality (3.28) as

$$2^n \geq \sum_{j=1}^{N} 2^{-n\ell_j} = \sum_{j=1}^{s} (k_j - k_{j-1}) 2^{-n\ell_{k_j}}. \tag{3.29}$$

Observe that the volume of a cube with side length $2^{-\ell_{k_j}}$ is $2^{-n\ell_{k_j}}$, while the volume of $[-1,1]^n$ is 2^n. It follows from simple volumetric considerations, that we can divide naturally the cube $[-1,1]^n$ into k_1 open non-overlaping cubes each with side length $2^{-\ell_{k_1}}$, $(k_2 - k_1)$ open non-overlaping cubes each with side length $2^{-\ell_{k_2}}, \ldots$, $(k_s - k_{s-1})$ open non-overlaping cubes each with side length $2^{-\ell_{k_s}}$, since, according to (3.29), the sum of the total volumes of these cubes does not exceed the total volume of $[-1,1]^n$. Thus, there exists a sequence of non-overlapping open cubes B^j,

$$B^j := B^j(y_j, 2^{-\ell_j - 1}) \subset (B_{\ell_\infty^n}, \|\cdot\|_{\ell_\infty^n}) := [-1,1]^n, \quad j = 1, \ldots, N,$$

with side length $2^{-\ell_j}$. Then, according to Lemma 3.2, the mapping $\Phi : (B_{\ell_\infty^n}, \|\cdot\|_{\ell_\infty^n}) \to \mathbf{c_0}$, defined as

$$\Phi(y) := \sum_{j=1}^{N} \sigma_j (1 - 2^{\ell_j+1} \|y_j - y\|_{\ell_\infty^n})_+ \cdot e_j$$

is a Lipschitz mapping. Its Lipschitz constant is $\sup_{j=1,\ldots,N} \{2^{\ell_j+1}\sigma_j\}$ and $\Phi(y_j) = \sigma_j e_j$, $j = 1, \ldots, N$. It follows from (3.27) that

$$\sup_{j=1,\ldots,N} 2^{\ell_j+1}\sigma_j \leq \gamma,$$

and therefore Φ is a γ-Lipschitz mapping. On the other hand, since

$$\sup_{j'\geq 1} \inf_{y\in B_n} \|\sigma_{j'} e_{j'} - \Phi(y)\|_{\ell_\infty} \leq \sup_{j'\geq 1} \inf_{j=1,\ldots,N} \|\sigma_{j'} e_{j'} - \Phi(y_j)\|_{\ell_\infty} = \sup_{j'\geq 1} \inf_{j=1,\ldots,N} \|\sigma_{j'} e_{j'} - \sigma_j e_j\|_{\ell_\infty} = \sigma_N,$$

and

$$\inf_{y\in B_n} \|0 - \Phi(y)\|_{\ell_\infty} \leq \inf_{j=1,\ldots,N} \|\Phi(y_j)\|_{\ell_\infty} = \inf_{j=1,\ldots,N} \|\sigma_j e_j\|_{\ell_\infty} = \sigma_N,$$

it follows that $d_n^{\gamma}(\mathcal{K}(\sigma))_X \leq \sigma_N$ and the proof is completed. □

Now, we are ready to state the main theorem in this section.

Theorem 3.12. *The compact set $\mathcal{K}(\sigma) \subset \mathbf{c_0}$, defined in (3.24), with $\sigma = (\sigma_j)_{j=1}^{\infty}$ being the sequence $\sigma_j = 1/\log_2(j+1)$ has inner entropy numbers*

$$\tilde{\varepsilon}_n(\mathcal{K}(\sigma)) \asymp \frac{1}{n},$$

and Lipschitz width

$$d_n^\gamma(\mathcal{K}(\sigma)) \asymp \frac{1}{n\log_2(n+1)}$$

for any for $\gamma > 2$.

Proof: The behavior of the entropy follows from (3.25) and the estimate from below for the Lipschitz widths follows from Theorem, 3.9, (i). We are only left to prove the upper estimate for the width. If we show that (3.26) holds for the choice of $\sigma_j = [\log_2(j+1)]^{-1}$, $j = 1, 2, \dots$, and $N = (n+1)^n$, where $\gamma > 2$ and n is sufficiently large, since $\sigma_1 = 1 \le \gamma/2$, we can use Lemma 3.11 to conclude that $d_n^\gamma(\mathcal{K}(\sigma)) \le \sigma_N = (n\log_2(n+1))^{-1}$, for $n \ge n_0$, depending only on γ. This could conclude the proof.

We now concentrate on proving (3.26) with $N = (n+1)^n$ for n sufficiently large. We start with defining $J = J(n)$ as

$$2^{J-1} \le (n+1)^n < 2^J,$$

and estimate

$$\sum_{j=1}^{(n+1)^n} \sigma_j^n \le \sum_{k=0}^{J-1} \sum_{j=2^k}^{2^{k+1}-1} \sigma_j^n \le 1 + \sum_{k=1}^{J-1} 2^k k^{-n} =: 1 + \sum_{k=1}^{J-1} q(k), \quad \text{where } q(t) := 2^t t^{-n}, \quad t \ge 1. \tag{3.30}$$

Simple calculation shows that $q(t)$ is decreasing on $[1, n/\ln 2]$ and increasing on $[n/\ln 2, \infty)$. Moreover, we have that

$$2e^{-n/t} < \frac{q(t+1)}{q(t)} = \frac{2}{\left(1+\frac{1}{t}\right)^n} < 2e^{-n/(t+1)} \le 1/2 \quad \text{for} \quad t \le \frac{n}{\ln 4} - 1. \tag{3.31}$$

It follows from (3.30) that for $n \ge 3$,

$$\sum_{j=1}^{(n+1)^n} \sigma_j^n \le 1 + \sum_{1\le k\le n/\ln 4} q(k) + \sum_{n/\ln 4 < k \le n/\ln 2} q(k) + \sum_{n/\ln 2 < k \le J-1} q(k) =: S_1(n) + S_2(n) + S_3(n). \tag{3.32}$$

We will provide upper bounds for each of S_1, S_2 and S_3. Clearly

$$S_1(n) = 1 + \sum_{1\le k\le n/\ln 4} q(k) < 1 + q(1) \cdot \sum_{1\le k\le n/\ln 4} 2^{-k+1} < 1 + 2 \cdot \sum_{k=1}^{\infty} 2^{-k+1} = 5,$$

since for this range of k's we have $q(k+1) < \frac{1}{2}q(k)$, see (3.31).

Next, note that q is a decreasing function for the range of k in S_2, and therefore

$$S_2(n) \le \left(\frac{n}{\ln 2} - \frac{n}{\ln 4}\right) \cdot 2^{n/\ln 4} \left(\frac{n}{\ln 4}\right)^{-n} < n\left(2^{1/\ln 4}\ln 4\right)^n n^{-n} < n\left(\frac{2.3}{n}\right)^n,$$

since $(n/\ln 2 - n/\ln 4) = n/(2\ln 2) < n$ and $2^{1/\ln 4}\ln 4 < 2.3$. For $n \ge 5$

$$2.3n^{1/n} < 2.3(1 + \frac{1}{\sqrt{2}}) < n \quad \Rightarrow \quad S_2(n) < n\left(\frac{2.3}{n}\right)^n = n\left(\frac{2.3n^{1/n}}{n \cdot n^{1/n}}\right)^n < 1.$$

So, we obtain

$$S_2(n) < 1 \quad \text{for} \quad n \geq 5.$$

To estimate S_3, we notice that the biggest summand is the last one,

$$\begin{aligned} S_3(n) &\leq J \cdot 2^{J-1}(J-1)^{-n} < (n\log_2(n+1)+1)(n+1)^n(n\log_2(n+1)-1)^{-n} \\ &= n\left[1+\frac{1}{n}\right]^n \left[\log_2(n+1)+\frac{1}{n}\right]\left[\log_2(n+1)-\frac{1}{n}\right]^{-n} \\ &< 2en\left[\log_2(n+1)-\frac{1}{n}\right]\left[\log_2(n+1)-\frac{1}{n}\right]^{-n} = 2en\left[\log_2(n+1)-\frac{1}{n}\right]^{1-n}. \end{aligned}$$

Let us now consider the functions

$$\ell(x) := x^{1/(x-1)}, \quad r(x) := \log_2(x+1) - \frac{1}{x}.$$

One can show that ℓ is a decreasing function on the interval $[5, \infty)$, while r is increasing function on the same interval. Therefore, for every $n \geq 5$

$$n^{1/(n-1)} = \ell(n) \leq \ell(5) = 5^{1/4} < \log_2 6 - \frac{1}{5} = r(5) \leq r(n) = \log_2(n+1) - \frac{1}{n},$$

and so

$$n < \left(\log_2(n+1) - \frac{1}{n}\right)^{n-1} \quad \Rightarrow \quad n\left(\log_2(n+1) - \frac{1}{n}\right)^{1-n} < 1.$$

The latter inequality combined with the estimate for S_3 gives that

$$S_3(n) < 2e, \quad \text{for} \quad n \geq 5.$$

Finally, combining (3.32) with all estimates for S_1, S_2 and S_3, we obtain that

$$\sum_{j=1}^{(n+1)^n} \sigma_j^n < S_1(n) + S_2(n) + S_3(n) < 6 + 2e \leq (\gamma/2)^n, \quad \text{provided} \quad n \geq \max\left\{5, \frac{\ln(6+2e)}{\ln\gamma - \ln 2}\right\}.$$

The proof is completed. □

4 Comparison between Lipschitz and Kolmogorov widths

If we fix the value of $n \geq 0$, the Kolmogorov n-width of $\mathcal{K}$ is defined as

$$d_0(\mathcal{K})_X = \sup_{f\in\mathcal{K}} \|f\|_X, \quad d_n(\mathcal{K})_X := \inf_{\dim(X_n)=n} \sup_{f\in\mathcal{K}} \operatorname{dist}(f, X_n)_X, \quad n \geq 1. \tag{4.1}$$

It tells us the optimal performance possible for the approximation of the model class $\mathcal{K}$ using linear spaces of dimension n. However, it does not tell us how to select a (near) optimal space Y of dimension n for this purpose. Let us note that in the definition of Kolmogorov width, we are not requiring that the mapping which sends $f \in \mathcal{K}$ into an approximation to f is a linear map. There is a concept of *linear* width which requires the linearity of the approximation map. Namely, given $n \geq 0$ and a model class $\mathcal{K} \subset X$, its *linear* width $d_n^L(K)_X$ is defined as

$$d_0^L(\mathcal{K})_X = \sup_{f\in\mathcal{K}} \|f\|_X, \quad d_n^L(\mathcal{K})_X := \inf_{L\in\mathcal{L}_n} \sup_{f\in\mathcal{K}} \|f - L(f)\|_X, \quad n \geq 1, \tag{4.2}$$

where the infimum is taken over the class $\mathcal{L}_n$ of all continuous linear maps from X into itself with rank at most n.

We prove in the next theorem the intuitive fact that the Lipschitz width is smaller than the Kolmogorov width.

Theorem 4.1. *For every compact set $\mathcal{K} \subset X$ and every $n \geq 1$, we have*

$$d_n^\gamma(\mathcal{K})_X \leq d_n(\mathcal{K})_X \leq d_n^L(\mathcal{K})_X, \quad \textit{for} \quad \gamma = d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K}). \tag{4.3}$$

Proof: It is clear that $d_n(\mathcal{K})_X \leq d_n^L(\mathcal{K})_X$ for every $n \geq 0$ since we can take X_n to be the n-dimensional linear space containing $L(X)$ when $L \in \mathcal{L}_n$, so we concentrate on the first inequality. We start with $\gamma > d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K})$, denote

$$\eta := \gamma - d_n(\mathcal{K})_X - \mathrm{rad}(\mathcal{K}) > 0,$$

and choose η_1 to be such that $0 < \eta_1 < \eta$. Let $X_n \subset X$ be an n-dimensional linear subspace in X such that,

$$\sup_{f\in\mathcal{K}} \inf_{g\in X_n} \|f - g\|_X < d_n(\mathcal{K})_X + \eta_1.$$

For every $f \in \mathcal{K}$, we denote by $g = g(f)$ the element in X_n for which

$$\|f - g(f)\|_X < d_n(\mathcal{K})_X + \eta_1, \tag{4.4}$$

and the collection of all such elements are denoted by

$$\mathcal{A} = \{g(f) : f \in \mathcal{K}\} \subset X_n.$$

Let us fix $g_0 \in X$ such that $\sup_{f\in\mathcal{K}} \|f - g_0\|_X < \mathrm{rad}\,\mathcal{K} + \eta - \eta_1$. Then, for every $f \in \mathcal{K}$,

$$\|g(f) - g_0\|_X \leq \|g(f) - f\|_X + \|f - g_0\|_X < d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K}) + \eta = \gamma,$$

and therefore

$$\mathrm{rad}(\mathcal{A}) < \gamma, \quad \text{and} \quad \mathcal{A} \subset B(g_0, \gamma) := \{g \in X_n : \|g - g_0\|_X \leq \gamma\}.$$

We now define the mapping $\Phi : (B_{X_n}, \|\cdot\|_X) \to X$ from the unit ball $B_{X_n} := \|g \in X_n : \|g\|_X \leq 1\}$ in X_n as $\Phi(g) = g_0 + \gamma g$. Clearly Φ is a γ-Lipschitz map. Moreover, since $\Phi(B_{X_n}) = B(g_0, \gamma)$ and $\mathcal{A} \subset B(g_0, \gamma)$, we have that

$$\sup_{f\in\mathcal{K}} \inf_{g\in B_{X_n}} \|f - \Phi(g)\|_X \leq \sup_{f\in\mathcal{K}} \inf_{g\in\mathcal{A}} \|f - g\| < d_n(\mathcal{K})_X + \eta_1,$$

where we have used (4.4) in the last inequality. Thus, using Remark 2.1 (iii), we obtain

$$d_n^\gamma(\mathcal{K})_X \leq d_n(\mathcal{K})_X + \eta_1,$$

and letting $\eta_1 \to 0$ gives

$$d_n^\gamma(\mathcal{K})_X \leq d_n(\mathcal{K})_X, \quad \text{for any } \gamma > d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K}).$$

Now (4.3) follows from Theorem 2.7 by taking $\gamma \to d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K})$. □

Corollary 4.2. *For every $n \geq 1$ and every compact set $\mathcal{K} \subset X$ we have*

$$d_n^{\gamma}(\mathcal{K})_X \leq d_n(\mathcal{K})_X, \quad \gamma = 2 \sup_{f \in \mathcal{K}} \|f\|_X. \tag{4.5}$$

Proof: The inequality follows from Theorem 4.1, Remark 2.1 (iii), and the fact that for every $n \geq 1$

$$d_n(\mathcal{K})_X + \mathrm{rad}(\mathcal{K}) \leq 2 \sup_{f \in \mathcal{K}} \|f\|_X.$$

□

As a result of this section, we can give the following improvement of Corollary 3.4.

Corollary 4.3. *If $\mathcal{K} \subset X$ is compact, then for every $n_0 \in \mathbb{N} \cup \{0\}$ and every $\gamma \geq d_{n_0}(\mathcal{K})_X + \mathrm{rad}(\mathcal{K})$, we have*

$$\lim_{n \to \infty} d_n^{\gamma}(\mathcal{K})_X = 0.$$

Proof: The statement follows from Theorem 4.1, Remark 2.1 (iii) and the fact that the sequence of Kolmogorov widths $(d_n(\mathcal{K})_X)$ of a compact set $\mathcal{K}$ is a non-increasing sequence of non-negative numbers that tends to zero, see e.g.[14, Prop 1.2]. □

4.1 Examples of different behavior of the Lipschitz and Kolmogorov widths

It is intuitively clear that the Lipschitz widths could be much smaller than the Kolmogorov widths. We illustrate this observation by discussing the following two examples.

Example 4.4. This example, borrowed from Albert Cohen, arises in some partial differential equations. We denote by χ_a the characteristic function of $[a, a+1]$, $a \in [0,1]$ and consider the univariate linear transport equation

$$\partial_t u_a + a \partial_x u_a = 0, \tag{4.6}$$

with constant velocity $a \in [0,1]$ and initial condition

$$u_0(x) = u_a(x, 0) = \chi_0(x). \tag{4.7}$$

We denote by

$$\mathcal{H} := \{\chi_a : a \in [0,1]\} \equiv \{u_a(x,1) : a \in [0,1]\}$$

the solution manifold to (4.6)-(4.7) evaluated at time $t = 1$. We prove the following lemma for the set $\mathcal{H}$.

Lemma 4.5. *The Kolmogorov width of $\mathcal{H} \subset L_1[0,2]$ satisfies*

$$(n+1)^{-1} \leq d_n(\mathcal{H})_{L_1[0,2]} \leq 4n^{-1}, \tag{4.8}$$

while its inner entropy numbers are given by

$$\tilde{\varepsilon}_n(\mathcal{H})_{L_1[0,2]} = 2^{-n+1}. \tag{4.9}$$

Proof: We first observe that $\|\chi_a - \chi_b\|_{L_1[0,2]} = 2|a-b|$. If we define

$$t_j := (2j+1)2^{-n-1}, \quad j = 0, 1, \dots, 2^n - 1,$$

to be the centers of the intervals $[j2^{-n}, (j+1)2^{-n}] \subset [0,1]$, we have

$$\chi_a \in B(\chi_{t_j}, 2^{-n+1}) \quad \Leftrightarrow \quad \|\chi_a - \chi_{t_j}\|_{L_1[0,2]} \le 2^{-n+1} \quad \Leftrightarrow \quad |a - t_j| \le 2^{-n},$$

where $B(\chi_{t_j}, 2^{-n+1})$ is the closed ball in $L_1[0,2]$ with center χ_{t_j} and radius 2^{-n+1}. So, those balls cover $\mathcal{H}$. This calculation also shows that if we have 2^n balls covering $\mathcal{H}$ and one of them has radius strictly smaller than 2^{-n+1} then some other one must have a radius strictly bigger than 2^{-n+1}. This proves (4.9).

To show (4.8), we first observe that the n-dimensional space

$$V_n := \operatorname{span}\{\bar{\chi}_j, \quad j = 0, \dots, n-1\},$$

where $\bar{\chi}_j$ is the characteristic function of the interval $[2j/n, 2(j+1)/n]$ provides an error at most $4n^{-1}$ for the elements from $\mathcal{H}$. Indeed, for each $\chi_a \in \mathcal{H}$, we have

$$\|\chi_a - \sum_{j=j_1}^{j_2} \bar{\chi}_j\|_{L_1[0,2]} \le 4n^{-1},$$

where $j_1 = j_1(a)$ and $j_2 = j_2(a)$ are defined as

$$j_1(a) = \max\{j : \ 2j/n \le a, \ 0 \le j \le n-1\}, \quad j_2(a) = \max\{j : \ 2j/n \le a+1, 0 \le j \le n-1\},$$

and therefore $d_n(\mathcal{H})_{L_1[0,2]} \le 4n^{-1}$. To prove the lower bound in (4.8), we use a well known result, see e.g. [14, Chap. II, Prop 1.3], which states that for any unit ball U in a Banach space X and any finite dimensional space $\mathcal{V}_{n+1}$ of dimension $n+1$, the Kolmogorov width

$$d_n(U \cap \mathcal{V}_{n+1})_X = 1. \tag{4.10}$$

We apply this result for the Banach space $X = L_1[0,2]$, the unit ball U in $L_1[0,2]$, and the linear space $\mathcal{V}_{n+1} \subset L_p[0,1]$, defined as

$$\mathcal{V}_{n+1} = \operatorname{span}\{\varphi_0, \dots, \varphi_n\}, \quad \varphi_j := \chi_{j/(n+1)} - \chi_{(j+1)/(n+1)}, \quad j = 0, \dots, n.$$

Another representation for the φ_j's is

$$\varphi_j := \chi_{[j/(n+1),(j+1)/(n+1)]} - \chi_{[1+j/(n+1),1+(j+1)/(n+1)]}, \quad j = 0, \dots, n,$$

and since they have disjoint supports, every $\varphi = \sum_{j=0}^n \alpha_j \varphi_j \in \mathcal{V}_{n+1}$ has norm

$$\|\varphi\|_{L_1[0,2]} = \|\sum_{j=0}^n \alpha_j \varphi_j\|_{L_1[0,2]} = 2(n+1)^{-1} \sum_{j=0}^n |\alpha_j|. \tag{4.11}$$

Therefore

$$\varphi = \sum_{j=0}^n \alpha_j \varphi_j \in U \cap \mathcal{V}_{n+1} \quad \Leftrightarrow \quad \sum_{j=0}^n |\alpha_j| \le \frac{1}{2}(n+1). \tag{4.12}$$

Let us fix an n dimensional subspace V_n and let $v_j \in V_n$ be such that

$$\mathrm{dist}(\chi_{j/(n+1)}, V_n)_{L_1[0,2]} = \|\chi_{j/(n+1)} - v_j\|_{L_1[0,2]}, \quad j = 0, \dots, n.$$

Then, for every $\varphi \in U \cap \mathcal{V}_{n+1}$, we have

$$\begin{aligned}
\mathrm{dist}(\varphi, V_n)_{L_1[0,2]} &\leq \|\sum_{j=0}^{n} \alpha_j \varphi_j - \sum_{j=0}^{n} \alpha_j (v_j - v_{j+1})\|_{L_1[0,2]} \\
&= \|\sum_{j=0}^{n} \alpha_j (\chi_{j/(n+1)} - v_j) - \sum_{j=0}^{n} \alpha_j (\chi_{(j+1)/(n+1)} - v_{j+1})\|_{L_1[0,2]} \\
&\leq 2\,\mathrm{dist}(\mathcal{H}, V)_{L_1[0,2]} \sum_{j=0}^{n} |\alpha_j| \leq (n+1)\,\mathrm{dist}(\mathcal{H}, V_n)_{L_1[0,2]},
\end{aligned}$$

where we have used (4.12). Therefore, it follows from (4.10) and the latter estimate that

$$\begin{aligned}
1 &= d_n(U \cap \mathcal{V}_{n+1})_{L_1[0,2]} = \inf_{V_n} \sup_{\varphi \in U \cap \mathcal{V}_{n+1}} \mathrm{dist}(\varphi, V_n)_{L_1[0,2]} \\
&\leq (n+1) \inf_{V_n} \mathrm{dist}(\mathcal{H}, V_n)_{L_1[0,2]} = (n+1) d_n(\mathcal{H})_{L_1[0,2]},
\end{aligned}$$

and the proof is completed. □

It follows then from Lemma 4.5 and Theorem 3.3 that the Lipschitz width of $\mathcal{H}$ decays exponentially, while its Kolmogorov width decays like n^{-1}. While this is a good example, one may argue that this different behavior is due to the fact that $\mathcal{H}$ is not convex. It is a well known fact that for every compact set $\mathcal{K}$ we have

$$d_n(\mathcal{K})_X = d_n(\mathcal{K}_c)_X, \quad \text{where} \quad \mathcal{K}_c = \mathrm{conv}(\mathcal{K} \cup (-\mathcal{K}))$$

is the minimal convex centrally symmetric set that contains $\mathcal{K}$. Therefore, a more suitable example would be one when $\mathcal{K}$ is a convex, centrally symmetric set. We discuss such case in Example 4.6.

Example 4.6. Consider the sequence $\sigma = (\sigma_j)_{j=1}^{\infty}$, with $\sigma_j = (\log_2(j+1))^{-1/2}$, and the corresponding linear map on sequences, $D_\sigma : \ell_1 \to \ell_2$, $\ell_1 := \{x = (x_1, x_2, \dots) : \sum_{j=1}^{\infty} |x_j| < \infty\}$, defined as

$$D_\sigma(x) = y, \quad \text{where} \quad y_j = \sigma_j x_j, \quad j = 1, 2, \dots.$$

Let us denote by $\mathcal{K}_\sigma \subset \ell_2$ the image of the unit ball in ℓ_1 under this map, namely,

$$\mathcal{K}_\sigma := \{y \in \ell_2 : \ y_j = \sigma_j x_j, \ \text{where} \sum_{j=1}^{\infty} |x_j| \leq 1\} = \{y \in \ell_2 \ : \ \sum_{j=1}^{\infty} |y_j| \sqrt{\log_2(j+1)} \leq 1\}. \quad (4.13)$$

The set $\mathcal{K}_\sigma$ is a convex, centrally symmetric subset of ℓ_2 for which

$$\left\{ \frac{\pm e_j}{\sqrt{\log_2(j+1)}} \right\}_{j=1}^{\infty} \subset \mathcal{K}_\sigma.$$

It follows from Proposition 3.1 in [11] that

$$\varepsilon_n(\mathcal{K}_\sigma)_{\ell_2} \asymp n^{-1/2}, \quad n = 1, 2, \dots, \tag{4.14}$$

which combined Theorem 3.3 shows that $d_n^\gamma(\mathcal{K}_\sigma)_{\ell_2} \le Cn^{-1/2}$ with $\gamma = 2\,\mathrm{rad}(\mathcal{K}_\sigma) = 2$. On the other hand, we show in the next lemma that its Kolmogorov width $d_n(\mathcal{K}_\sigma)_{\ell_2}$ behaves as $d_n(\mathcal{K}_\sigma)_{\ell_2} \asymp (\log_2 n)^{-1/2}$.

Lemma 4.7. *The Kolmogorov width of the compact set $\mathcal{K}_\sigma$ defined in* (4.13) *is*

$$d_n(\mathcal{K}_\sigma)_{\ell_2} \asymp (\log_2 n)^{-1/2}, \quad n = 2, 3, \dots.$$

Proof: Clearly,

$$d_n(\mathcal{K}_\sigma)_{\ell_2} \le \sup_{x \in \mathcal{K}_\sigma} \mathrm{dist}(x, \mathrm{span}\{e_j\}_{j=1}^n)_{\ell_2} = \frac{1}{\sqrt{\log_2(n+2)}}.$$

To prove the inequality from below, we fix $\epsilon > 0$ and denote by X_n the n dimensional subspace for which

$$\sup_{x \in \mathcal{K}_\sigma} \mathrm{dist}(x, X_n)_{\ell_2} \le (1+\epsilon) d_n(\mathcal{K}_\sigma)_{\ell_2}.$$

If P is the orthogonal projection onto $\ell_2^{2n} := \mathrm{span}\{e_j\}_{j=1}^{2n}$ and $\widetilde{X}_n := P(X_n)$, then

$$d_n(P(\mathcal{K}_\sigma))_{\ell_2} \le \sup_{x \in P(\mathcal{K}_\sigma)} \mathrm{dist}(x, \widetilde{X}_n)_{\ell_2} \le (1+\epsilon) d_n(\mathcal{K}_\sigma)_{\ell_2}. \tag{4.15}$$

Since $\mathcal{P}_n := \frac{1}{\sqrt{\log_2(2n+1)}} \mathrm{conv}\{\pm e_j\}_{j=1}^{2n} \subset P(\mathcal{K}_\sigma)$, we have

$$d_n(\mathcal{P}_n)_{\ell_2} \le d_n(P(\mathcal{K}_\sigma))_{\ell_2}, \tag{4.16}$$

and from Stechkin's theorem [12, Ch. 13 Th.3.3] we know that

$$d_n(\mathcal{P}_n)_{\ell_2} = \frac{1}{\sqrt{2}} \left(\log_2(2n+1)\right)^{-1/2}. \tag{4.17}$$

Combining (4.15), (4.16) and (4.17) gives

$$\frac{1}{\sqrt{2}(1+\varepsilon)} \left(\log_2(2n+1)\right)^{-1/2} \le d_n(\mathcal{K}_\sigma)_{\ell_2}.$$

Since $\epsilon > 0$ is arbitrary, we obtain

$$C(\log_2 n)^{-1/2} \le d_n(\mathcal{K}_\sigma)_{\ell_2},$$

which completes the proof. $\square$

5 Comparison between Lipschitz and stable manifold widths

Let us recall the definition of *manifold* width $\delta_n(\mathcal{K})_X$ for the compact set $\mathcal{K} \subset X$, see [7, 8],

$$\delta_n(\mathcal{K})_X := \inf_{a,M} \sup_{f \in \mathcal{K}} \|f - M(a(f))\|_X, \tag{5.1}$$

where the infimum is taken over all mappings $a : \mathcal{K} \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to X$ with a continuous on $\mathcal{K}$ and M continuous on $\mathbb{R}^n$. A comparison between manifold widths and other types of nonlinear widths was given in [8]. There is also another concept, called *stable manifold* width $\delta^*_{n,\gamma}(\mathcal{K})_X$ of the compact set $\mathcal{K} \subset X$, see [5], defined as

$$\delta^*_{n,\gamma}(\mathcal{K})_X := \inf_{a,M,\|\cdot\|_{Y_n}} \sup_{f \in \mathcal{K}} \|f - M(a(f))\|_X, \tag{5.2}$$

where now the infimum is taken over all maps $a : \mathcal{K} \to (\mathbb{R}^n, \|\cdot\|_{Y_n})$, $M : (\mathbb{R}^n, \|\cdot\|_{Y_n}) \to X$, and norms $\|\cdot\|_{Y_n}$ on $\mathbb{R}^n$, with a, M being γ-Lipschitz. We discuss in this section the relation between stable manifold widths and the Lipschitz widths. The next theorem shows that for any compact set $\mathcal{K} \subset X$, the Lipschitz widths are smaller than the stable manifold widths.

Theorem 5.1. *For every compact set $\mathcal{K} \subset X$, every $n \geq 1$, and every $\gamma > 0$, we have*

$$d_n^{\gamma^2 \mathrm{diam}(\mathcal{K})}(\mathcal{K})_X \leq \delta^*_{n,\gamma}(\mathcal{K})_X. \tag{5.3}$$

Proof: We choose $\epsilon > 0$, and let $a : \mathcal{K} \to (\mathbb{R}^n, \|\cdot\|_{Y_n})$ and $M : (\mathbb{R}^n, \|\cdot\|_{Y_n}) \to X$ be two γ-Lipschitz mappings with respect to a norm $\|\cdot\|_{Y_n}$ in $\mathbb{R}^n$ such that for every $f \in \mathcal{K}$,

$$\|f - M \circ a(f)\|_X \leq \delta^*_{n,\gamma}(\mathcal{K})_X + \epsilon. \tag{5.4}$$

For every $f_1, f_2 \in \mathcal{K}$ we have

$$\|a(f_1) - a(f_2)\|_{Y_n} \leq \gamma \|f_1 - f_2\|_X,$$

which implies

$$\mathrm{diam}(\mathcal{A}) \leq \gamma\, \mathrm{diam}(\mathcal{K}), \quad \text{where} \quad \mathcal{A} := a(\mathcal{K}).$$

We fix an element $f_0 \in \mathcal{K}$ and define the mapping $\Phi : (B_{Y_n}, \|\cdot\|_{Y_n}) \to X$ as

$$\Phi(y) := M(a(f_0) + \gamma\, \mathrm{diam}(\mathcal{K}) y), \quad a(f_0) \in \mathbb{R}^n.$$

Note that Φ is a $\gamma^2\, \mathrm{diam}(\mathcal{K})$-Lipschitz mapping. For each $f \in \mathcal{K}$ we define

$$y(f) := \frac{1}{\gamma\, \mathrm{diam}(\mathcal{K})}(a(f) - a(f_0)) \in B_{Y_n}.$$

An easy calculation shows that $\Phi(y(f)) = M \circ a(f)$, so

$$\|f - \Phi(y(f))\|_X = \|f - M \circ a(f)\|_X \leq \delta^*_{n,\gamma}(\mathcal{K})_X + \epsilon,$$

where we have used (5.4). Therefore we obtain

$$\sup_{f \in \mathcal{K}} \inf_{y \in B_{Y_n}} \|f - \Phi(y)\|_X \leq \delta^*_{n,\gamma}(\mathcal{K})_X + \epsilon \quad \Rightarrow \quad d_n^{\gamma^2\, \mathrm{diam}(\mathcal{K})}(\mathcal{K})_X \leq \delta^*_{n,\gamma}(\mathcal{K})_X + \epsilon.$$

Since ϵ is arbitrary, (5.3) holds and the proof is completed. □

Remark 5.2. *It was shown in [5] that in the case of Hilbert space H*

$$\delta^*_{26n,2}(\mathcal{K})_H \le 3\varepsilon_n(\mathcal{K})_H.$$

Thus, the above result and Theorem 5.1 lead to the inequality

$$d_{26n}^{4\mathrm{diam}(\mathcal{K})}(\mathcal{K})_H \le 3\varepsilon_n(\mathcal{K})_H,$$

which is an estimate in the spirit of Theorem 3.3. Note that direct estimates between stable manifold widths and entropy numbers are known only in the Hilbert space case, while direct estimates between Lipschitz widths and entropy numbers, as seen from Theorem 3.3, hold for any Banach space.

Theorem 5.3. *For every Banach space X and for every $n \in \mathbb{N}$, there exist compact sets $\mathcal{K} \subset X$ such that for every $\gamma > 0$,*

$$\delta^*_{n,\gamma}(\mathcal{K})_X \ge \delta_n(\mathcal{K})_X \ge 1, \quad \textit{while} \quad \lim_{\gamma\to\infty} d_n^\gamma(\mathcal{K})_X = 0.$$

Proof: Let us fix $n \in \mathbb{N}$ and consider the compact set

$$\mathcal{K} := S_{n+1} \subset X_{n+1} \subset X,$$

where S_{n+1} is the boundary of the unit sphere of an $(n+1)$-dimensional subspace $(X_{n+1}, \|\cdot\|_X)$ of X. By the Borsuk theorem, see [2, 12], we have that for any continuous map $a : \mathcal{K} \to \mathbb{R}^n$, there exists $f_0 \in \mathcal{K}$ such that $a(f_0) = a(-f_0)$, and thus for any map $M : \mathbb{R}^n \to X$ we have $M(a(f_0)) = M(a(-f_0))$. Then, since $\|f_0\|_X = 1$ and $f_0, -f_0 \in \mathcal{K}$, we have the inequality

$$\begin{aligned} 2 &= \|f_0 - (-f_0)\|_X = \|f_0 - M(a(f_0)) + (M(a(-f_0)) - (-f_0)\|_X \\ &\le \|f_0 - M(a(f_0))\|_X + \|M(a(-f_0)) - (-f_0)\|_X \le 2\sup_{f\in\mathcal{K}} \|f - M(a(f))\|_X \end{aligned} \tag{5.5}$$

for all mappings $a : \mathcal{K} \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to X$ with a continuous on $\mathcal{K}$ and M continuous on $\mathbb{R}^n$. So $\delta_n(\mathcal{K})_X \ge 1$, and therefore $\delta^*_{n,\gamma}(\mathcal{K})_X \ge \delta_n(\mathcal{K})_X \ge 1$ for any $\gamma > 0$. On the other hand, since $\mathcal{K}$ is compact (and thus totally bounded), we have that $\lim_{\gamma\to\infty} d_n^\gamma(\mathcal{K})_X = 0$ because of Lemma 2.8. $\square$

Probably one of the main differences between stable manifold widths and the Lipschitz widths is how they depend on the Lipschitz constant γ. Even though we touch upon this in the above theorem, we will present a simple example where this difference will be seen quite clearly.

Example 5.4. We consider the set $\mathcal{K} \subset \mathbb{R}^2$ to be the Euclidean unit ball $\mathcal{K} = \{(x, y) \,:\, x^2+y^2 \le 1\}$.

- Manifold widths: it follows from Borsuk's theorem applied to the restriction of the mapping a on $\partial\mathcal{K}$ (which is a continuous function) that there exists a point $(x^*, y^*) \in \partial\mathcal{K}$ such that $a(x^*, y^*) = a(-x^*, -y^*)$. Then, no matter what the size of γ is, the fact that $M(a(x^*, y^*)) = M(a(-x^*, -y^*)) =: (\hat{x}, \hat{y})$ implies that we can not approximate both (x^*, y^*) and $(-x^*, -y^*)$ by $(\hat{x}, \hat{y})$ with accuracy better than 1.

- Lipschitz widths: we fix k and find an upper bound for $d_1^{\gamma}(\mathcal{K})_{\mathbb{R}^2}$ with $\gamma = (k+1)\pi$. For that purpose, we consider the mapping $\Phi_1 : [-1,1] \to \mathbb{R}^2$, defined as

$$\Phi_1(t) = \left(\frac{1}{2}(t+1)\cos(k\pi(t+1)), \frac{1}{2}(t+1)\sin(k\pi(t+1))\right)^T, \quad t \in [-1,1].$$

Clearly, $\Phi_1([-1,1])$ is a spiral (an Archimedean spiral) starting at $\Phi_1(-1) = (0,0)^T$ and ending at $\Phi_1(1) = (1,0)^T$. A ray $r(\cos(k\pi(t^*+1)), \sin(k\pi(t^*+1)))^T$ from the origin with fixed $t^* \in [-1, -1+\frac{2}{k}]$ intersects successive turnings of the spiral in points with a constant separation distance $\frac{1}{k}$. Therefore, $\Phi_1([-1,1])$ approximates $\mathcal{K}$ with accuracy at most $\frac{1}{k}$. Note that Φ_1 is a γ-Lipschitz mapping, which gives the upper bound $d_1^{\gamma}(\mathcal{K})_{\mathbb{R}^2} \leq \frac{1}{k}$. Thus, because of monotonicity, $d_n^{\gamma}(\mathcal{K})_{\mathbb{R}^2} \leq \frac{1}{k}$, and we see that as $k \to \infty$, we have $d_n^{\gamma}(\mathcal{K})_{\mathbb{R}^2} \to 0$ when $\gamma \to \infty$.

6 Relation to neural networks

In this section, we discuss deep neural network approximation (DNNA) by feed-forward ReLU neural networks (NN) of constant width $W > 2$ and depth n, whose parameters have absolute values bounded by 1. We will show that the approximation tools provided by these NNs are in fact Lipschitz mappings

$$\Phi : (B_{\ell_\infty^{\tilde{n}}}, \|\cdot\|_{\ell_\infty^{\tilde{n}}}) \to C(\Omega), \quad \Omega = [0,1]^d, \quad \tilde{n} = Cn, \quad C = C(W),$$

with Lipschitz constant $\gamma(n) = C'nW^n$. It follows from Remark 2.1 (iii) that

$$d_{\tilde{n}}^{\tilde{\gamma}_n}(\mathcal{K})_X \leq d_{\tilde{n}}^{\gamma(n)}(\mathcal{K})_X \leq d_{\tilde{n}}^{\hat{\gamma}_n}(\mathcal{K})_X, \quad \text{with} \quad \tilde{\gamma}_n = C'W^{2n}, \quad \hat{\gamma}_n = C'W^n. \tag{6.1}$$

Therefore, a theoretical benchmark for the performance of the DNNA for a class $\mathcal{K} \subset X$ is given by the Lipschitz width $d_{\tilde{n}}^{\gamma_n}(\mathcal{K})_X$ with $\gamma_n = C'\lambda^n$, where $C' > 0$ and $\lambda > 2$ are fixed constants. This observation motivates our investigation of Lipschitz widths whose Lipschitz constant depends on n.

6.1 Deep neural networks as Lipschitz mappings

Let us first recall that a DNNA of a function $f \in C(\Omega)$, $\Omega \subset \mathbb{R}^d$, via feed-forward NN with activation function $\sigma : \mathbb{R} \to \mathbb{R}$, constant width W and depth n is in fact an approximation to f by the family of functions

$$\Sigma_n := \{\Phi(y) : \ y \in \mathbb{R}^{\tilde{n}}, \ \tilde{n} = \tilde{n}(W,n) = Cn\} \subset C(\Omega).$$

For each $y \in \mathbb{R}^{\tilde{n}}$, $\Phi(y) \in C(\Omega)$ is a continuous function $\Phi(y) : \Omega \to \mathbb{R}$ of the form

$$\Phi(y) := A^{(n)} \circ \bar{\sigma} \circ A^{(n-1)} \circ \ldots \circ \bar{\sigma} \circ A^{(0)}, \tag{6.2}$$

with $A^{(0)} : \mathbb{R}^d \to \mathbb{R}^W$, $A^{(\ell)} : \mathbb{R}^W \to \mathbb{R}^W$, $\ell = 1, \ldots, n-1$, and $A^{(n)} : \mathbb{R}^W \to \mathbb{R}$ being affine mappings, and $\bar{\sigma} : \mathbb{R}^W \to \mathbb{R}^W$ given by

$$\bar{\sigma}(x_{j+1}, \ldots, x_{j+W}) = (\sigma(x_{j+1}), \ldots, \sigma(x_{j+W})).$$

The argument y of Φ is a vector in $\mathbb{R}^{\tilde{n}}$ that consists of the entries of the matrices and offset vectors (biases) of the affine mappings $A^{(\ell)}$, $\ell = 0, \dots, n$. We order these entries in such a way that the entries of $A^{(\ell)}$ appear before those of $A^{(\ell+1)}$ and the ordering for each $A^{(\ell)}$ is done in the same way. Before going further, we need to specify a norm $\|\cdot\|_{Y_{\tilde{n}}}$ to be used for $\mathbb{R}^{\tilde{n}}$. We take this norm to be the $\ell_\infty^{\tilde{n}} := \ell_\infty(\mathbb{R}^{\tilde{n}})$ norm, that is, $\|y\|_{\ell_\infty^{\tilde{n}}} := \max_{1\le i\le \tilde{n}} |y_i|$. This choice is not optimal for obtaining the best constants in our estimates but it will simplify the exposition that follows. Also, when considering vector functions $g = (g_1, \dots, g_W)$ from $C(\Omega)$, we use the notation

$$\|g\| := \max_{1\le i\le W} \|g_i\|_{C(\Omega)}.$$

It was proven in [6] that if B is any finite ball in $\ell_\infty(\mathbb{R}^{\tilde{n}})$ and $\sigma(t) = \mathrm{ReLU}(t) = \max\{t, 0\} = t_+$, then $\Phi : B \to C(\Omega)$ is a γ-Lipschitz mapping with γ depending only on B, W, n, and d. In fact, Φ is a γ-Lipschitz map on any bounded set. Here, we will investigate in detail the Lipschitz constant γ in the case when B is the unit ball $(B_{\ell_\infty^{\tilde{n}}}, \|\cdot\|_{\ell_\infty^{\tilde{n}}})$. More precisely, the following theorem holds.

Theorem 6.1. *The mapping* $\Phi : (B_{\ell_\infty^{\tilde{n}}}, \|\cdot\|_{\ell_\infty^{\tilde{n}}}) \to C(\Omega)$, *with* $\Omega = [0,1]^d \subset \mathbb{R}^d$, *defined in* (6.2) *with* $\sigma = \mathrm{ReLU}$ *is a* $C'nW^n$*-Lipschitz mapping, that is*

$$\|\Phi(y) - \Phi(y')\|_{C(\Omega)} \le C'nW^n \|y - y'\|_{\ell_\infty^{\tilde{n}}}, \quad y, y' \in B_{\ell_\infty^{\tilde{n}}},$$

where $C' = C'(d)$ *is a constant depending on* d.

Proof: Let y and y' be the entries of the affine mappings $A^{(j)}(\cdot) := A_j(\cdot) + b^{(j)}$, $j = 0, \dots, n$, and $A'^{(j)}(\cdot) := A'_j(\cdot) + b'^{(j)}$, $j = 0, \dots, n$, respectively, ordered in a predetermined way. We fix $x \in \Omega$ and denote by

$$\eta^{(0)}(x) := \overline{\mathrm{ReLU}}(A_0 x + b^{(0)}), \quad \eta'^{(0)}(x) := \overline{\mathrm{ReLU}}(A'_0 x + b'^{(0)}),$$

$$\eta^{(j)} := \overline{\mathrm{ReLU}}(A_j \eta^{(j-1)} + b^{(j)}), \quad \eta'^{(j)} := \overline{\mathrm{ReLU}}(A'_j \eta'^{(j-1)} + b'^{(j)}), \quad j = 1, \dots, n-1,$$

where $A_j, A'_j, b^{(j)}, b'^{(j)}$, $j = 1, \dots, n-1$, are the respective $W \times W$ matrices and bias vectors, associated to y and y', and

$$\eta^{(n)} := A_n \eta^{(n-1)} + b^{(n)}, \quad \eta'^{(n)} := A'_n \eta'^{(n-1)} + b'^{(n)}.$$

Note that since $\|y\|_{\ell_\infty^{\tilde{n}}} \le 1$,

$$\|\eta'^{(0)}\| \le (d+1)\|y\|_{\ell_\infty^{\tilde{n}}} \le d+1, \quad \|\eta'^{(j)}\| \le (W\|\eta'^{(j-1)}\| + 1)\|y\|_{\ell_\infty^{\tilde{n}}} \le W\|\eta'^{(j-1)}\| + 1, \quad j = 1, \dots, n.$$

One can show by induction that for $j = 1, \dots, n$,

$$\|\eta'^{(j)}\| \le W^j d + \sum_{k=0}^{j} W^k \le (d+2)W^j. \tag{6.3}$$

Note that the above inequality also holds for $j = 0$. Next, since ReLU is a Lip 1 function, we have

$$\|\eta^{(0)}(x) - \eta'^{(0)}(x)\| \le \|(A_0 - A'_0)x\| + \|b^{(0)} - b'^{(0)}\| \le (d+1)\|y - y'\|_{\ell_\infty^{\tilde{n}}} =: C_0 \|y - y'\|_{\ell_\infty^{\tilde{n}}},$$

and therefore $\|\eta^{(0)} - \eta'^{(0)}\| \leq C_0\|y - y'\|_{\ell_\infty^{\tilde{n}}}$. Suppose we have proved that

$$\|\eta^{(j-1)} - \eta'^{(j-1)}\| \leq C_{j-1}\|y - y'\|_{\ell_\infty^{\tilde{n}}}.$$

It follows that

$$\begin{aligned}
\|\eta^{(j)}(x) - \eta'^{(j)}(x)\| &\leq \|A_j\eta^{(j-1)}(x) + b^{(j)} - A'_j\eta'^{(j-1)}(x) - b'^{(j)}\| \\
&\leq \|A_j(\eta^{(j-1)}(x) - \eta'^{(j-1)}(x))\| + \|(A_j - A'_j)\eta'^{(j-1)}(x)\| + \|b^{(j)} - b'^{(j)}\| \\
&\leq W\|y\|_{\ell_\infty^{\tilde{n}}}\|\eta^{(j-1)} - \eta'^{(j-1)}\| + W\|y - y'\|_{\ell_\infty^{\tilde{n}}}\|\eta'^{(j-1)}\| + \|y - y'\|_{\ell_\infty^{\tilde{n}}} \\
&\leq (WC_{j-1} + (d+2)W^j + 1)\|y - y'\|_{\ell_\infty^{\tilde{n}}} \\
&=: C_j\|y - y'\|_{\ell_\infty^{\tilde{n}}},
\end{aligned}$$

where we have used the induction hypothesis, the fact that $\|y\|_{\ell_\infty^{\tilde{n}}} \leq 1$, and the bound (6.3) for $\|\eta'^{(j)}\|$. Thus, we have obtained that $\|\eta^{(j)} - \eta'^{(j)}\| \leq C_j\|y - y'\|_{\ell_\infty^{\tilde{n}}}$, and therefore the following recursive relation

$$C_j = WC_{j-1} + (d+2)W^j + 1, \quad j = 1, \dots, n,$$

between the constants C_j, $j = 1, \dots, n$, where $C_0 = d + 1$. We then obtain

$$C_n < (n+1)(d+2)W^n + \sum_{k=0}^{n-1} W^k < C'nW^n, \quad \text{with } C' = C'(d).$$

Finally, we write

$$\|\Phi(y) - \Phi(y')\|_{C(\Omega)} = \|\eta^{(n)} - \eta'^{(n)}\| \leq C_n\|y - y'\|_{\ell_\infty^{\tilde{n}}} < C'nW^n\|y - y'\|_{\ell_\infty^{\tilde{n}}},$$

and the proof is completed. □

We next discuss a Carl's type inequality that is similar to Lemma 3.7, but is for the case when the Lipschitz constant γ depends on n.

Remark 6.2. *If one follows the proof of Lemma* 3.7 *with the condition that γ is not a constant, but $\gamma = \gamma_n = C'\lambda^n$, where $C' > 0$ and $\lambda > 1$, one can show that*

$$d_n^{\gamma_n}(\mathcal{K})_X < c_0\frac{[\log_2 n]^\beta}{n^{2\alpha}}, \quad \beta \in \mathbb{R}, \;\alpha > 0 \quad \Rightarrow \quad \varepsilon_m(\mathcal{K})_X < C\frac{[\log_2 m]^\beta}{m^\alpha}, \quad \textit{where } m = cn^2,$$

and C, c are fixed constants, depending only on c_0, β, α, λ and C'. Indeed, the proof follows from the fact that for $\varepsilon = c_0[\log_2 n]^\beta n^{-2\alpha}$ we have

$$N_{2\varepsilon}(\mathcal{K}) \leq \left(\frac{3\gamma}{\varepsilon}\right)^n = (3C'c_0^{-1}\lambda^n[\log_2 n]^{-\beta}n^{2\alpha})^n < 2^{cn^2},$$

and therefore

$$\varepsilon_{cn^2}(\mathcal{K})_X < 2c_0[\log_2 n]^\beta n^{-2\alpha}.$$

Setting $m = cn^2$ i.e. $n = \sqrt{m/c}$ gives

$$\varepsilon_m(\mathcal{K})_X \leq 2c_0[\log_2 \sqrt{m/c}]^\beta(m/c)^{-\alpha} = 2c_0c^\alpha 2^{-\beta}\frac{[\log_2 m - \log_2 c]^\beta}{m^\alpha} < C''\frac{[\log_2 m]^\beta}{m^\alpha},$$

which is what we wanted to show.

6.2 Lower bound for $d_n^{\gamma_n}(\mathcal{K})_X$

Now that we know that DNNA is an approximation to a function f by a particular $\gamma(n)$-Lipschitz mapping with $\gamma(n) = C'nW^n$, we can ask the question what are the limits of such approximation. This question is answered by providing a lower bound for the Lipschitz width $d_n^{\gamma_n}(\mathcal{K})_X$, $\gamma_n = C'W^{2n}$, see (6.1), via the next theorem which is a modification of Theorem 3.9.

Theorem 6.3. *For any compact set $\mathcal{K} \subset X$ we consider the Lipschitz width $d_n^{\gamma_n}(\mathcal{K})_X$ with Lipschitz constant $\gamma_n = C'\lambda^n$, $\lambda > 1$ and $C' > 0$ being fixed constants. Then the following holds:*

(i) if for some constants $c_1 > 0, \alpha > 0$ and $\beta \in \mathbb{R}$ we have

$$\varepsilon_n(\mathcal{K})_X > c_1 \frac{(\log_2 n)^\beta}{n^\alpha}, \quad n = 1, 2, \dots,$$

then there exists a constant $C > 0$ such that

$$d_n^{\gamma_n}(\mathcal{K})_X \geq C \frac{(\log_2 n)^\beta}{n^{2\alpha}}, \quad n = 1, 2, \dots. \tag{6.4}$$

(ii) if for some constants $c_1 > 0, \alpha > 0$ we have

$$\varepsilon_n(\mathcal{K})_X > c_1 (\log_2 n)^{-\alpha}, \quad n = 1, 2, \dots,$$

then there exists a constant $C > 0$ such that

$$d_n^{\gamma_n}(\mathcal{K})_X \geq C (\log_2 n)^{-\alpha}, \quad n = 1, 2, \dots. \tag{6.5}$$

Proof: We prove the theorem by contradiction. We first concentrate on the proof of (i). If (6.4) does not hold for some constant C, then there exists a strictly increasing sequence of integers $(n_k)_{k=1}^\infty$, such that

$$p_k := \frac{d_{n_k}^{\gamma_{n_k}}(\mathcal{K}) n_k^{2\alpha}}{(\log_2 n_k)^\beta} \to 0 \quad \text{as} \quad k \to \infty.$$

Thus we can write

$$d_{n_k}^{\gamma_{n_k}}(\mathcal{K}) = \frac{p_k \left[\log_2 n_k\right]^\beta}{n_k^{2\alpha}} < \frac{2p_k \left[\log_2 n_k\right]^\beta}{n_k^{2\alpha}} =: \delta_k, \quad k = 1, 2, \dots.$$

Now we apply Proposition 3.8 with $\eta_n = c_1 (\log_2 n)^\beta n^{-\alpha}$ and obtain

$$c_1 \left[\log_2(n_k \log_2(3\gamma_{n_k}\delta_k^{-1}))\right]^\beta {n_k}^{-\alpha} \left[\log_2(3\gamma_{n_k}\delta_k^{-1})\right]^{-\alpha} \leq 4 \frac{p_k \left[\log_2 n_k\right]^\beta}{{n_k}^{2\alpha}},$$

which we rewrite as

$$p_k^{-1} \left[\log_2 n_k + \log_2 \log_2(3\gamma_{n_k}\delta_k^{-1})\right]^\beta \left[\log_2(3\gamma_{n_k}\delta_k^{-1})\right]^{-\alpha} \leq C_1 \left[\log_2 n_k\right]^\beta n_k^{-\alpha}, \quad C_1 = 4/c_1. \tag{6.6}$$

Observe that

$$\begin{aligned} \log_2(3\gamma_{n_k}\delta_k^{-1}) &= \log_2(1.5\gamma_{n_k}) + \log_2(p_k^{-1}) + 2\alpha \log_2 n_k - \beta \log_2(\log_2 n_k) \\ &= \log_2(1.5C'\lambda^{n_k}) + \log_2(p_k^{-1}) + 2\alpha \log_2 n_k - \beta \log_2(\log_2 n_k)\ , \end{aligned}$$

and therefore for k big enough we obtain

$$\log_2(3\gamma_{n_k}\delta_k^{-1}) \leq 2\left[\log_2(p_k^{-1}) + An_k\right]. \tag{6.7}$$

The latter inequality and (6.6) give

$$p_k^{-1}\left[\log_2 n_k + \log_2(\log_2(3\gamma_{n_k}\delta_k^{-1}))\right]^{\beta}\left[\log_2(p_k^{-1}) + An_k\right]^{-\alpha} \leq 2^{\alpha}C_1\left[\log_2 n_k\right]^{\beta} n_k^{-\alpha},$$

which is equivalent to

$$p_k^{-1}\left[\log_2 n_k + \log_2(\log_2(3\gamma_{n_k}\delta_k^{-1}))\right]^{\beta} \leq 2^{\alpha}C_1\left[\frac{\log_2(p_k^{-1})}{n_k} + A\right]^{\alpha}\left[\log_2 n_k\right]^{\beta}. \tag{6.8}$$

Case 1: $\beta \geq 0$.
Note that since $\delta_k \to 0$ and $\gamma_{n_k} \to \infty$, for k big enough we have $\log_2(\log_2(3\gamma_{n_k}\delta_k^{-1})) > 0$. Since $\beta \geq 0$, we have

$$\left[\log_2 n_k\right]^{\beta} \leq \left[\log_2 n_k + \log_2(\log_2(3\gamma_{n_k}\delta_k^{-1}))\right]^{\beta},$$

and therefore it follows from (6.8) that

$$p_k^{-1} \leq 2^{\alpha}C_1\left[\frac{\log_2(p_k^{-1})}{n_k} + A\right]^{\alpha} < C[\log_2(p_k^{-1})]^{\alpha},$$

which contradicts the fact that p_k tends to zero (and thus $p_k^{-1} \to \infty$).

Case 2: $\beta < 0$.
In this case we rewrite (6.8) and use (6.7) to obtain

$$\begin{aligned} p_k^{-1} &\leq 2^{\alpha}C_1\left[\frac{\log_2(p_k^{-1})}{n_k} + A\right]^{\alpha}\left[1 + \frac{\log_2(\log_2(3\gamma_{n_k}\delta_k^{-1}))}{\log_2 n_k}\right]^{-\beta} \\ &\leq 2^{\alpha}C_1\left[\frac{\log_2(p_k^{-1})}{n_k} + A\right]^{\alpha}\left[1 + \frac{\log_2(2An_k + 2\log_2(p_k^{-1}))}{\log_2 n_k}\right]^{-\beta}. \end{aligned}$$

Next, we consider the following 2 cases.

Case 2.1: If for infinitely many values of k we have $p_k^{-1} \leq cn_k$, then the above inequality becomes

$$p_k^{-1} \leq C,$$

which contradicts with the fact that $p_k^{-1} \to \infty$ as $k \to \infty$.

Case 2.2: If for infinitely many values of k we have $p_k^{-1} \geq cn_k$, then the above inequality becomes

$$p_k^{-1} \leq C\left[\log_2(p_k^{-1})\right]^{\alpha}\left[\log_2(\log_2(p_k^{-1}))\right]^{-\beta},$$

which also contradicts with the fact that $p_k^{-1} \to \infty$ as $k \to \infty$.

To prove (ii) we repeat the argument for (i), namely, we assume that (ii) does not hold. Therefore there exists a strictly increasing sequence of integers $(n_k)_{k=1}^{\infty}$, such that

$$e_k := d_{n_k}^{\gamma_n}(\mathcal{K})[\log_2 n_k]^{\alpha} \to 0 \quad \text{as} \quad k \to \infty.$$

We write

$$d_{n_k}^{\gamma_{n_k}}(\mathcal{K}) = e_k[\log_2 n_k]^{-\alpha} < 2e_k[\log_2 n_k]^{-\alpha} =: \delta_k, \quad k = 1, 2, \dots, \tag{6.9}$$

and use Proposition 3.8 with $\eta_n = c_1(\log_2 n)^{-\alpha}$ to derive

$$c_1 \left[\log_2(n_k \log_2(3\gamma_{n_k}\delta_k^{-1}))\right]^{-\alpha} \leq 4e_k[\log_2 n_k]^{-\alpha}.$$

The latter inequality is equivalent to

$$e_k^{-1} \leq C_1 \left[1 + \frac{\log_2(\log_2(3\gamma_{n_k}\delta_k^{-1}))}{\log_2 n_k}\right]^{\alpha} \leq C_1 \left[1 + c\frac{\log_2(n_k + \log_2(e_k^{-1}))}{\log_2 n_k}\right]^{\alpha}, \quad C_1 = 4/c_1, \tag{6.10}$$

where we have used inequality similar to (6.7).

Case 1: If for infinitely many values of k we have $e_k^{-1} \leq cn_k$, then the above inequality becomes

$$e_k^{-1} \leq C,$$

which contradicts with the fact that $e_k^{-1} \to \infty$ as $k \to \infty$.

Case 2: If for infinitely many values of k we have $e_k^{-1} \geq cn_k$, then the above inequality becomes

$$e_k^{-1} \leq C \left[\log_2(e_k^{-1})\right]^{\alpha},$$

which also contradicts with the fact that $e_k^{-1} \to \infty$ as $k \to \infty$. □

6.3 Summary

In this section we summarize our results for the Lipschitz widths $d_n^{\gamma_n}(\mathcal{K})_X$ and give several examples. The following corollary holds.

Corollary 6.4. *Let $\mathcal{K} \subset X$ be a compact subset of a Banach space X, $n \in \mathbb{N}$, and $d_n^{\gamma_n}(\mathcal{K})_X$ be the Lipschitz width for $\mathcal{K}$ with Lipschitz constant $\gamma_n = C'\lambda^n$, where $C' > 0$ and $\lambda > 2$.*

(i) For $\alpha > 0$, $\beta \in \mathbb{R}$, we have

$$\varepsilon_n(\mathcal{K})_X \asymp \frac{[\log_2 n]^{\beta}}{n^{\alpha}} \quad \Rightarrow \quad d_n^{\gamma_n}(\mathcal{K})_X \asymp \frac{[\log_2 n]^{\beta}}{n^{2\alpha}}; \tag{6.11}$$

(ii) For $\alpha > 0$, we have

$$\varepsilon_n(\mathcal{K})_X \asymp \frac{1}{[\log_2 n]^{\alpha}} \quad \Rightarrow \quad d_n^{\gamma_n}(\mathcal{K})_X \asymp \frac{1}{[\log_2 n]^{\alpha}}. \tag{6.12}$$

Proof: We first prove (i). Let us assume that

$$\varepsilon_n(\mathcal{K})_X \leq C\frac{[\log_2 n]^{\beta}}{n^{\alpha}},$$

holds. After using (3.3) from Theorem 3.3, we obtain

$$d_n^{2^n \operatorname{rad}(\mathcal{K})}(\mathcal{K})_X \leq 2^{\beta}C\frac{[\log_2 n]^{\beta}}{n^{2\alpha}}. \tag{6.13}$$

We now fix n_0 such that $C'\lambda^n \geq 2^n \operatorname{rad}(\mathcal{K})$ for all $n \geq n_0$ (recall that $\lambda > 2$). We apply Remark 2.1 (iii) to derive

$$d_n^{\gamma_n}(\mathcal{K})_X \leq d_n^{2^n \operatorname{rad}(\mathcal{K})}(\mathcal{K})_X, \quad n \geq n_0. \tag{6.14}$$

Finally, it follows from (6.13) and (6.14) that

$$d_n^{\gamma_n}(\mathcal{K})_X \leq C' \frac{[\log_2 n]^\beta}{n^{2\alpha}}, \quad \text{for all } n,$$

provided the constant C' is chosen appropriately. The other direction in (6.11) is the statement of Theorem 6.3, (i). The proof of (ii) is similar and we omit it. □

Corollary 6.4 provides a tool for giving lower bounds on how well a compact set (model class) $\mathcal{K}$ can be approximated by a DNN that has the additional restriction that all weights and biases are from the unit ball of some norm $\|\cdot\|_{Y_{\tilde{n}}}$. A standard technique to obtain lower bounds is using the VC dimension, see [6], §5.9, and the references therein, which is restricted to the case when approximation error is measured in the norm $\|\cdot\|_{C(\Omega)}$. Note that Corollary 6.4 can be applied in the case of L_p approximation when $p \neq \infty$. For example, if $B_q^s(L_\tau(\Omega))$, $\Omega = [0,1]^d$, is any Besov space that lies above the Sobolev embedding line for $L_p(\Omega)$, then it is proven in [9] that

$$\varepsilon_n(U(B_q^s(L_\tau(\Omega))))_{L_p(\Omega)} \asymp n^{-s/d},$$

where $U(B_q^s(L_\tau(\Omega)))$ is the unit ball of $B_q^s(L_\tau(\Omega))$. Then, according to Theorem 6.1 and Corollary 6.4, we have

$$\operatorname{dist}(U(B_q^s(L_\tau(\Omega))), \Sigma_n)_{L_p(\Omega)} \geq d_{\tilde{n}}^{\gamma_n}(U(B_q^s(L_\tau(\Omega))))_{L_p(\Omega)} \geq C n^{-2s/d}.$$

In particular, we derive the estimate

$$\operatorname{dist}(U(B_q^s(L_\infty(\Omega))), \Sigma_n)_{C(\Omega)} \geq C n^{-2s/d},$$

which was proved in [6], see (5.18), in the case of general DNN. Note that here Σ_n is the family of functions that is generated by a DNN for which the parameters are from the unit ball of some norm $\|\cdot\|_{Y_{\tilde{n}}}$. This is the main difference between the framework considered here and the results in [16] or [13], which give upper bounds $\mathcal{O}(n^{-2s/d})$ (modulo logarithmic factors in some cases) for the error of approximation of the unit ball of $C^s([0,1]^d) \subset C([0,1]^d)$, $s > 0$, by ReLU DNN with no restrictions on the parameters used in the network. These upper bounds are optimal (up to logarithmic factor in some cases), see [6], section 8.7.1. It would be interesting to obtain upper bounds for the DNN approximation when there are restrictions imposed on its parameters. We are currently working on estimates from below on the Lipschitz widths associated with DNN whose parameters are allowed to grow as we increase the depth of the network.

Lastly, we want to mention that, in contrast to stable manifold widths, the Lipschitz widths do not shed a light on the numerical aspect of this approximation, that is, they do not give even a theoretical algorithm of how to design the approximant.

Acknowledgment: We would like to thank the referee for the essential remarks and suggestions which helped improve the quality and readability of the paper. Among other things, the referee pointed out to us a simpler proof of Lemma 2.8 and a simpler example that is discussed in Remark 3.9.

References

[1] Y. Benyamini, J. Lindenstrauss, *Geometric Nonlinear Functional Analysis*, Vol. 1, American Mathematical Society Colloquium Publications, **48**(2000), AMS, Providence, RI.

[2] K. Borsuk, *Drei Sätze über die n-dimensionale euklidische Sphäre* Fund. Math., **20** (1933), 177-191.

[3] B. Carl, *Entropy numbers, s-numbers, and eigenvalue problems,* J. Funct. Anal., **41** (1981), 290–306.

[4] B. Carl, I. Stephani, *Entropy, compactness and the approximation of operators,* Cambridge University Press, 1990.

[5] A Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Optimal Stable Nonlinear Approximation*, Journal of FoCM, to appear.

[6] R. DeVore, B. Hanin, and G. Petrova, *Neural Network Approximation*, Acta Numerica, **30** (2021), 327–444.

[7] R. DeVore, R. Howard, C. Micchelli, *Optimal nonlinear approximation*, Manuscripta Mathematica, **63**(4) (1989), 469–478.

[8] R. DeVore, G. Kyriazis, D. Leviatan, V. Tichomirov, *Wavelet compression and nonlinear-widths*, Advances in Computational Mathematics, **1**(2) (1993), 197–214.

[9] R. DeVore, R. Sharpley, *Besov spaces on domains in* $\mathbb{R}^d$, Transactions of the American Mathematical Society, **335** (2) (1993), 843–864.

[10] A.N. Kolmogorov, V.M. Tihomirov, *ε-entropy and ε-capacity of sets in function spaces,* (in Russian) Uspehi Mat. Nauk **14** no. 2 (86) (1959), 386.

[11] T. Kühn, *Entropy Numbers of General Diagonal Operators*, Rev. Mat. Complut., **18**(2) (2005), 479–491.

[12] G.G. Lorentz, M.Golitschek, Y.Makovoz, *Constructive Approximation, Advanced Problems, vol II* ,Grundlehren der mathematischen Wissenschaften 304, Springer Verlag, 1996.

[13] J. Lu, Z. Shen, H. Yang, S. Zhang, *Deep network approximation for smooth functions*, SIAM Journal on Mathematical Analysis, **53**(5) (2020), 5465–5506.

[14] A. Pinkus, *n-Widths in Approximation Theory,* Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer Verlag, 1985.

[15] P. Wojtaszczyk, *Banach spaces for analysts,* Cambridge studies in advanced mathematics 25, Cambridge University Press, 1991.

[16] D. Yarotsky *Optimal approximation of continuous functions by very deep relu networks*, Proceedings of the 31st Conference On Learning Theory, PMLR **75** (2018), 639–649.

Affiliations:
Guergana Petrova, Department of Mathematics, Texas A&M University, College Station, TX 77843, gpetrova@math.tamu.edu

Przemysław Wojtaszczyk, Institut of Mathematics Polish Academy of Sciences, ul. Śniadeckich 8, 00-656 Warszawa, Poland, wojtaszczyk@impan.pl