Check for
updates

Phase Transitions in Rate Distortion Theory and Deep Learning

Philipp Grohs[1,2] · Andreas Klotz[1] · Felix Voigtlaender[3,4]

Abstract

Rate distortion theory is concerned with optimally encoding signals from a given signal class $\mathcal{S}$ using a budget of R bits, as $R \to \infty$. We say that $\mathcal{S}$ *can be compressed at rate s* if we can achieve an error of at most $\mathcal{O}(R^{-s})$ for encoding the given signal class; the supremal compression rate is denoted by $s^*(\mathcal{S})$. Given a fixed coding scheme, there usually are *some* elements of $\mathcal{S}$ that are compressed at a higher rate than $s^*(\mathcal{S})$ by the given coding scheme; in this paper, we study the size of this set of signals. We show that for certain "nice" signal classes $\mathcal{S}$, a *phase transition* occurs: We construct a probability measure $\mathbb{P}$ on $\mathcal{S}$ such that for *every* coding scheme $\mathcal{C}$ and any $s > s^*(\mathcal{S})$, the set of signals encoded with error $\mathcal{O}(R^{-s})$ by $\mathcal{C}$ forms a $\mathbb{P}$-null-set. In particular, our results apply to all unit balls in Besov and Sobolev spaces that embed compactly into $L^2(\Omega)$ for a bounded Lipschitz domain Ω. As an application, we show that several existing sharpness results concerning function approximation using deep neural networks are

✉ Felix Voigtlaender
  felix.voigtlaender@ku.de

  Philipp Grohs
  philipp.grohs@univie.ac.at

  Andreas Klotz
  andreas.klotz@univie.ac.at

[1] Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

[2] Research Platform Data Science, University of Vienna, Vienna, Austria

[3] Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching bei München, Germany

[4] Present Address: Catholic University of Eichstätt–Ingolstadt, Mathematisch–Geographische Fakultät, Ostenstraße 26, 85072 Eichstätt, Germany

in fact *generically sharp*. In addition, we provide quantitative and non-asymptotic bounds on the probability that a random $f \in S$ can be encoded to within accuracy ε using R bits. This result is subsequently applied to the problem of approximately representing $f \in S$ to within accuracy ε by a (quantized) neural network with at most W nonzero weights. We show that for any $s > s^*(S)$ there are constants c, C such that, no matter what kind of "learning" procedure is used to produce such a network, the probability of success is bounded from above by $\min\left\{1, 2^{C \cdot W \lceil \log_2(1+W)\rceil^2 - c \cdot \varepsilon^{-1/s}}\right\}$.

1 Introduction

Let S be a *signal class*, that is, a relatively compact subset of a Banach space $(\mathbf{X}, \|\cdot\|_{\mathbf{X}})$. Rate distortion theory is concerned with the question of how well the elements of S can be encoded using a prescribed number R of bits. In many cases of interest, the best achievable coding error scales like R^{-s^*}, where s^* is the *optimal compression rate* of the signal class S. We show that a phase transition occurs: the set of elements $\mathbf{x} \in S$ that can be encoded using a *strictly larger* exponent than s^* is thin; precisely, it is a null-set with respect to a suitable probability measure $\mathbb{P}$. Crucially, the measure $\mathbb{P}$ is *independent* of the chosen coding scheme.

In order to rigorously formulate these results, we first review the needed notions of rate-distortion theory, see also [3,4,13,15]. For later use, we state the definitions here in the setting of general Banach spaces, although our main results only focus on the Hilbert space $L^2(\Omega)$.

1.1 A Crash Course in Rate Distortion Theory

To formalize the notion of encoding a signal class $S \subset \mathbf{X}$, we define the set $\mathrm{Enc}^R_{S,\mathbf{X}}$ of encoding/decoding pairs (E, D) of *code-length* $R \in \mathbb{N}$ as

$$\mathrm{Enc}^R_{S,\mathbf{X}} := \left\{ (E, D) \quad : \quad E : S \to \{0, 1\}^R \quad \text{and} \quad D : \{0, 1\}^R \to \mathbf{X} \right\}.$$

We are interested in choosing $(E, D) \in \mathrm{Enc}^R_{S,\mathbf{X}}$ such as to minimize the *(maximal) distortion* $\delta_{S,\mathbf{X}}(E, D) := \sup_{\mathbf{x} \in S} \|\mathbf{x} - D(E(\mathbf{x}))\|_{\mathbf{X}}$.

The intuition behind these definitions is that the encoder E converts any signal $\mathbf{x} \in S$ into a bitstream of code-length R (i.e., consisting of R bits), while the decoder D produces from a given bitstream $b \in \{0, 1\}^R$ a signal $D(b) \in \mathbf{X}$. The goal of *rate distortion theory* is to determine the minimal distortion that can be achieved by any encoder/decoder pair of code-length $R \in \mathbb{N}$. Typical results concerning the relation between code-length and distortion are formulated in an asymptotic sense: One assumes that for every code-length $R \in \mathbb{N}$, one is given an encoding/decoding pair

$(E_R, D_R) \in \mathrm{Enc}_{\mathcal{S},\mathbf{X}}^R$ and then, studies the asymptotic behavior of the corresponding distortion $\delta_{\mathcal{S},\mathbf{X}}(E_R, D_R)$ as $R \to \infty$.

We refer to a sequence $\big((E_R, D_R)\big)_{R \in \mathbb{N}}$ of encoding/decoding pairs as a *codec*, so that the set of all codecs is

$$\mathrm{Codecs}_{\mathcal{S},\mathbf{X}} := \underset{R \in \mathbb{N}}{\times} \mathrm{Enc}_{\mathcal{S},\mathbf{X}}^R.$$

For a given signal class $\mathcal{S}$ in a Banach space $\mathbf{X}$, it is of great interest to find an asymptotically optimal codec; that is, a sequence $\big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}}$ such that the asymptotic decay of $\big(\delta_{\mathcal{S},\mathbf{X}}(E_R, D_R)\big)_{R \in \mathbb{N}}$ is, in a sense, maximal. To formalize this, for each $s \in [0, \infty)$ define the class of subsets of $\mathbf{X}$ that *admit compression rate s* as

$$\mathrm{Comp}_{\mathbf{X}}^s := \Big\{ \mathcal{S} \subset \mathbf{X} : \exists \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}} : \delta_{\mathcal{S},\mathbf{X}}(E_R, D_R) \lesssim R^{-s} \Big\}.$$

For a given (bounded) signal class $\mathcal{S} \subset \mathbf{X}$, we aim to determine the *optimal compression rate* for $\mathcal{S}$ in $\mathbf{X}$, that is

$$s_{\mathbf{X}}^*(\mathcal{S}) := \sup \big\{ s \in [0, \infty) : \mathcal{S} \in \mathrm{Comp}_{\mathbf{X}}^s \big\} \in [0, \infty]. \tag{1.1}$$

Although the calculation of the quantity $s_{\mathbf{X}}^*(\mathcal{S})$ may appear daunting for a given signal class $\mathcal{S}$, there exists in fact a large body of literature addressing this topic. A landmark result in this area states that the JPEG2000 compression standard represents an optimal codec for the compression of piecewise smooth signals [26]. This optimality is typically stated more generally for the signal class $\mathcal{S} = \mathrm{Ball}\big(0, 1; B_{p,q}^\alpha(\Omega)\big)$, the unit ball in the Besov space $B_{p,q}^\alpha(\Omega)$, considered as a subset of $\mathbf{X} = \mathcal{H} = L^2(\Omega)$, for "sufficiently nice" bounded domains $\Omega \subset \mathbb{R}^d$; see [10].

For a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}}$, instead of considering the maximal distortion of $\mathcal{C}$ over the *entire* signal class $\mathcal{S}$, one can also measure the approximation rate that the codec $\mathcal{C}$ achieves for each *individual* $\mathbf{x} \in \mathcal{S}$. Precisely, the *class of elements with compression rate s under $\mathcal{C}$* is

$$\mathcal{A}_{\mathcal{S},\mathbf{X}}^s(\mathcal{C}) := \Big\{ \mathbf{x} \in \mathcal{S} \quad : \quad \sup_{R \in \mathbb{N}} \big[R^s \cdot \big\| \mathbf{x} - D_R(E_R(\mathbf{x})) \big\|_{\mathbf{X}} \big] < \infty \Big\}. \tag{1.2}$$

If the signal class $\mathcal{S}$ is "sufficiently regular"—for instance if $\mathcal{S}$ is compact and convex— then one can prove (see Proposition 3) that the following *dichotomy* is valid:

$$\begin{aligned} s < s_{\mathbf{X}}^*(\mathcal{S}) &\Longrightarrow \exists \mathcal{C} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}} \, \forall \mathbf{x} \in \mathcal{S} \, : \quad \mathbf{x} \in \mathcal{A}_{\mathcal{S},\mathbf{X}}^s(\mathcal{C}), \\ s > s_{\mathbf{X}}^*(\mathcal{S}) &\Longrightarrow \forall \mathcal{C} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}} \, \exists \mathbf{x}^* \in \mathcal{S} : \quad \mathbf{x}^* \notin \mathcal{A}_{\mathcal{S},\mathbf{X}}^s(\mathcal{C}). \end{aligned} \tag{1.3}$$

Thus, all signals in $\mathcal{S}$ can be approximated at any compression rate lower than the optimal rate for $\mathcal{S}$ using a *common* codec. Furthermore, for any approximation rate s larger than the optimal rate for $\mathcal{S}$, and for any codec $\mathcal{C}$, there exists some $\mathbf{x}^* = \mathbf{x}^*(s, \mathcal{C}) \in \mathcal{S}$ that is *not* compressed at rate s by $\mathcal{C}$.

Remark 1 (*Encoding/decoding schemes vs. discretization maps*) As the above considerations suggest, the crucial objects for our investigations are not the encoding/decoding pairs $(E, D) \in \mathrm{Enc}_{\mathcal{S}, \mathbf{X}}^R$, but the *distortion* they cause for each $\mathbf{x} \in \mathcal{S}$. Therefore, we could equally well restrict our attention to the *discretization map* $D \circ E : \mathcal{S} \to \mathbf{X}$, which has the crucial property $|\mathrm{range}(D \circ E)| \leq 2^R$. Conversely, given any (discretization) map $\Delta : \mathcal{S} \to \mathbf{X}$ with $|\mathrm{range}(\Delta)| \leq 2^R$, one can construct an encoding/decoding pair $(E, D) \in \mathrm{Enc}_{\mathcal{S}, \mathbf{X}}^R$, by choosing a surjection $D : \{0, 1\}^R \to \mathrm{range}(\Delta)$, and then setting

$$E : \quad \mathcal{S} \to \{0, 1\}^R, \quad \mathbf{x} \mapsto \mathrm{argmin}_{c \in \{0,1\}^R} \|\mathbf{x} - D(c)\|_{\mathbf{X}} ,$$

which ensures that $\|\mathbf{x} - D(E(\mathbf{x}))\|_{\mathbf{X}} \leq \|\mathbf{x} - \Delta(\mathbf{x})\|_{\mathbf{X}}$ for all $\mathbf{x} \in \mathcal{S}$. Thus, all our results could be rephrased in terms of such discretization maps rather than in terms of encoding/decoding pairs. For more details on this connection, see also Lemma 10. ◄

1.2 Our Contributions

1.2.1 Phase Transition

We improve on the dichotomy (1.3) by measuring the size of the class $\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$ of elements with compression rate s under the codec $\mathcal{C}$. Then, a phase transition occurs: the class of elements that cannot be encoded at a "larger than optimal" rate is *generic*. We prove this when the signal class is a ball in a Besov- or Sobolev space, as long as this ball forms a compact subset of $L^2(\Omega)$ for a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$.

More precisely, for each such signal class $\mathcal{S}$, we construct a probability measure $\mathbb{P}$ on $\mathcal{S}$ such that the compressibility exhibits a *phase transition* as in the following definition.

Definition 1 A Borel probability measure $\mathbb{P}$ on a subset $\mathcal{S}$ of a Banach space $\mathbf{X}$ *exhibits a compressibility phase transition* if it satisfies the following:

$$\begin{aligned} &\text{if } s < s_{\mathbf{X}}^*(\mathcal{S}) \text{ then } \exists\, \mathcal{C} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}} : \mathbb{P}\big(\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})\big) = 1; \\ &\text{if } s > s_{\mathbf{X}}^*(\mathcal{S}) \text{ then } \forall\, \mathcal{C} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}} : \mathbb{P}^*\big(\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})\big) = 0. \end{aligned} \qquad (1.4)$$

In this definition, since the set $\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$ is not necessarily measurable for $s > s_{\mathbf{X}}^*(\mathcal{S})$, instead of the measure $\mathbb{P}$, we use the associated *outer measure* $\mathbb{P}^*$. Generally, given a measure space $(\mathcal{S}, \mathscr{A}, \mu)$ the *outer measure* $\mu^* : 2^{\mathcal{S}} \to [0, \infty]$ induced by μ is defined as

$$\mu^*(M) := \inf \Big\{ \sum_{n=1}^{\infty} \mu(M_n) \quad : \quad (M_n)_{n \in \mathbb{N}} \subset \mathscr{A} \text{ with } M \subset \bigcup_{n=1}^{\infty} M_n \Big\}. \quad (1.5)$$

In general, μ^* is not a measure, but it is always *σ-subadditive*, meaning that it satisfies $\mu^*(\bigcup_{n=1}^{\infty} M_n) \leq \sum_{n=1}^{\infty} \mu^*(M_n)$ for arbitrary $M_n \subset \mathcal{S}$; see [14, Section 1.4]. Further-

more, it is easy to see for $M \in \mathscr{A}$ that $\mu^*(M) = \mu(M)$; that is, μ^* is an extension of μ.

As seen in (1.4), we are mostly interested in μ^*-*null-sets*; that is, subsets $N \subset \mathcal{S}$ satisfying $\mu^*(N) = 0$. This holds if and only if there is $N' \in \mathscr{A}$ satisfying $N \subset N'$ and $\mu(N') = 0$. Directly from the σ-subadditivity of μ^*, it follows that a countable union of μ^*-null-sets is again a μ^*-null-set.

We note that the first implication in (1.4) is always satisfied, as a consequence of (1.3). The second part of (1.4) states that for any $s > s_{\mathbf{X}}^*(\mathcal{S})$ and any codec $\mathcal{C}$, almost every $\mathbf{x} \in \mathcal{S}$ *cannot* be compressed by $\mathcal{C}$ at rate s. In other words, whenever $\mathbb{P}$ exhibits a compressibility phase transition on $\mathcal{S}$, *the property of not being compressible at a "larger than optimal" rate is a generic property.*

> **Remark 2** (i) We emphasize that the measure $\mathbb{P}$ in Definition 1 is required to satisfy the second property in (1.4) *universally* for *any* choice of codec $\mathcal{C}$.
> In fact, if $\mathbb{P}$ would be allowed to depend on $\mathcal{C}$, one could simply choose $\mathbb{P} = \delta_{\mathbf{x}}$, where $\mathbf{x} = \mathbf{x}(\mathcal{C}, s) \in \mathcal{S}$ is a single element that is not approximated at rate s by $\mathcal{C}$; for $s > s_{\mathbf{X}}^*(\mathcal{S})$ such an element exists under mild assumptions on $\mathcal{S}$; see Proposition 3.
> (ii) Any measure $\mathbb{P}$ satisfying (1.4) also satisfies $\mathbb{P}^*(\{\mathbf{x}\}) = 0$ (and hence $\mathbb{P}(\{\mathbf{x}\}) = 0$) for each $\mathbf{x} \in \mathcal{S}$, as can be seen by taking $\mathcal{C} = \left((E_R, D_R)\right)_{R \in \mathbb{N}}$ with $D_R : \{0, 1\}^R \to \mathcal{S}, c \mapsto \mathbf{x}$, so that $\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C}) = \{\mathbf{x}\}$ for all $s > 0$. Thus, any probability measure $\mathbb{P}$ exhibiting a compressibility phase transition is *atom free*.
> (iii) Measures satisfying (1.4) are quite special—in fact, Proposition 4 shows under mild assumptions on $\mathcal{S}$ that the set of measures *not* satisfying (1.4) is *generic* in the set of atom-free probability measures. ◀

Our first main result establishes the existence of critical measures for all Sobolev- and Besov balls (denoted by $\mathrm{Ball}(0, 1; W^{k,p}(\Omega; \mathbb{R}))$, respectively $\mathrm{Ball}(0, 1; B_{p,q}^\tau(\Omega; \mathbb{R}))$; see Appendix C) that are compact subsets of $L^2(\Omega)$:

Theorem 1 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Consider either of the following two settings:*

- *$\mathcal{S} := \mathrm{Ball}\left(0, 1; B_{p,q}^\tau(\Omega; \mathbb{R})\right)$ and $s^* := \frac{\tau}{d}$, where $p, q \in (0, \infty]$ and $\tau \in \mathbb{R}$ with $\tau > d \cdot (\frac{1}{p} - \frac{1}{2})_+$, or*
- *$\mathcal{S} := \mathrm{Ball}\left(0, 1; W^{k,p}(\Omega)\right)$ and $s^* := \frac{k}{d}$, where $p \in [1, \infty]$ and $k \in \mathbb{N}$ with $k > d \cdot (\frac{1}{p} - \frac{1}{2})_+$.*

In either case, $s_{L^2(\Omega)}^*(\mathcal{S}) = s^*$, and there exists a Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ that exhibits a compressibility phase transition as in Definition 1.

Proof This follows from Theorems 7, 8, and 4. □

Since Remark 2 shows that the measure $\mathbb{P}$ from the preceding theorem satisfies $\mathbb{P}(M) = 0$ for each countable set $M \subset \mathcal{S}$, we get the following strengthening of the dichotomy (1.3).

Corollary 1 *Suppose the assumptions of Theorem* 1 *are satisfied. Then, given any codec* $C \in \text{Codecs}_{\mathcal{S}, L^2(\Omega)}$ *the set* $\mathcal{S} \setminus \bigcup_{s > s^*} \mathcal{A}^s_{\mathcal{S}, L^2(\Omega)}(C)$, *which consists of all signals that* can*not be encoded by* C *at compression rate s for some $s > s^*$, is uncountable.*

In words, Corollary 1 states that for every codec the set of signals in $\mathcal{S}$ that cannot be approximated at any compression rate larger than the optimal rate for $\mathcal{S}$ is uncountable. In contrast, previous results (such as Proposition 3) only state the existence of a single such "badly approximable" signal.

1.2.2 Quantitative Lower Bounds

The statement of Theorem 1 is purely qualitative; it shows that the set of elements that are approximable at a "better than optimal rate" forms a null-set with respect to the measure $\mathbb{P}$. In fact, the measure $\mathbb{P}$ constructed in (the proof of) Theorem 1 satisfies a stronger, quantitative condition: If one randomly draws a function $f \sim \mathbb{P}$ using the probability measure $\mathbb{P}$, one can precisely bound the probability that a given encoding/decoding pair (E_R, D_R) of code-length R achieves a given error ε for f. To underline this probabilistic interpretation, we define, for any property Q of elements $f \in \mathcal{S}$,

$$\Pr(f \text{ satisfies } Q) := \mathbb{P}^*(\{f \in \mathcal{S} \colon f \text{ satisfies } Q\}), \qquad (1.6)$$

where $\mathbb{P}^*$ denotes the outer measure induced by $\mathbb{P}$.

Theorem 2 *Let $\mathcal{S}$ and s^* as in Theorem* 1. *Let the probability measure $\mathbb{P}$ on $\mathcal{S}$ as constructed in (the proof of) Theorem* 1, *and let us use the notation from Eq.* (1.6).
Then, for any $s > s^*$ there exist $c, \varepsilon_0 > 0$ (depending on $\mathcal{S}, s$) such that for arbitrary $R \in \mathbb{N}$ and $(E_R, D_R) \in \text{Enc}^R_{\mathcal{S}, L^2(\Omega)}$ it holds that

$$\Pr\big(\|f - D_R(E_R(f))\|_{L^2(\Omega)} \le \varepsilon\big) \le 2^{R - c \cdot \varepsilon^{-1/s}} \qquad \forall \varepsilon \in (0, \varepsilon_0).$$

Proof This follows from Theorem 4 and (the proof of) Theorem 1. □

Theorem 2 is interesting due to its nonasymptotic nature. Indeed, given a fixed budget of R bits and a desired accuracy ε, it provides a partial answer to the question:

How likely is one to succeed in describing a random $f \in \mathcal{S}$ to within accuracy ε using R bits?

Figure 1 provides an illustration of the phase transition behavior in dependence of ε and R; it graphically shows that the transition is quite sharp.

1.2.3 Lower Bounds for Neural Network Approximation

As an application, we draw a connection between the previously described results and the approximation properties of neural networks. In a nutshell, a neural network

$$E_s(R, \varepsilon) = \min\{1, 2^{R-(1/\varepsilon)^{1/s}}\} \text{ where } s = 2.002$$

$$\log_2(E_s(R, \varepsilon)) \text{ where } E_s(R, \varepsilon) = \min\{1, 2^{R-(1/\varepsilon)^{1/s}}\} \text{ and } s = 2.002$$
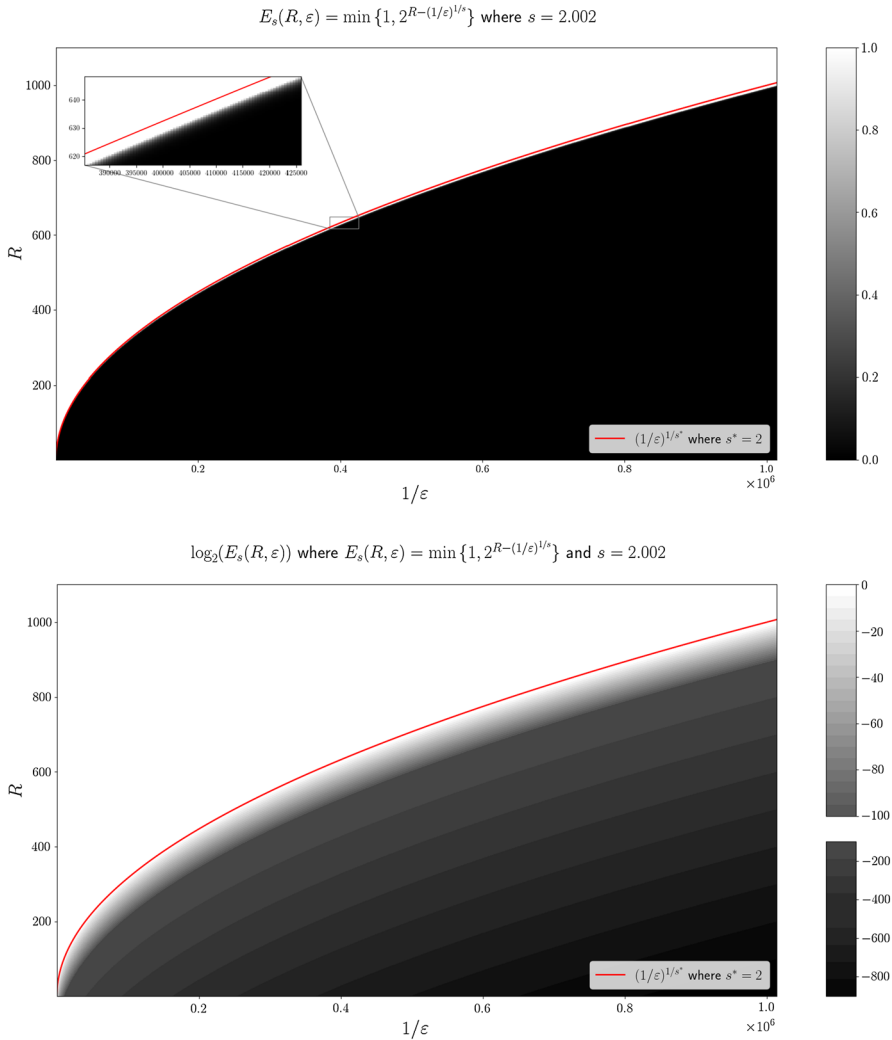
Fig. 1 For $\mathcal{S}$ a Sobolev or Besov ball, Theorem 2 provides bounds on the probability of being able to describe a random function $f \in \mathcal{S}$ to within accuracy ε using R bits. This probability is, for every $s > s^*$ and $\varepsilon \in (0, \varepsilon_0)$ (s^* denoting the optimal compression rate of $\mathcal{S}$), upper bounded by $E_s(R, \varepsilon) := \min\{1, 2^{R-c\cdot\varepsilon^{-1/s}}\}$. In this figure, we show two plots of the function E_s over the $(R, 1/\varepsilon)$-plane. Both grayscale plots show E_s for $s = 2.002 > s^* = 2$ and $c = 1$, while the red curve indicates the critical region where $R = (1/\varepsilon)^{1/s^*}$. We see that a sharp phase transition occurs in the sense that above and slightly below the critical curve $R = \varepsilon^{-1/2}$ (white area) the upper bound E_s does not rule out the possibility that it is always possible to describe $f \in \mathcal{S}$ to within accuracy ε using R bits; but even slightly below the critical curve (dark area) the bound E_s shows that such a compression is almost impossible. The sharpness of the phase transition is more clearly shown in the zoomed part of the figure. The bottom plot further illustrates the quantitative behavior by using a logarithmic colormap. Note that in the bottom plot two different colormaps are used for the range $[-100, 0]$ and the remaining range $[-1000, -100)$

alternatingly applies affine-linear maps and a so-called activation function $\varrho : \mathbb{R} \to \mathbb{R}$ that acts componentwise on vectors, meaning $\varrho((x_1, \ldots, x_m)) = (\varrho(x_1), \ldots, \varrho(x_m))$. More precisely, we will use the following mathematical formalization of (fully connected, feed forward) neural networks [29].

Definition 2 Let $d, L \in \mathbb{N}$ and $\mathbf{N} = (N_0, \ldots, N_L) \subset \mathbb{N}$ with $N_0 = d$. We say that $\mathbf{N}$ is a *network architecture*, where L describes the number of layers of the network and N_ℓ is the number of neurons in the ℓ-th layer.

A *neural network* (NN) with architecture $\mathbf{N}$ is a tuple

$$\Phi = ((A_1, b_1), \ldots, (A_L, b_L))$$

of matrices $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and bias vectors $b_\ell \in \mathbb{R}^{N_\ell}$. Given a function $\varrho : \mathbb{R} \to \mathbb{R}$, called the *activation function*, the mapping computed by the network Φ is defined as

$$R_\varrho \Phi : \quad \mathbb{R}^d \to \mathbb{R}^{N_L}, \quad R_\varrho \Phi = T_L \circ (\varrho \circ T_{L-1}) \circ \cdots \circ (\varrho \circ T_1),$$
$$\text{where} \quad T_\ell x = A_\ell x + b_\ell.$$

Here, we use the convention $\varrho((x_1, \ldots, x_m)) = (\varrho(x_1), \ldots, \varrho(x_m))$, i.e., ϱ acts componentwise on vectors.

The complexity of Φ is mainly described by the number $L(\Phi) := L$ of layers, the number $W(\Phi) := \sum_{\ell=1}^{L} \left(\|A_\ell\|_{\ell^0} + \|b_\ell\|_{\ell^0} \right)$ of weights (or connections), and the number $N(\Phi) := \sum_{\ell=0}^{L} N_\ell$ of neurons of Φ. Here, for a matrix or vector A, we denote by $\|A\|_{\ell^0}$ the number of nonzero entries of A. Furthermore, we set $d_{\text{in}}(\Phi) := N_0$ and $d_{\text{out}}(\Phi) := N_L$.

In addition to the number of layers, neurons, and weights of a network, we will also be interested in the complexity of the *individual* weights and biases of the network, i.e., of the entries of the (weight) matrices A_ℓ and the bias vectors b_ℓ. Thus, define

$$G_{\sigma, W} := \left[-W^{\sigma \lceil \log_2 W \rceil}, W^{\sigma \lceil \log_2 W \rceil} \right] \cap 2^{-\sigma \lceil \log_2 W \rceil^2} \mathbb{Z} \quad \text{for} \quad \sigma, W \in \mathbb{N}. \quad (1.7)$$

We say that Φ is (σ, W)-*quantized* if all entries of the matrices A_ℓ and the vectors b_ℓ belong to $G_{\sigma, W}$.

The set $G_{\sigma, W}$ in (1.7) contains all real numbers that belong to the grid $2^{-\sigma \lceil \log_2 W \rceil^2} \mathbb{Z}$ and simultaneously to the interval $\left[-W^{\sigma \lceil \log_2 W \rceil}, W^{\sigma \lceil \log_2 W \rceil} \right]$; thus, the grid gets arbitrarily fine and the interval gets arbitrarily large as $W \to \infty$, where the parameter σ determines how fast this happens. We note that *in applications one necessarily deals with quantized NNs* due to the necessity to store and process the weights on a digital computer. Regarding function approximation by such quantized neural networks, we have the following result:

Theorem 3 *Let* $\varrho : \mathbb{R} \to \mathbb{R}$ *be measurable and let* $d, \sigma \in \mathbb{N}$. *For* $W \in \mathbb{N}$, *define*

$$\mathcal{NN}_{d,W}^{\sigma, \varrho} := \left\{ R_\varrho \Phi \ : \ \begin{matrix} \Phi \text{ is a } (\sigma, W)\text{-quantized NN} \\ \text{and } W(\Phi) \le W, \, d_{\text{in}}(\Phi) = d, \, d_{\text{out}}(\Phi) = 1 \end{matrix} \right\}.$$

Let $\mathcal{S}$, s^*, and $\mathbb{P}$ as in Theorem 1. Then, using the notation (1.6), the following hold:

1. *There exists $C = C(d, \sigma) \in \mathbb{N}$ such that for each $s > s^*$ there are $c, \varepsilon_0 > 0$ satisfying*

$$\Pr\left(\min_{g \in \mathcal{NN}^{\sigma,\varrho}_{d,W}} \|f - g\|_{L^2(\Omega)} \leq \varepsilon\right) \leq 2^{C \cdot W \lceil \log_2(1+W)\rceil^2 - c \cdot \varepsilon^{-1/s}} \quad \forall \varepsilon \in (0, \varepsilon_0) \quad (1.8)$$

In fact, one can choose $C = 4 + 4\lceil\log_2(4ed)\rceil + 8\sigma$.

2. *If we define $W_{\sigma,\varrho}(f; \varepsilon) \in \mathbb{N} \cup \{\infty\}$ by*

$$W_{\sigma,\varrho}(f; \varepsilon) := \inf\left\{W \in \mathbb{N} \; : \; \exists g \in \mathcal{NN}^{\sigma,\varrho}_{d,W} \text{ such that } \|f - g\|_{L^2(\Omega)} \leq \varepsilon\right\}$$

and furthermore set

$$\mathcal{A}^*_{\mathcal{NN},\varrho} := \left\{f \in \mathcal{S} \; : \; \begin{matrix} \exists \tau \in (0, \frac{1}{s^*}), \sigma \in \mathbb{N}, C > 0 \\ \forall \varepsilon \in (0, 1) : W_{\sigma,\varrho}(f; \varepsilon) \leq C \cdot \varepsilon^{-\tau} \end{matrix}\right\},$$

then $\mathbb{P}^*(\mathcal{A}^*_{\mathcal{NN},\varrho}) = 0$.

Proof The proof of this theorem is deferred to Appendix F. ☐

In a nutshell, Eq. (1.8) states the following: If one draws a random function f from $\mathbb{P}$, then with probability at least $1 - 2^{C \cdot W \lceil \log_2(1+W)\rceil^2 - c \cdot \varepsilon^{-1/s}}$, the function f will have L^2 distance at least ε to *every* network from the class $\mathcal{NN}^{\sigma,\varrho}_{d,W}$. Consequently, Eq. (1.8) implies that the network size W has to scale at least like $W \gtrsim \varepsilon^{-1/s^*}$ (up to log factors) to succeed with high probability if $\mathcal{S}$ is a Sobolev- or Besov ball with optimal exponent $s^* = s^*_{L^2(\Omega)}(\mathcal{S})$. In particular, if one uses *any* learning procedure Learn $: \mathcal{S} \to \mathcal{NN}^{\sigma,\varrho}_{d,W}$— not necessarily a typical algorithm like (stochastic) gradient descent— and hopes to achieve $\|f - \text{Learn}(f)\|_{L^2(\Omega)} \leq \varepsilon$, then Eq. (1.8) provides an upper bound on the probability of success.

Remark 3 (*Further discussion of Theorem 3*) It might seem peculiar at first sight that the depth of the approximating networks does not seem to play a role in Theorem 3, even though deeper networks should intuitively have better approximation properties. To understand how this can be, note that the theorem only provides a *hardness* result: it gives a lower bound on how well *arbitrarily deep* (quantized) networks can approximate functions from the class $\mathcal{S}$.

Whether this lower bound is sharp (i.e., whether the critical rate s^* can be attained using suitable networks) then depends on the chosen activation function and the network depth. For instance, for approximating C^k functions, a result by Mhaskar [27] shows that *shallow* networks with *smooth*, non-polynomial activation functions already attain the optimal rates. For networks with the (non-smooth) ReLU activation function $\varrho(x) = \max\{0, x\}$, on the other hand, it is known that (somewhat) deep networks are necessary to attain the optimal rates; more precisely, one needs ReLU networks of depth $\mathcal{O}(1 + \frac{k}{d})$ to achieve the optimal approximation rates for C^k functions on $[0, 1]^d$; see [29, Theorem C.6] and [31,32].

In fact, the optimal rates predicted by Theorem 3 are attained (up to log factors) by sufficiently deep ReLU networks and domain $\Omega = [0, 1]^d$. Precisely, there exist $C = C(\mathcal{S}) > 0$ and $\sigma = \sigma(\mathcal{S}) \in \mathbb{N}$ such that

$$\forall f \in \mathcal{S} \quad \forall \varepsilon \in (0, \tfrac{1}{2}): \qquad W_{\sigma,\varrho}(f; \varepsilon) \leq C \cdot \varepsilon^{-1/s^*} \cdot \log_2(1/\varepsilon) \lesssim \varepsilon^{-\tau},$$

where $\tau \in (0, \frac{1}{s^*})$ is arbitrary. This follows from results in [13,34]. Since the details are mainly technical, the proof is deferred to Appendix F. We remark that by similar arguments as in [13,34], one can also prove the sharpness for other activation functions than the ReLU and other domains than $[0, 1]^d$.   ◄

1.3 Related Literature and Concepts

1.3.1 Minimax Optimality in Approximation Theory

Many (optimality) results in approximation theory are formulated in a *minimax sense*, meaning that one precisely characterizes the asymptotic decay of

$$d_{\mathbf{X}}(\mathcal{S}, M_n) = \sup_{f \in \mathcal{S}} \inf_{g \in M_n} \|f - g\|_{\mathbf{X}},$$

where $\mathcal{S} \subset \mathbf{X}$ is the class of signals to be approximated, and $M_n \subset \mathbf{X}$ contains all functions "of complexity n," for example polynomials of degree n or shallow neural networks with n neurons, etc. As recent examples of such results related to neural networks, we mention [4,29,41].

A minimax lower bound of the form $d_{\mathbf{X}}(\mathcal{S}, M_n) \gtrsim n^{-s^*}$, however, only makes a claim about the possible *worst case* of approximating elements $f \in \mathcal{S}$. In other words, such an estimate in general only guarantees that there is *at least one* "hard to approximate" function $f^* \in \mathcal{S}$ satisfying $\inf_{g \in M_n} \|f^* - g\|_{\mathbf{X}} \gtrsim n^{-s}$ for each $s > s^*$, but *nothing is known about how "massive" this set of "hard to approximate" functions is*, or about the "average case."

1.3.2 Results Quantifying the "mass" of Hard-To-Approximate Functions

The first paper to address this question—and one of the main sources of inspiration for the present paper—is [24]. In that paper, Maiorov, Meir, and Ratsaby consider essentially the "L^2-Besov-space type" signal class $\mathcal{S} = \mathcal{S}_r$ of functions $f \in L^2(\mathbb{B}_d)$ (with $\mathbb{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$) that satisfy

$$\mathrm{dist}_{L^2}(f, \mathscr{P}_{2^N}) \leq 2^{-rN},$$

where $\mathscr{P}_K = \mathrm{span}\{x^\alpha : \alpha \in \mathbb{N}_0^d \text{ with } |\alpha| \leq K\}$ denotes the space of d-variate polynomials of degree at most K. On this signal class, they construct a probability measure

$\mathbb{P}$ such that given the subset of functions

$$M_n = \left\{ x \mapsto \sum_{i=1}^{n} g_i(\langle a_i, x \rangle) : a_i \in \mathbb{S}^{d-1} \text{ and } g_i \in L^2([-1, 1]) \right\},$$

one obtains the minimax asymptotic $d_{L^2}(\mathcal{S}_r, M_n) \asymp n^{-r/(d-1)}$, but furthermore there exists $c > 0$ such that

$$\mathbb{P}\left(\left\{ f \in \mathcal{S}_r : \ \text{dist}_{L^2}(f, M_n) \geq c \cdot n^{-r/(d-1)} \right\} \right) \geq 1 - e^{-c \cdot n^{d/(d-1)}}.$$

In other words, the measure of the set of functions for which the minimax asymptotic is sharp tends to 1 for $n \to \infty$. In this context, we would also like to mention the recent article [23], in which the results of [24] are extended to cover more general signal classes and approximation in stronger norms than the L^2 norm.

While we draw heavily on the ideas from [24] for the construction of the measure $\mathbb{P}$ in Theorem 1, it should be noted that we are interested in phase transitions for general encoding/decoding schemes, while [23,24] exclusively focus on approximation using the ridge function classes M_n.

1.3.3 Baire Category

Extending the scale of C^k spaces ($k \in \mathbb{N}_0$) to the scale of Hölder-spaces C^β, $\beta \geq 0$, it is well-known that C^β is of first category (or *meager*) in C^η if $\beta > \eta$. Similarly, under mild regularity conditions on the signal class $\mathcal{S} \subset \mathbf{X}$, one can show for every codec $\mathcal{C} \in \text{Codecs}_{\mathcal{S}, \mathbf{X}}$ that the set $\bigcup_{s > s^*} \mathcal{A}^s_{\mathcal{S}, \mathbf{X}}(\mathcal{C})$ of signals $\mathbf{x} \in \mathcal{S}$ that are encodable by $\mathcal{C}$ at a "better than optimal" rate $s > s^* = s^*_{\mathbf{X}}(\mathcal{S})$ is meager in $\mathcal{S}$; for instance, this holds if $\mathcal{S}$ is compact and convex; see Proposition 3.

Thus, if one could construct a Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ satisfying $\mathbb{P}^*(M) = 0$ for *every* set $M \subset \mathcal{S}$ that is meager in $\mathcal{S}$, then $\mathbb{P}$ would automatically satisfy the phase transition (1.4). In most cases, however, it turns out that no such measure exists.

Indeed, assuming that such a measure exists and that $\mathcal{S}$ has no isolated points, every singleton $\{\mathbf{x}\} \subset \mathcal{S}$ is meager in $\mathcal{S}$, so that $\mathbb{P}(\{\mathbf{x}\}) = 0$ for every $\mathbf{x} \in \mathcal{S}$. Therefore, if $\mathcal{S}$, equipped with the topology induced by $\| \cdot \|_{\mathbf{X}}$, has a base whose cardinal has measure zero (see below for details), then [28, Theorem 16.5] shows that one can write $\mathcal{S} = N \cup M$, where $M \subset \mathcal{S}$ is meager and $N \subset \mathcal{S}$ satisfies $\mathbb{P}(N) = 0$, leading to $1 = \mathbb{P}(\mathcal{S}) \leq \mathbb{P}^*(N) + \mathbb{P}^*(M) = 0$, a contradiction. Regarding the assumption on the existence of a base whose cardinal has measure zero, note that if $\mathcal{S}$ is relatively compact (which always holds if $s^*_{\mathbf{X}}(\mathcal{S}) > 0$), then $\mathcal{S}$ is separable and hence has a countable base, whose cardinal thus has measure zero; see [28, Page 63]. In summary, *if $\mathcal{S} \subset \mathbf{X}$ is relatively compact and has no isolated points, then there does* not *exist a Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ satisfying $\mathbb{P}^*(M) = 0$ for every set $M \subset \mathcal{S}$ of first category.*

One could further ask how "special" measures satisfying the phase transition (1.4) are; more precisely: *Is the set of probability measures $\mathbb{P}$ satisfying (1.4) generic in the set of (atom-free) probability measures?* This is not the case; in fact, Proposition 4

shows under very mild assumptions that if one equips the set of atom-free probability measures on $\mathcal{S}$ with the total variation metric, then the set of measures satisfying (1.4) is *meager*. In other words, the set of measures *not* satisfying (1.4) is generic as a subset of all atom-free Borel probability measures on $\mathcal{S}$.

1.3.4 Small ball probabilities and Gaussian measures

An important notion that we introduce and study in this article are measures of *logarithmic growth order* s_0, which are measures satisfying a certain *small ball condition*; see Eq. (1.9) below. Such small ball conditions have been extensively studied in the theory of Gaussian measures; see for instance [21,22]. An important result in that area of research shows that the small ball probability of a Gaussian measure μ is closely related to the behavior of the entropy numbers of the unit ball K_μ of a certain reproducing kernel Hilbert space H_μ associated with μ.

In seeming similarity, we are concerned with constructing a probability measure $\mathbb{P}$ supported on a given set $\mathcal{S}$ such that $\mathbb{P}$ satisfies a certain small ball property, depending on the optimal exponent $s_\mathcal{H}^*(\mathcal{S})$ of $\mathcal{S}$, which is intimately related to the behavior of the entropy numbers of $\mathcal{S}$. As far as we can tell, however, the similarity is only superficial, meaning that the main similarity is simply that both results are concerned with measures satisfying small ball properties and the relation to entropy numbers.

To see that the questions considered in [21,22] are different from the ones studied here, note that the Gaussian measures considered in [21,22] are *not* supported on K_μ and furthermore that the entropy numbers of K_μ always satisfy $H(K_\mu, \varepsilon) \in o(\varepsilon^{-2})$ as $\varepsilon \to 0$, a property that is in general *not* shared by the signal classes $\mathcal{S} = \mathrm{Ball}\left(0,1; B_{p,q}^\tau(\Omega;\mathbb{R})\right)$ and $\mathcal{S} := \mathrm{Ball}\left(0,1; W^{k,p}(\Omega)\right)$ that we consider.

Finally, we mention that a (non-trivial) modification of our proof shows that the measure $\mathbb{P}$ constructed in Theorem 1 can be chosen to be (the restriction of) a suitable centered Gaussian measure.

1.3.5 Optimality Results for Neural Network Approximation

We emphasize that our lower bounds for neural network approximation consider networks with *quantized weights*, as in [4,29]. The main reason is that without such an assumption, even networks with two hidden layers and *a fixed number of neurons* can approximate any function arbitrarily well if the activation function is chosen suitably; see [25, Theorem 4] and [40]. Moreover, even if one considers the popular ReLU activation function, it was recently observed that the optimal approximation rates for networks with quantized weights can in fact be *doubled* by using arbitrarily deep ReLU networks with highly complex (non-quantized) weights [41].

1.4 Structure of the Paper and Proof Ideas

In Sect. 2, we introduce and study a class of probability measures with a certain growth behavior. More precisely, we say that $\mathbb{P}$ is of *logarithmic growth order* s_0 on $\mathcal{S} \subset \mathbf{X}$

if for each $s > s_0$ there exist $\varepsilon_0 = \varepsilon_0(s) > 0$ and $c = c(s) > 0$ satisfying

$$\mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})\big) \le 2^{-c \cdot \varepsilon^{-1/s}} \qquad \forall \mathbf{x} \in \mathbf{X} \text{ and } \varepsilon \in (0, \varepsilon_0). \qquad (1.9)$$

Here, as in the rest of the paper, $\mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})$ is the *closed* ball around $\mathbf{x}$ of radius ε with respect to $\|\cdot\|_{\mathbf{X}}$. A probability measure has *critical growth* if its logarithmic growth order equals the optimal compression rate $s_{\mathbf{X}}^*(\mathcal{S})$. We show in particular that every critical probability measure exhibits a compressibility phase transition as in Definition 1, and we show how critical probability measures can be transported from one set to another.

Intuitively, the natural way to construct a probability measure $\mathbb{P}$ satisfying (1.9) is to make the measure "as uniform as possible," so that each ball $\mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})$ contains roughly the same *(small)* volume. At first sight, it is thus natural to choose $\mathbb{P}$ to be *translation invariant*. It is well-known, however, (see, e.g., [17, Page 218]) that there *does not exist* any non-trivial locally finite, translation invariant measure on an infinite-dimensional, separable Banach space.

Therefore, we construct a measure $\mathbb{P}$ satisfying (1.9) in the setting of certain *sequence spaces*, where we can exploit the product structure of the signal class $\mathcal{S}$ to make the measure as uniform as possible—this technique was pioneered in [24]. More precisely, in Sect. 3, we study the sequence spaces $\ell_{\mathcal{P},\alpha}^{p,q}$, which are essentially the coefficient spaces associated with Besov spaces. By modifying the construction given in [24], we construct probability measures of critical growth on the unit balls $\mathcal{S}_{\mathcal{P},\alpha}^{p,q}$ of the spaces $\ell_{\mathcal{P},\alpha}^{p,q}$, for the range of parameters for which the embedding $\ell_{\mathcal{P},\alpha}^{p,q} \hookrightarrow \ell^2$ is compact. For the case $q = \infty$, we directly use the product structure of the spaces to construct the measure; the construction for the case $q < \infty$ uses the measure from the case $q = \infty$, combined with a technical trick (namely, introducing an additional weight).

The construction of critical measures on the unit balls of Besov and Sobolev spaces is then accomplished in Sect. 4, by using wavelet systems to transfer the critical measure from the sequence spaces to the function spaces. This makes heavy use of the transfer results established in Sect. 2.

A host of more technical proofs are deferred to the appendices.

1.5 Notation

We write $\mathbb{N} := \{1, 2, 3, \dots\}$ for the set of natural numbers, and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$ for the natural numbers including zero. The number of elements of a set M is denoted by $|M| \in \mathbb{N}_0 \cup \{\infty\}$. For $n \in \mathbb{N}_0$, we define $[n] := \{k \in \mathbb{N}: k \le n\}$; in particular, $[0] = \emptyset$.

For $x \in \mathbb{R}$, we write $x_+ := \max\{0, x\}$ and $x_- := (-x)_+ = \max\{0, -x\}$.

We assume all vector spaces to be over $\mathbb{R}$, unless explicitly stated otherwise.

For a given (quasi-)normed vector space $(\mathbf{X}, \|\cdot\|)$, we denote the *closed ball* of radius $r \ge 0$ around $\mathbf{x} \in \mathbf{X}$ by

$$\mathrm{Ball}(\mathbf{x}, r) := \mathrm{Ball}(\mathbf{x}, r; \mathbf{X}) := \big\{\mathbf{y} \in \mathbf{X}: \|\mathbf{y} - \mathbf{x}\| \le r\big\}.$$

If we want to emphasize the quasi-norm (for example, if multiple quasi-norms are considered on the same space $\mathbf{X}$), we write $\mathrm{Ball}(\mathbf{x}, r; \|\cdot\|)$ instead.

We say that a subset $\mathcal{S}$ of a topological space $\mathbf{X}$ is *relatively compact*, if the closure $\overline{\mathcal{S}}$ of $\mathcal{S}$ in $\mathbf{X}$ is compact. If $\mathbf{X}$ is a complete metric space with metric d, then this holds if and only if $\mathcal{S}$ is *totally bounded*, meaning that for every $\varepsilon > 0$ there exist finitely many $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbf{X}$ satisfying $\mathcal{S} \subset \bigcup_{i=1}^{N} \{\mathbf{x} \in \mathbf{X} \colon d(\mathbf{x}, \mathbf{x}_i) \leq \varepsilon\}$; see [14, Theorem 0.25].

For an index set $\mathcal{I}$ and an integrability exponent $p \in (0, \infty]$, the sequence space $\ell^p(\mathcal{I}) \subset \mathbb{R}^{\mathcal{I}}$ is

$$\ell^p(\mathcal{I}) = \left\{\mathbf{x} = (x_i)_{i \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}} \quad : \quad \|\mathbf{x}\|_{\ell^p} < \infty\right\},$$

where $\|\mathbf{x}\|_{\ell^p} := \left(\sum_{i \in \mathcal{I}} |x_i|^p\right)^{1/p}$ if $p < \infty$, while $\|\mathbf{x}\|_{\ell^\infty} := \sup_{i \in \mathcal{I}} |x_i|$.

A Comment on Measurability: Given a (not necessarily measurable) subset $M \subset \mathbf{X}$ of a Banach space $\mathbf{X}$, we will always equip M with the *trace σ-algebra*

$$M \cap \mathscr{B}_{\mathbf{X}} := \{M \cap B \colon B \in \mathscr{B}_{\mathbf{X}}\}$$

of the Borel σ-algebra $\mathscr{B}_{\mathbf{X}}$. A *Borel measure* on M is then a measure defined on $M \cap \mathscr{B}_{\mathbf{X}}$.

Note that if $(\Omega, \mathscr{A})$ is any measurable space, then $\Phi \colon \Omega \to M$ is measurable if and only if it is measurable considered as a map $\Phi \colon \Omega \to (\mathbf{X}, \mathscr{B}_{\mathbf{X}})$.

2 General Results on Phase Transitions in Banach Spaces

In this section, we establish an abstract version of the phase transition considered in (1.4) for signal classes in general Banach spaces and a class of measures that satisfy a uniform growth property that we term "critical" (see Definition 3). We will show in Sect. 2.1 that such critical measures automatically induce a phase transition behavior. We furthermore show in Sect. 2.2 that criticality is preserved under pushforward by "nice" mappings. The *existence* of critical measures is by no means trivial; quite the opposite, their construction for a class of sequence spaces in Sect. 3—and for Besov and Sobolev spaces on domains in Sect. 4—constitutes an essential part of the present article.

2.1 Measures of Logarithmic Growth

Definition 3 Let $\mathcal{S} \neq \emptyset$ be a subset of a Banach space $\mathbf{X}$, and let $s_0 \in [0, \infty)$.

A Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ has *(logarithmic) growth order s_0 (with respect to $\mathbf{X}$)* if for every $s > s_0$, there are constants $\varepsilon_0, c > 0$ (depending on $s, s_0, \mathbb{P}, \mathcal{S}, \mathbf{X}$) such that

$$\mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})\big) \leq 2^{-c \cdot \varepsilon^{-1/s}} \qquad \forall \, \mathbf{x} \in \mathbf{X} \text{ and } \varepsilon \in (0, \varepsilon_0). \tag{2.1}$$

We say that $\mathbb{P}$ is *critical for $\mathcal{S}$ (with respect to $\mathbf{X}$)* if $\mathbb{P}$ has logarithmic growth order $s_{\mathbf{X}}^*(\mathcal{S})$, with the optimal compression rate $s_{\mathbf{X}}^*(\mathcal{S})$ as defined in Eq. (1.1).

Remark (1) If $\mathbb{P}$ has growth order s_0, then $\mathbb{P}$ also has growth order σ, for arbitrary $\sigma > s_0$.

(2) Instead of Property (2.1), one could *equivalently* only require that the measure of balls *centered at points of* $\mathcal{S}$ decays rapidly. More precisely, (2.1) is valid for certain $\varepsilon_0, c > 0$ (depending on $s, s_0, \mathbb{P}, \mathcal{S}, \mathbf{X}$) if and only if there exist $\varepsilon_1, \omega > 0$ (depending on $s, s_0, \mathbb{P}, \mathcal{S}, \mathbf{X}$) satisfying

$$\mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})\big) \leq 2^{-\omega \cdot \varepsilon^{-1/s}} \qquad \forall\, \mathbf{x} \in \mathcal{S} \text{ and } \varepsilon \in (0, \varepsilon_1). \qquad (2.2)$$

Indeed, if (2.1) holds, then so does (2.2) (with $\omega = c$ and $\varepsilon_1 = \varepsilon_0$). Conversely, suppose (2.2) holds for certain $\varepsilon_1, \omega > 0$. Set $c := \omega / 2^{1/s} > 0$ and $\varepsilon_0 := \varepsilon_1 / 2$ and let $\varepsilon \in (0, \varepsilon_0)$ and $\mathbf{x} \in \mathbf{X}$ be arbitrary. First, if $\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X}) = \emptyset$, then trivially $\mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})\big) = 0 \leq 2^{-c \cdot \varepsilon^{-1/s}}$. Otherwise, there exists $\mathbf{y} \in \mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})$ and then $\mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X}) \subset \mathrm{Ball}(\mathbf{y}, 2\varepsilon; \mathbf{X})$ and $2\varepsilon < \varepsilon_1$. Hence, (2.2) shows

$$\mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{x}, \varepsilon; \mathbf{X})\big) \leq \mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{y}, 2\varepsilon; \mathbf{X})\big) \leq 2^{-\omega \cdot (2\varepsilon)^{-1/s}} = 2^{-c \cdot \varepsilon^{-1/s}}$$

by our choice of c. ◄

The motivation for considering the growth order of a measure is that it leads to bounds regarding the measure of elements $\mathbf{x} \in \mathcal{S}$ that are well-approximated by a given codec; see Eq. (2.3) below. Furthermore, as we will see in Corollary 2, if $\mathbb{P}$ is a probability measure of growth order s_0, then necessarily $s_0 \geq s_{\mathbf{X}}^*(\mathcal{S})$, so *critical measures have the minimal possible growth order*.

The following theorem summarizes our main structural results, showing that critical measures always exhibit a compressibility phase transition.

Theorem 4 *Let the signal class $\mathcal{S}$ be a subset of the Banach space $\mathbf{X}$, let $\mathbb{P}$ be a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $\mathbf{X}$, and set $s^* := s_{\mathbf{X}}^*(\mathcal{S})$. Then, the following hold:*

(i) *Let $s > s^*$ and let $c = c(s) > 0$ and $\varepsilon_0 = \varepsilon_0(s)$ as in Eq. (2.1). Then, for any $R \in \mathbb{N}$ and $(E_R, D_R) \in \mathrm{Enc}_{\mathcal{S}, \mathbf{X}}^R$, we have*

$$\mathrm{Pr}\big(\|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq \varepsilon\big) \leq 2^{R - c \cdot \varepsilon^{-1/s}} \qquad \forall\, \varepsilon \in (0, \varepsilon_0), \qquad (2.3)$$

where we use the notation from Eq. (1.6).

(ii) *For every $s > s^*$ and every codec $\mathcal{C} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}}$, the set $\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$ is a $\mathbb{P}^*$-null-set:*

$$\mathrm{Pr}\left(\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})\right) = 0.$$

(iii) *For every $0 \leq s < s^*$, there exists a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}}$ with distortion*

$$\delta_{\mathcal{S}, \mathbf{X}}(E_R, D_R) \leq C \cdot R^{-s} \qquad \forall\, R \in \mathbb{N},$$

for a constant $C = C(s, \mathcal{C}) > 0$. In particular, the set of s-compressible signals $\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C})$ defined in Eq. (1.2) satisfies $\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C}) = \mathcal{S}$ and hence $\mathbb{P}(\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C})) = 1$.

Remark (1) Note that the theorem does not make any statement about the case $s = s^*$. In this case, the behavior depends on the specific choices of $\mathcal{S}$ and $\mathbb{P}$.

(2) As noted above, the question of the existence of a critical probability measure $\mathbb{P}$ is nontrivial. ◀

The proof of Theorem 4 is divided into several auxiliary results. Part (i) is contained in the following lemma.

Lemma 1 *Let $\mathcal{S} \neq \emptyset$ be a subset of a Banach space $\mathbf{X}$, and let $\mathbb{P}$ be a Borel probability measure on $\mathcal{S}$ that is of logarithmic growth order $s_0 \geq 0$ with respect to $\mathbf{X}$.*

Let $s > s_0$ and let $c = c(s) > 0$ and $\varepsilon_0 = \varepsilon_0(s)$ as in Eq. (2.1). Then, for any $R \in \mathbb{N}$ and $(E_R, D_R) \in \mathrm{Enc}^R_{\mathcal{S},\mathbf{X}}$, we have

$$\mathbb{P}^*\big(\{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq \varepsilon\}\big) \leq 2^{R - c \cdot \varepsilon^{-1/s}} \qquad \forall \varepsilon \in (0, \varepsilon_0).$$

Further, for any given $s > s_0$ and $K > 0$ there exists a minimal code-length $R_0 = R_0(s, s_0, K, \mathbb{P}, \mathcal{S}, \mathbf{X}) \in \mathbb{N}$ such that every $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}}$ satisfies

$$\mathbb{P}^*\big(\{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq K \cdot R^{-s}\}\big) \leq 2^{-R} \qquad \forall R \geq R_0. \qquad (2.4)$$

Remark The lemma states that the measure of the subset of points $\mathbf{x} \in \mathcal{S}$ with approximation error $\mathcal{E}_R(\mathbf{x}) := \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}}$ satisfying $\mathcal{E}_R(\mathbf{x}) \leq K \cdot R^{-s}$ for some $s > s_0$ decreases exponentially with R. In fact, the proof shows that the approximation error is decreasing asymptotically *superexponentially*. ◀

Proof Let $s > s_0$ and let c, ε_0 as in Eq. (2.1). For $R \in \mathbb{N}$ and $\varepsilon \in (0, \varepsilon_0)$, define $A(R, \varepsilon) := \{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq \varepsilon\}$. By definition,

$$A(R, \varepsilon) \subset \bigcup_{\mathbf{y} \in \mathrm{range}(D_R)} \big[\mathcal{S} \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{X})\big].$$

Since $\mathbb{P}$ is of growth order s_0 and because of $|\mathrm{range}(D_R)| \leq 2^R$, we can apply (2.1) and the definition of the outer measure $\mathbb{P}^*$ (see Eq. (1.5)) to deduce

$$\mathbb{P}^*\big(A(R, \varepsilon)\big) \leq \sum_{\mathbf{y} \in \mathrm{range}(D_R)} \mathbb{P}\big(\mathcal{S} \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{X})\big) \leq 2^R \cdot 2^{-c \cdot \varepsilon^{-1/s}}.$$

This proves the first part of the lemma.

To prove the second part, let $s > s_0$, and choose $\sigma = \frac{s+s_0}{2}$, noting that $\sigma \in (s_0, s)$. Therefore, the first part of the lemma, applied with σ instead of s, yields $c, \varepsilon_0 > 0$ such that $\mathbb{P}^*(\{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq \varepsilon\}) \leq 2^{R - c \cdot \varepsilon^{-1/\sigma}}$ for all $R \in \mathbb{N}$ and $\varepsilon \in (0, \varepsilon_0)$.

Note that $\varepsilon := K \cdot R^{-s} \leq \frac{\varepsilon_0}{2} < \varepsilon_0$ holds as soon as $R \geq \lceil (2K/\varepsilon_0)^{1/s} \rceil =: R_1$. Finally, since $s/\sigma > 1$ we can find a code-length $R_2 = R_2(s, s_0, \varepsilon_0, K) \in \mathbb{N}$ such that

$$R - c \cdot \varepsilon^{-1/\sigma} = R - c \cdot K^{-1/\sigma} \cdot R^{s/\sigma} \leq -R \quad \text{for } R \geq R_2 .$$

Overall, we thus see that (2.4) holds, with $R_0 := \max\{R_1, R_2\}$. □

Proposition 1 *Let $\mathcal{S} \neq \emptyset$ be a subset of the Banach space $\mathbf{X}$. If $\mathbb{P}$ is a Borel probability measure on $\mathcal{S}$ that is of growth order $s_0 \in [0, \infty)$, then, for every $s > s_0$ and every codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \text{Codecs}_{\mathcal{S},\mathbf{X}}$, we have*

$$\mathbb{P}^*\big(\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C})\big) = 0 .$$

Proof First, note that

$$\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C}) = \bigcup_{K \in \mathbb{N}} \big\{ \mathbf{x} \in \mathcal{S} \colon \forall R \in \mathbb{N} \colon \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_\mathbf{X} \leq K \cdot R^{-s} \big\}$$

$$= \bigcup_{K \in \mathbb{N}} \bigcap_{R \in \mathbb{N}} A^{(s)}_{K,R},$$

where $A^{(s)}_{K,R} := \{ \mathbf{x} \in \mathcal{S} \colon \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_\mathbf{X} \leq K \cdot R^{-s} \}$.

By σ-subadditivity of $\mathbb{P}^*$, it is thus enough to show $\mathbb{P}^*\big(\bigcap_{R \in \mathbb{N}} A^{(s)}_{K,R}\big) = 0$ for each $K \in \mathbb{N}$. To see that this holds, note that Lemma 1 shows

$$0 \leq \mathbb{P}^*\left(\bigcap_{R \in \mathbb{N}} A^{(s)}_{K,R} \right) \leq \mathbb{P}^*\big(A^{(s)}_{K,R}\big) \leq 2^{-R} \quad \forall R \geq R_0(s, s_0, K, \mathbb{P}, \mathcal{S}, \mathbf{X}).$$

This easily implies $\mathbb{P}^*\big(\bigcap_{R \in \mathbb{N}} A^{(s)}_{K,R}\big) = 0$. □

The proof of Theorem 4 merely consists of combining the preceding lemmas.

Proof of Theorem 4 **Proof of (i)**: This is contained in the statement of Lemma 1.
Proof of (ii): This follows from Proposition 1.
Proof of (iii): This follows from the definition of the optimal compression rate: for $0 \leq s < s^*$ there exists a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \text{Codecs}_{\mathcal{S},\mathbf{X}}$ such that

$$R^s \cdot \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_\mathbf{X} \leq C \quad \forall R \in \mathbb{N},$$

for a constant $C > 0$ and all $\mathbf{x} \in \mathcal{S}$. In particular, this implies $\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C}) = \mathcal{S}$, and therefore $\mathbb{P}\big(\mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C})\big) = 1$. □

We close this subsection by showing that if $\mathbb{P}$ is a probability measure with logarithmic growth order s_0, then this growth order is at least as large as the optimal compression rate of the set on which $\mathbb{P}$ is defined. This justifies the nomenclature of "critical measures" as introduced in Definition 3.

Corollary 2 *Let $\mathcal{S} \neq \emptyset$ be a subset of $\mathbf{X}$, and $\mathbb{P}$ be a Borel probability measure on $\mathcal{S}$ of growth order $s_0 \geq 0$. Then, $s_0 \geq s_{\mathbf{X}}^*(\mathcal{S})$, with $s_{\mathbf{X}}^*(\mathcal{S})$ as defined in Eq. (1.1).*

Proof Suppose for a contradiction that $0 \leq s_0 < s_{\mathbf{X}}^*(\mathcal{S})$, and let $s \in (s_0, s_{\mathbf{X}}^*(\mathcal{S}))$. By definition of $s_{\mathbf{X}}^*(\mathcal{S})$, there exists a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \text{Codecs}_{\mathcal{S}, \mathbf{X}}$ such that $\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C}) = \mathcal{S}$. By Proposition 1, we thus obtain the desired contradiction $1 = \mathbb{P}(\mathcal{S}) = \mathbb{P}(\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})) = 0$. □

2.2 Transferring Critical Measures

Our main goal in this paper is to prove a phase transition as in (1.4) for $\mathcal{S}$ being the unit ball of suitable Besov- or Sobolev spaces. To do so, we will first prove (in Sect. 3) that such a phase-transition occurs for a certain class of sequence spaces and then, transfer this result to the Besov- and Sobolev spaces, essentially by discretizing these function spaces using suitable wavelet systems. In the present subsection, we formulate a general result that allows such a transfer from a phase transition as in (1.4) from one space to another.

The precise (very general, but slightly technical) transference result reads as follows:

Theorem 5 *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be Banach spaces, and let $\mathcal{S}_{\mathbf{X}} \subset \mathbf{X}$, $\mathcal{S}_{\mathbf{Y}} \subset \mathbf{Y}$, and $\mathcal{S} \subset \mathbf{Z}$. Assume that*

1. *$s_{\mathbf{X}}^*(\mathcal{S}_{\mathbf{X}}) = s_{\mathbf{Y}}^*(\mathcal{S}_{\mathbf{Y}})$;*
2. *there exists a Lipschitz continuous map $\Phi : \mathcal{S}_{\mathbf{X}} \subset \mathbf{X} \to \mathbf{Z}$ satisfying $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$;*
3. *there exists a Borel probability measure $\mathbb{P}$ on $\mathcal{S}_{\mathbf{Y}}$ that is critical for $\mathcal{S}_{\mathbf{Y}}$ with respect to $\mathbf{Y}$;*
4. *there exists a (not necessarily surjective) measurable map $\Psi : \mathcal{S}_{\mathbf{Y}} \to \mathcal{S}$ that is expansive, meaning that there exists $\kappa > 0$ satisfying*
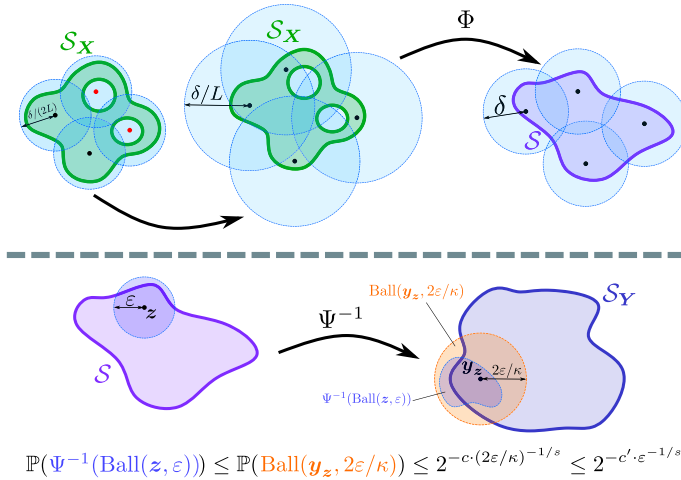
$$\|\Psi(\mathbf{x}) - \Psi(\mathbf{x}')\|_{\mathbf{Z}} \geq \kappa \cdot \|\mathbf{x} - \mathbf{x}'\|_{\mathbf{Y}} \quad \forall \mathbf{x}, \mathbf{x}' \in \mathcal{S}_{\mathbf{Y}}.$$

Then, $s_{\mathbf{Z}}^*(\mathcal{S}) = s_{\mathbf{X}}^*(\mathcal{S}_{\mathbf{X}})$, and the push-forward measure $\mathbb{P} \circ \Psi^{-1}$ is a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $\mathbf{Z}$.

Remark (1) As mentioned in Sect. 1.5, regarding the measurability of Ψ, $\mathcal{S}_{\mathbf{Y}}$ is equipped with the trace σ-algebra of the Borel σ-algebra on $\mathbf{Y}$, and analogously for $\mathcal{S}$.

(2) In most practical applications of this theorem, one is given a Lipschitz continuous map $\Phi : \mathbf{X} \to \mathbf{Z}$ and a (not necessarily surjective) measurable, expansive map $\Psi : \mathbf{Y} \to \mathbf{Z}$ satisfying $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$ and $\Psi(\mathcal{S}_{\mathbf{Y}}) \subset \mathcal{S}$. For greater generality, in the above theorem we only assume that Φ, Ψ are defined on $\mathcal{S}_{\mathbf{X}}$ and $\mathcal{S}_{\mathbf{Y}}$, respectively. ◄

Proof The proof is given in Appendix A. □

$$\mathbb{P}\big(\Psi^{-1}(\mathrm{Ball}(z,\varepsilon))\big) \leq \mathbb{P}\big(\mathrm{Ball}(y_z, 2\varepsilon/\kappa)\big) \leq 2^{-c\cdot(2\varepsilon/\kappa)^{-1/s}} \leq 2^{-c'\cdot\varepsilon^{-1/s}}$$

Fig. 2 Geometric intuition behind Theorem 5. Top: An encoder/decoder pair (E_R, D_R) with distortion at most $\delta/(2L)$ corresponds to a covering of $\mathcal{S}_{\mathbf{X}}$ by 2^R balls of radius $\delta/(2L)$, not necessarily centered inside $\mathcal{S}_{\mathbf{X}}$ (red centers; also see Lemma 10). By doubling the radius, one can "move the centers inside $\mathcal{S}_{\mathbf{X}}$." Using the Lipschitz map Φ satisfying $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$, this yields a covering of $\mathcal{S}$ by 2^R balls of radius δ, and hence, an encoder/decoder pair with distortion at most δ. This entails $s_{\mathbf{Z}}^*(\mathcal{S}) \geq s_{\mathbf{X}}^*(\mathcal{S}_{\mathbf{X}})$. Bottom: Since Ψ is expansive, the inverse image under Ψ of a ball of radius ε is contained in a ball of radius $2\varepsilon/\kappa$, ensuring that the push-forward measure $\mathbb{P}\circ\Psi^{-1}$ is of logarithmic growth order $s_{\mathbf{Y}}^*(\mathcal{S}_{\mathbf{Y}})$ on $\mathcal{S}$, which implies $s_{\mathbf{Z}}^*(\mathcal{S}) \leq s_{\mathbf{Y}}^*(\mathcal{S}_{\mathbf{Y}})$; see Corollary 2

3 Proof of the Phase Transition in $\ell^2(\mathcal{I})$

In this section, we provide the proof of the phase transition for a class of sequence spaces associated with Sobolev- and Besov spaces; these sequence spaces are defined in Sect. 3.1, where we also formulate the main result (Theorem 6) concerning the compressibility phase transition for these spaces. Section 3.2 establishes elementary embedding results for these spaces and provides a lower bound for their optimal compression rate; the latter essentially follows by adapting results by Leopold [20] to our setting. The construction of the critical probability measure for the sequence spaces is presented in Sect. 3.3, while the proof of Theorem 6 is given in Sect. 3.4.

3.1 Main Result

Definition 4 (*d*-**regular partitions**) Let $\mathcal{I}$ be a countably infinite index set, and $\mathscr{P} = (\mathcal{I}_m)_{m\in\mathbb{N}}$ be a partition of $\mathcal{I}$; that is, $\mathcal{I} = \biguplus_{m=1}^{\infty} \mathcal{I}_m$, where the union is disjoint. For $d \in \mathbb{N}$, we call $\mathscr{P}$ a *d-regular partition*, if there are $0 < a < A < \infty$ satisfying

$$a \cdot 2^{dm} \leq |\mathcal{I}_m| \leq A \cdot 2^{dm} \quad \text{for all } m \in \mathbb{N}. \tag{3.1}$$

Convention: We will always assume that $\mathcal{I}$, $\mathscr{P}$ and d have this meaning.

Associated with a *d*-regular partition, we now define the following family of weighted sequence spaces.

Definition 5 (**Sequence spaces**) Let $p, q \in (0, \infty]$ and $\alpha, \theta \in \mathbb{R}$. For any sequence $\mathbf{x} = (x_i)_{i \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$, we define

$$\mathbf{x}_m := \mathbf{x}|_{\mathcal{I}_m} = (x_i)_{i \in \mathcal{I}_m}$$

$$\text{and} \quad \|\mathbf{x}\|_{\ell^{p,q}_{\mathcal{P},\alpha,\theta}} := \left\| \left(2^{\alpha m} \cdot m^\theta \cdot \|\mathbf{x}_m\|_{\ell^p(\mathcal{I}_m)} \right)_{m \in \mathbb{N}} \right\|_{\ell^q(\mathbb{N})}. \tag{3.2}$$

The *mixed-norm sequence space* $\ell^{p,q}_{\mathcal{P},\alpha,\theta}$ is

$$\ell^{p,q}_{\mathcal{P},\alpha,\theta} := \left\{ \mathbf{x} \in \mathbb{R}^{\mathcal{I}} \ : \ \|\mathbf{x}\|_{\ell^{p,q}_{\mathcal{P},\alpha,\theta}} < \infty \right\}.$$

For brevity, we also define $\ell^{p,q}_{\mathcal{P},\alpha} := \ell^{p,q}_{\mathcal{P},\alpha,0}$ and

$$\mathcal{S}^{p,q}_{\mathcal{P},\alpha,\theta} := \text{Ball}\left(0, 1; \ell^{p,q}_{\mathcal{P},\alpha,\theta}\right), \quad \text{as well as} \quad \mathcal{S}^{p,q}_{\mathcal{P},\alpha} := \mathcal{S}^{p,q}_{\mathcal{P},\alpha,0}.$$

In the remainder of this section, we will prove the existence of a critical measure on each of the sets $\mathcal{S}^{p,q}_{\mathcal{P},\alpha}$, provided that $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$. In the proof, the (otherwise not really important) spaces $\ell^{p,q}_{\mathcal{P},\alpha,\theta}$ will play an essential role. The main result of this section is thus the following theorem, the proof of which is given in Sect. 3.4 below.

Theorem 6 *Let $p, q \in (0, \infty]$ and $\alpha \in \mathbb{R}$, and assume that $\alpha > d \cdot \left(\frac{1}{2} - \frac{1}{p}\right)_+$. Then, $\mathcal{S}^{p,q}_{\mathcal{P},\alpha} \subset \ell^2(\mathcal{I})$ is compact and hence, Borel measurable, its optimal compression rate is given by $s^*_{\ell^2(\mathcal{I})}\left(\mathcal{S}^{p,q}_{\mathcal{P},\alpha}\right) = \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$, and there exists a Borel probability measure $\mathbb{P}^{p,q}_{\mathcal{P},\alpha}$ on $\mathcal{S}^{p,q}_{\mathcal{P},\alpha}$ that is critical for $\mathcal{S}^{p,q}_{\mathcal{P},\alpha}$ with respect to $\ell^2(\mathcal{I})$. In particular, the phase transition described in Theorem 4 holds.*

Remark Explicitly, the proof shows for any $s > s^* = \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$ that

$$\mathbb{P}^{p,q}_{\mathcal{P},\alpha}\left(\mathcal{S} \cap \text{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\right) \le 2^{-c \cdot \varepsilon^{-1/s}} \quad \forall \varepsilon \in (0, \varepsilon_0),$$

where $\varepsilon_0 = \varepsilon_0^{(0)} \cdot (s - s^*)^{2/q} \cdot e^{-s \cdot (d+1)}$ and $c = c^{(0)} \cdot 2^{-d} \cdot (s - s^*)^{2/(sq)}$ for constants $\varepsilon_0^{(0)} = \varepsilon_0^{(0)}(p, q, a, A) > 0$ and $c^{(0)} = c^{(0)}(s^*, p, q, a, A) > 0$ with a, A as in (3.1). This provides control on how fast c, ε_0 deteriorate as $s \downarrow s^*$ or $d \to \infty$. These bounds are probably not optimal. ◀

3.2 Embedding Results and a Lower Bound for the Compression Rate

Having introduced the signal classes $\mathcal{S}^{p,q}_{\mathcal{P},\alpha}$, we now collect two technical ingredients needed to construct the measures $\mathbb{P}^{p,q}_{\mathcal{P},\alpha}$ on these sets: A lower bound for the optimal compression rate of $\mathcal{S}^{p,q}_{\mathcal{P},\alpha}$ (Proposition 2) and certain elementary embeddings between the spaces $\ell^{p,q}_{\mathcal{P},\alpha,\theta}$ for different choices of the parameters (Lemma 2).

Lemma 2 *Let $p, q, r \in (0, \infty]$ and $\alpha, \theta, \vartheta \in \mathbb{R}$. If $q > r$ and $\vartheta > \frac{1}{r} - \frac{1}{q}$, then $\ell^{p,q}_{\mathscr{P},\alpha,\theta+\vartheta} \hookrightarrow \ell^{p,r}_{\mathscr{P},\alpha,\theta}$. More precisely, there exists a constant $\kappa = \kappa(r, q, \vartheta) \geq 1$ such that $\|\mathbf{x}\|_{\ell^{p,r}_{\mathscr{P},\alpha,\theta}} \leq \kappa \cdot \|\mathbf{x}\|_{\ell^{p,q}_{\mathscr{P},\alpha,\theta+\vartheta}}$ for all $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$.*

Proof The claim follows by an elementary application of Hölder's inequality; the details can be found in Appendix H. □

The following result shows that the supremal compression rate of the class $\mathcal{S}^{p,q}_{\mathscr{P},\alpha}$ identified in Theorem 6 is realized by a suitable codec—thus, the supremum in Eq. (1.1) is attained in this setting.

Proposition 2 *Let $p, q \in (0, \infty]$ and $\alpha \in (0, \infty)$, and assume $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$. Then, we have $\ell^{p,q}_{\mathscr{P},\alpha} \hookrightarrow \ell^2(\mathcal{I})$, and the set $\mathcal{S}^{p,q}_{\mathscr{P},\alpha} \subset \ell^2(\mathcal{I})$ is compact with $s^*_{\ell^2(\mathcal{I})}(\mathcal{S}^{p,q}_{\mathscr{P},\alpha}) \geq \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$.*

Furthermore, there exists a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}^{p,q}_{\mathscr{P},\alpha}, \ell^2(\mathcal{I})}$ satisfying

$$\delta_{\mathcal{S}^{p,q}_{\mathscr{P},\alpha}, \ell^2(\mathcal{I})}(E_R, D_R) \lesssim R^{-\left(\frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})\right)} \quad \forall\, R \in \mathbb{N}. \tag{3.3}$$

Proof In essence, this an entropy estimate for sequence spaces; see [12]. Since the precise proof mainly consists in translating the results in [20] to our setting, it is deferred to Appendix B. □

3.3 Construction of the Measure

We now come to the technical heart of this section—the construction of the measures $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}$. We will provide different constructions for $q = \infty$ and for $q < \infty$: Since for $q = \infty$ the class $\mathcal{S}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ has a natural product structure (Lemma 3), we define the measure as a product measure (Definition 6). We then use the embedding result of Lemma 2 to transfer the measure on $\mathcal{S}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ to the general signal classes $\mathcal{S}^{p,q}_{\mathscr{P},\alpha}$; see Definition 7.

We start with the elementary observation that the balls $\mathcal{S}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ can be written as infinite products of finite-dimensional balls.

Lemma 3 *The balls of the mixed-norm sequence spaces $\ell^{p,\infty}_{\mathscr{P},\alpha,\theta}$ satisfy (up to canonical identifications) the factorization*

$$\mathcal{S}^{p,\infty}_{\mathscr{P},\alpha,\theta} = \mathrm{Ball}\big(0, 1; \ell^{p,\infty}_{\mathscr{P},\alpha,\theta}\big) = \bigtimes_{m \in \mathbb{N}} \mathrm{Ball}\big(0, 2^{-\alpha m} m^{-\theta}; \ell^p(\mathcal{I}_m)\big).$$

Proof We identify $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$ with $(\mathbf{x}_m)_{m \in \mathbb{N}} \in \bigtimes_{m \in \mathbb{N}} \mathbb{R}^{\mathcal{I}_m}$, as defined in Eq. (3.2). Set $w_m := m^{\theta} \cdot 2^{\alpha m}$ for $m \in \mathbb{N}$. The statement of the lemma then follows by recalling that

$$\|\mathbf{x}\|_{\ell^{p,\infty}_{\mathscr{P},\alpha,\theta}} = \sup_{m \in \mathbb{N}} \Big(w_m \cdot \|\mathbf{x}_m\|_{\ell^p(\mathcal{I}_m)}\Big).$$

□

With Lemma 3 in hand, we can readily define $\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}$ as a product measure.

Definition 6 (**Measures for** $q = \infty$) Let $\mathscr{P} = (\mathcal{I}_m)_{m \in \mathbb{N}}$ be a d-regular partition of $\mathcal{I}$. Let $\mathscr{B}_m$ be the Borel σ-algebra on $\mathbb{R}^{\mathcal{I}_m}$ and denote the Lebesgue measure on $(\mathbb{R}^{\mathcal{I}_m}, \mathscr{B}_m)$ by μ_m.

For $p \in (0, \infty]$ and $w_m > 0$ define the probability measure $\mathbb{P}_m^{p,w_m}$ on $(\mathbb{R}^{\mathcal{I}_m}, \mathscr{B}_m)$ by

$$\mathbb{P}_m^{p,w_m}: \quad \mathscr{B}_m \to [0,1], \quad A \mapsto \frac{\mu_m\big(A \cap \text{Ball}\,\big(0, w_m^{-1}; \ell^p(\mathcal{I}_m)\big)\big)}{\mu_m\big(\text{Ball}\,\big(0, w_m^{-1}; \ell^p(\mathcal{I}_m)\big)\big)}. \qquad (3.4)$$

Given $p \in (0, \infty]$ and $\alpha, \theta \in \mathbb{R}$ define $w_m := m^\theta \cdot 2^{\alpha m}$, let $\mathscr{B}_{\mathcal{I}}$ denote the product σ-algebra on $\mathbb{R}^{\mathcal{I}}$ and define $\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}$ as the product measure of the family $\big(\mathbb{P}_m^{p,w_m}\big)_{m \in \mathbb{N}}$ (see, e.g., [11, Section 8.2]):

$$\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty} := \bigotimes_{m \in \mathbb{N}} \mathbb{P}_m^{p,w_m}: \quad \mathscr{B}_{\mathcal{I}} \to [0,1]. \qquad (3.5)$$

With the help of the preceding results, we can now describe the construction of the measure $\mathbb{P}_{\mathscr{P},\alpha}^{p,q}$ on $\mathcal{S}_{\mathscr{P},\alpha}^{p,q}$, also for $q < \infty$. A crucial tool will be the embedding result from Lemma 2.

Definition 7 (**Measures for** $q < \infty$) Let the notation be as in Definition 6.

For given $q \in (0, \infty]$, choose (according to Lemma 2) a constant $\kappa = \kappa(q) \geq 1$ (with $\kappa = 1$ if $q = \infty$) such that $\|\mathbf{x}\|_{\ell_{\mathscr{P},\alpha,0}^{p,q}} \leq \kappa \cdot \|\mathbf{x}\|_{\ell_{\mathscr{P},\alpha,2/q}^{p,\infty}}$ for all $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$, and define

$$\mathbb{P}_{\mathscr{P},\alpha}^{p,q}: \quad \mathscr{B}_{\mathcal{I}} \to [0,1], \quad A \mapsto \mathbb{P}_{\mathscr{P},\alpha,2/q}^{p,\infty}(\kappa \cdot A).$$

In the following, we verify that the measures defined according to Definitions 6 and 7 are indeed (Borel) probability measures on the signal classes $\mathcal{S}_{\mathscr{P},\alpha,\theta}^{p,\infty}$ and $\mathcal{S}_{\mathscr{P},\alpha}^{p,q}$, respectively. To do so, we first show that the signal classes are measurable with respect to the product σ-algebra $\mathscr{B}_{\mathcal{I}}$, and we compare this σ-algebra to the Borel σ-algebra on $\ell^2(\mathcal{I})$.

Lemma 4 *Let $\mathscr{B}_{\mathcal{I}}$ denote the product σ-algebra on $\mathbb{R}^{\mathcal{I}}$ and let $p, q \in (0, \infty]$ and $\alpha, \theta \in \mathbb{R}$. Then, the (quasi)-norm $\|\cdot\|_{\ell_{\mathscr{P},\alpha,\theta}^{p,q}}: \mathbb{R}^{\mathcal{I}} \to [0, \infty]$ is measurable with respect to $\mathscr{B}_{\mathcal{I}}$. In particular, $\mathcal{S}_{\mathscr{P},\alpha,\theta}^{p,q} \in \mathscr{B}_{\mathcal{I}}$.*

Further, the Borel σ-algebra $\mathscr{B}_{\ell^2}$ on $\ell^2(\mathcal{I})$ coincides with the trace σ-algebra $\ell^2(\mathcal{I}) \cap \mathscr{B}_{\mathcal{I}}$.

Proof The (mainly technical) proof is deferred to Appendix H.   □

Lemma 5 (a) *The measure $\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}$ is a probability measure on the measurable space $\big(\mathcal{S}_{\mathscr{P},\alpha,\theta}^{p,\infty}, \mathcal{S}_{\mathscr{P},\alpha,\theta}^{p,\infty} \cap \mathscr{B}_{\mathcal{I}}\big)$.*

(b) *If $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$, then $\mathcal{S}_{\mathcal{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$, and the measure $\mathbb{P}_{\mathcal{P},\alpha}^{p,q}$ is a probability measure on $(\mathcal{S}_{\mathcal{P},\alpha}^{p,q}, \mathcal{S}_{\mathcal{P},\alpha}^{p,q} \cap \mathcal{B}_{\ell^2})$, where $\mathcal{B}_{\ell^2}$ denotes the Borel σ-algebra on $\ell^2(\mathcal{I})$.*

Proof For the first part, Lemma 4 implies that $\mathcal{S}_{\mathcal{P},\alpha,\theta}^{p,\infty} \in \mathcal{B}_{\mathcal{I}}$, so that $\mathbb{P}_{\mathcal{P},\alpha,\theta}^{p,\infty}$ is a measure on $\mathcal{S}_{\mathcal{P},\alpha,\theta}^{p,\infty} \cap \mathcal{B}_{\mathcal{I}}$. Furthermore, Lemma 3 and Definition 6 show $\mathbb{P}_{\mathcal{P},\alpha,\theta}^{p,\infty}(\mathcal{S}_{\mathcal{P},\alpha,\theta}^{p,\infty}) = 1$.

For the second part, recall from Proposition 2 that $\mathcal{S}_{\mathcal{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$, so that Lemma 4 implies $\mathcal{S}_{\mathcal{P},\alpha}^{p,q} \cap \mathcal{B}_{\ell^2} = \mathcal{S}_{\mathcal{P},\alpha}^{p,q} \cap \mathcal{B}_{\mathcal{I}}$, which easily implies that $\mathbb{P}_{\mathcal{P},\alpha}^{p,q}$ is a measure on $\mathcal{S}_{\mathcal{P},\alpha}^{p,q} \cap \mathcal{B}_{\ell^2}$. Finally, observe that $\mathcal{S}_{\mathcal{P},\alpha,2/q}^{p,\infty} \subset \kappa \cdot \mathcal{S}_{\mathcal{P},\alpha}^{p,q}$ by choice of κ in Definition 7, and hence

$$1 \geq \mathbb{P}_{\mathcal{P},\alpha}^{p,q}(\mathcal{S}_{\mathcal{P},\alpha}^{p,q}) \geq \mathbb{P}_{\mathcal{P},\alpha}^{p,q}(\kappa^{-1} \mathcal{S}_{\mathcal{P},\alpha,2/q}^{p,\infty}) = \mathbb{P}_{\mathcal{P},\alpha,2/q}^{p,\infty}(\mathcal{S}_{\mathcal{P},\alpha,2/q}^{p,\infty}) = 1.$$

$\square$

3.4 Proof of Theorem 6

In this subsection, we prove that the measures $\mathbb{P}_{\mathcal{P},\alpha}^{p,q}$ constructed in Definition 7 are critical, provided that $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$. An essential ingredient for the proof is the following estimate for the volumes of balls in $\ell^p([m])$.

Lemma 6 *Let $m \in \mathbb{N}$ and $p \in (0, \infty]$. The m-dimensional Lebesgue measure of $\mathrm{Ball}(0, 1; \ell^p([m]))$ is*

$$\lambda_m\big(\mathrm{Ball}(0, 1; \ell^p([m]))\big) = \frac{2^m \cdot \big(\Gamma(1 + \frac{1}{p})\big)^m}{\Gamma(1 + \frac{m}{p})}. \tag{3.6}$$

For every $p \in (0, \infty]$, there exist constants $c_p \in (0, 1]$ and $C_p \in [1, \infty)$, such that

$$c_p^m \cdot m^{-m(\frac{1}{2} - \frac{1}{p})} \leq \frac{\lambda_m\big(\mathrm{Ball}(0, 1; \ell^2([m]))\big)}{\lambda_m\big(\mathrm{Ball}(0, 1; \ell^p([m]))\big)} \leq C_p^m \cdot m^{-m(\frac{1}{2} - \frac{1}{p})} \quad \forall m \in \mathbb{N}. \tag{3.7}$$

Proof A proof of (3.6) can be found, e.g., in [18, Theorem 5].

For proving (3.7), we use that in [18, Lemma 4] it is shown that for each $p \in (0, \infty)$ there are constants $\eta_p, \omega_p > 0$ satisfying

$$\eta_p \cdot x^{1/p} \leq \Big[\Gamma\Big(1 + \frac{x}{p}\Big)\Big]^{1/x} \leq \omega_p \cdot x^{1/p} \quad \forall x \in [1, \infty). \tag{3.8}$$

It is clear that this remains true also for $p = \infty$; in fact, since $\Gamma(1) = 1$, one can simply choose $\eta_\infty = \omega_\infty = 1$ in this case.

By (3.6), we see that

$$\frac{\lambda_m\big(\mathrm{Ball}(0,1;\ell^2([m]))\big)}{\lambda_m\big(\mathrm{Ball}(0,1;\ell^p([m]))\big)} = \left(\frac{\Gamma(1+\frac{1}{2})}{\Gamma(1+\frac{1}{p})}\right)^m \cdot \frac{\Gamma(1+\frac{m}{p})}{\Gamma(1+\frac{m}{2})},$$

and the estimate (3.8) implies

$$\frac{\eta_p^m \cdot m^{m/p}}{\omega_2^m \cdot m^{m/2}} \leq \frac{\Gamma(1+\frac{m}{p})}{\Gamma(1+\frac{m}{2})} \leq \frac{\omega_p^m \cdot m^{m/p}}{\eta_2^m \cdot m^{m/2}}.$$

Hence, we can choose $C_p := \max\left\{1, \frac{\Gamma(1+\frac{1}{2})}{\Gamma(1+\frac{1}{p})}\frac{\omega_p}{\eta_2}\right\}$ and $c_p := \min\left\{1, \frac{\Gamma(1+\frac{1}{2})}{\Gamma(1+\frac{1}{p})}\frac{\eta_p}{\omega_2}\right\}$.

□

We are finally equipped to prove Theorem 6.

Proof of Theorem 6 **Step 1:** We show for $s^* := \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$ and arbitrary $\theta \in [0, \infty)$ that the measure $\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ has growth order s^* with respect to $\ell^2(\mathcal{I})$.

To this end, let $s > s^*$ be arbitrary, and let $\varepsilon \in (0, \varepsilon_0)$ (for a suitable $\varepsilon_0 > 0$ to be chosen below), and $\mathbf{x} \in \ell^2(\mathcal{I})$. We estimate the measure $\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,\theta}(\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I})))$ by estimating the measure of certain finite-dimensional projections of the ball, exploiting the product structure of the measure: Recall the identification $\mathbf{x} = (\mathbf{x}_m)_{m\in\mathbb{N}}$, where $\mathbf{x}_m = \mathbf{x}|_{\mathcal{I}_m}$. Set $w_m := m^\theta \cdot 2^{\alpha m}$ for $m \in \mathbb{N}$, as in Definition 6. For arbitrary $m \in \mathbb{N}$, we have

$$\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I})) \subset \underset{t=1}{\overset{m-1}{\times}} \mathbb{R}^{\mathcal{I}_t} \times \mathrm{Ball}(\mathbf{x}_m, \varepsilon; \ell^2(\mathcal{I}_m)) \times \underset{t=m+1}{\overset{\infty}{\times}} \mathbb{R}^{\mathcal{I}_t}.$$

Using the product structure of $\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ (cf. Eq. (3.5)) and the constant $C_p \geq 1$ from Lemma 6, we thus see for each $m \in \mathbb{N}$ that

$$\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,\theta}\big(\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big)$$
$$\leq \mathbb{P}^{p,w_m}_m\left(\mathrm{Ball}\left(\mathbf{x}_m, \varepsilon; \ell^2(\mathcal{I}_m)\right)\right)$$
$$\leq \frac{\mu_m\left(\mathrm{Ball}\left(\mathbf{x}_m, \varepsilon; \ell^2(\mathcal{I}_m)\right)\right)}{\mu_m\left(\mathrm{Ball}\left(0, w_m^{-1}; \ell^p(\mathcal{I}_m)\right)\right)} \qquad \text{by Equation (3.4),}$$
$$= \varepsilon^{n_m} w_m^{n_m} \cdot \frac{\lambda_{n_m}\left(\mathrm{Ball}(0,1;\ell^2(\mathcal{I}_m))\right)}{\lambda_{n_m}\left(\mathrm{Ball}(0,1;\ell^p(\mathcal{I}_m))\right)} \qquad \text{for } n_m := |\mathcal{I}_m|,$$
$$\leq \left(C_p \cdot \varepsilon\, w_m \cdot n_m^{-(\frac{1}{2}-\frac{1}{p})}\right)^{n_m} \qquad \text{by Lemma 6.}$$

From (3.1), we see that $n_m = 2^{dm}\eta_m$ for a certain $\eta_m \in [a, A]$ and hence $\eta_m^{-(\frac{1}{2}-\frac{1}{p})} \leq K_1^{(0)} = K_1^{(0)}(a, A, p)$. Furthermore, a straightforward curve discussion of the function

$x \mapsto x^\theta \cdot e^{-\delta x}$ for $\delta = (s - s^*) \cdot \ln 2$ shows that $x^\theta \leq \left(\frac{\theta}{e \cdot \ln 2 \cdot (s-s^*)}\right)^\theta \cdot 2^{x(s-s^*)}$ for all $x > 0$, with the convention $0^0 = 1$. Combining these observations, we see for $K_1^{(1)} := (\theta/(e \ln 2))^\theta$ that

$$
\begin{aligned}
w_m \, n_m^{-(\frac{1}{2} - \frac{1}{p})} &= m^\theta \, 2^{\alpha m} \, 2^{-md(\frac{1}{2} - \frac{1}{p})} \, \eta_m^{-(\frac{1}{2} - \frac{1}{p})} \\
&= m^\theta \, 2^{mds^*} \, \eta_m^{-(\frac{1}{2} - \frac{1}{p})} \\
&\leq K_1^{(0)} K_1^{(1)} \cdot (s - s^*)^{-\theta} \cdot 2^{m(s-s^*)} 2^{dms^*} \\
&\leq K_1 \cdot (s - s^*)^{-\theta} \cdot 2^{md(s-s^*)} 2^{dms^*} \\
&= K_1 \cdot (s - s^*)^{-\theta} \cdot 2^{mds},
\end{aligned}
$$

where $K_1 = K_1(\theta, p, a, A) \geq 1$.

For $K_2^{(0)} := C_p K_1 \geq 1$ and $K_2 := K_2^{(0)}/(s - s^*)^\theta$, we thus see that $K_2^{(0)} = K_2^{(0)}(\theta, p, a, A)$ and

$$
\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}\big(\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big) \leq \left(K_2 \cdot \varepsilon \cdot 2^{mds}\right)^{2^{md} \eta_m} \qquad \forall\, m \in \mathbb{N}. \tag{3.9}
$$

A good candidate for an upper bound for $\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}(\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I})))$ is associated with a positive integer close to

$$
\widetilde{m}(\varepsilon) := \operatorname*{argmin}_{m \in \mathbb{R}} \left(K_2 \cdot \varepsilon \cdot 2^{msd}\right)^{2^{md}} = -\frac{\log_2(K_2 \cdot \varepsilon)}{ds} - \frac{\log_2 e}{d}.
$$

Choose a positive $\varepsilon_0 > 0$ so small that $\widetilde{m}(\varepsilon) > 1$ for all $\varepsilon \in (0, \varepsilon_0)$. An easy calculation shows that one can choose $\varepsilon_0 = 1/(K_2 \cdot e^{s(d+1)}) = \varepsilon_0^{(0)} \cdot (s - s^*)^\theta \cdot e^{-s(d+1)}$, where $\varepsilon_0^{(0)} = \varepsilon_0^{(0)}(\theta, p, a, A) > 0$.

Set $m_0 := \lfloor \widetilde{m}(\varepsilon) \rfloor \in \mathbb{N}$. Note that $2^{ds \cdot \widetilde{m}(\varepsilon)} = e^{-s}/(K_2 \cdot \varepsilon)$, and hence $K_2 \, \varepsilon \, 2^{ds \cdot m_0} \leq e^{-s} < 1$. For the exponent in (3.9), observe that

$$
2^{d\,m_0} \, \eta_{m_0} \geq a \cdot 2^{d \cdot (\widetilde{m}(\varepsilon) - 1)} = \frac{a}{2^d} \cdot \left(2^{ds\widetilde{m}(\varepsilon)}\right)^{1/s} = \frac{a}{2^d \cdot e \cdot K_2^{1/s}} \cdot \varepsilon^{-1/s} \geq K_3 \cdot \varepsilon^{-1/s},
$$

where $K_3 = K_3^{(0)} \cdot 2^{-d} \cdot (s - s^*)^{\theta/s}$ and $K_3^{(0)} = K_3^{(0)}(s^*, \theta, p, a, A) > 0$ is given by $K_3^{(0)} = a/\big(e \cdot (K_2^{(0)})^{1/s^*}\big)$. Now, (3.9) can be estimated further, yielding

$$
\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}\big(\mathrm{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big) \leq \left(K_2 \cdot \varepsilon \cdot 2^{m_0 ds}\right)^{K_3 \cdot \varepsilon^{-1/s}} \leq e^{-K_3 \cdot s \cdot \varepsilon^{-1/s}} \leq 2^{-K_4 \cdot \varepsilon^{-1/s}},
$$

for $K_4 := s^* \cdot K_3$, so that $K_4 = K_4^{(0)} \cdot 2^{-d} \cdot (s - s^*)^{\theta/s}$ for a suitable constant $K_4^{(0)} = K_4^{(0)}(s^*, \theta, p, a, A) > 0$. Since $s > s^*$ was arbitrary, this shows that $\mathbb{P}_{\mathscr{P},\alpha,\theta}^{p,\infty}$ is of logarithmic growth order s^*; see Definition 3.

Step 2: We show that $\mathbb{P}_{\mathscr{P},\alpha}^{p,q}$ is of growth order s^* with respect to $\ell^2(\mathcal{I})$ on $\mathcal{S}_{\mathscr{P},\alpha}^{p,q}$.

To see this, let $s > s^*$ be arbitrary and choose (by virtue of Step 1) $\varepsilon_0, c > 0$ such that $\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,2/q}\big(\text{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big) \leq 2^{-c \cdot \varepsilon^{-1/s}}$ for all $\mathbf{x} \in \ell^2(\mathcal{I})$ and $\varepsilon \in (0, \varepsilon_0)$. From the explicit formulas given in Step 2, we see that $\varepsilon_0 = \varepsilon_0^{(0)} \cdot (s - s^*)^{2/q} \cdot e^{-s(d+1)}$ for $\varepsilon_0^{(0)} = \varepsilon_0^{(0)}(p, q, a, A) > 0$, and furthermore that $c = c^{(0)} \cdot 2^{-d} \cdot (s - s^*)^{2/(sq)}$ for $c^{(0)} = c^{(0)}(s^*, p, q, a, A) > 0$. Recall from Definition 7 that $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}(M) = \mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,2/q}(\kappa M)$ for a suitable $\kappa = \kappa(q) \geq 1$. Define $\varepsilon_0' := \varepsilon_0/\kappa$ and $c' := c \cdot \kappa^{-1/s^*}$.

Now, if $\varepsilon \in (0, \varepsilon_0')$, then $\kappa\varepsilon \in (0, \varepsilon_0)$ and hence

$$
\begin{aligned}
\mathbb{P}^{p,q}_{\mathscr{P},\alpha}\big(\text{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big) &= \mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,2/q}\big(\kappa \, \text{Ball}(\mathbf{x}, \varepsilon; \ell^2(\mathcal{I}))\big) \\
&= \mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,2/q}\big(\text{Ball}(\kappa\mathbf{x}, \kappa\varepsilon; \ell^2(\mathcal{I}))\big) \\
&\leq 2^{-c \cdot (\kappa\varepsilon)^{-1/s}} \leq 2^{-c' \cdot \varepsilon^{-1/s}},
\end{aligned}
$$

where the last step used that $\kappa \geq 1$ and $s > s^*$. Overall, we have shown that $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}$ is of growth order s^* with respect to $\ell^2(\mathcal{I})$.

Step 3: *(Completing the proof):* By Proposition 2, $\mathcal{S}^{p,q}_{\mathscr{P},\alpha} \subset \ell^2(\mathcal{I})$ is compact with $s^*_{\ell^2(\mathcal{I})}\big(\mathcal{S}^{p,q}_{\mathscr{P},\alpha}\big) \geq s^*$. By Step 2 and Lemma 5, $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}$ is a Borel probability measure on $\mathcal{S}^{p,q}_{\mathscr{P},\alpha}$ of growth order s^* with respect to $\ell^2(\mathcal{I})$. Thus, Lemma 9 shows that $s^*_{\ell^2(\mathcal{I})}\big(\mathcal{S}^{p,q}_{\mathscr{P},\alpha}\big) = s^*$ and that $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}$ is critical for $\mathcal{S}^{p,q}_{\mathscr{P},\alpha}$ with respect to $\ell^2(\mathcal{I})$. □

Remark The proof borrows its main idea (using the product measure structure of $\mathbb{P}^{p,\infty}_{\mathscr{P},\alpha,\theta}$ to work on finite dimensional projections) from [24]. ◄

4 Examples

4.1 Besov Spaces on Bounded Open Sets $\Omega \subset \mathbb{R}^d$

For Besov spaces on bounded domains, we obtain the following consequence of Theorem 6, by using suitable wavelet bases to transport the measures $\mathbb{P}^{p,q}_{\mathscr{P},\alpha}$ to the Besov spaces. For a review of the definition of Besov spaces (on $\mathbb{R}^d$ and on domains), and the characterization of these spaces by wavelets, we refer to Appendices C.1 and C.2 .

Theorem 7 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be open and bounded, let $p, q \in (0, \infty]$, and $\tau \in \mathbb{R}$ with $\tau > d \cdot (\frac{1}{p} - \frac{1}{2})_+$.*
Then,

(i) *$\mathcal{S} := \text{Ball}\big(0, 1; B^\tau_{p,q}(\Omega; \mathbb{R})\big)$ is a compact subset of $L^2(\Omega)$, with optimal compression rate given by $s^*_{L^2(\Omega)}(\mathcal{S}) = \frac{\tau}{d}$;*
(ii) *there exists a Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$;*
(iii) *there exists a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \text{Codecs}_{\mathcal{S}, L^2(\Omega)}$ satisfying*

$$\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-\frac{\tau}{d}}.$$

Remark In the discussion following Theorem 4, we observed that the existence of a critical measure in general leaves open what happens for $s = s^*$. In the case of Besov spaces, the above theorem shows that the compression rate $s = s^* = \frac{\tau}{d}$ is actually achieved by a suitable codec. ◀

Proof Define $\alpha := \tau + d \cdot (\frac{1}{2} - \frac{1}{p})$, noting that

$$\alpha > d \cdot \left[(\tfrac{1}{p} - \tfrac{1}{2})_+ + (\tfrac{1}{2} - \tfrac{1}{p}) \right] = d \cdot (\tfrac{1}{2} - \tfrac{1}{p})_+ \,,$$

so that α satisfies the assumptions of Theorem 6.

Using the wavelet characterization of Besov spaces, it is shown in Appendix C.3 [1] that there are countably infinite index sets J^{ext}, J^{int} with associated d-regular partitions $\mathscr{P}^{\text{ext}} = (\mathcal{I}_m^{\text{ext}})_{m \in \mathbb{N}}$ and $\mathscr{P}^{\text{int}} = (\mathcal{I}_m^{\text{int}})_{m \in \mathbb{N}}$ and such that there are linear maps

$$Q_{\text{int}} : \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q} \to B_{p,q}^{\tau}(\Omega; \mathbb{R}) \qquad \text{and} \qquad Q_{\text{ext}} : \ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q} \to B_{p,q}^{\tau}(\Omega; \mathbb{R})$$

with the following properties:

1. $\ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q} \hookrightarrow \ell^2(J^{\text{int}})$ and $\ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q} \hookrightarrow \ell^2(J^{\text{ext}})$; this follows from Proposition 2.
2. There is some $\gamma > 0$ such that $\|Q_{\text{int}} \mathbf{c}\|_{L^2(\Omega)} = \gamma \cdot \|\mathbf{c}\|_{\ell^2} < \infty$ and furthermore $\|Q_{\text{int}} \mathbf{c}\|_{B_{p,q}^{\tau}(\Omega)} \le \|\mathbf{c}\|_{\ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q}}$ for all $\mathbf{c} \in \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q}$.
3. There is $\varrho > 0$ such that $\|Q_{\text{ext}} \mathbf{c}\|_{L^2(\Omega)} \le \varrho \cdot \|\mathbf{c}\|_{\ell^2} < \infty$ for all $\mathbf{c} \in \ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q}$, and

$$\text{Ball}\left(0, 1; B_{p,q}^{\tau}(\Omega; \mathbb{R})\right) \subset Q_{\text{ext}}\left(\text{Ball}\left(0, 1; \ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q}\right)\right) \subset L^2(\Omega). \quad (4.1)$$

Furthermore, Theorem 6 shows that

$$s_{\ell^2(J^{\text{int}})}^*\left(\mathcal{S}_{\mathscr{P}^{\text{int}},\alpha}^{p,q}\right) = s_{\ell^2(J^{\text{ext}})}^*\left(\mathcal{S}_{\mathscr{P}^{\text{ext}},\alpha}^{p,q}\right) = \frac{\alpha}{d} - (\tfrac{1}{2} - \tfrac{1}{p}) = \frac{\tau}{d}$$

and that there exists a Borel probability measure $\mathbb{P}_0$ on $\text{Ball}(0, 1; \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q})$ that is critical for $\text{Ball}(0, 1; \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q})$ with respect to $\ell^2(J^{\text{int}})$. Therefore, we can apply Theorem 5 with the choices $\mathbf{X} = \ell^2(J^{\text{ext}})$, $\mathbf{Y} = \ell^2(J^{\text{int}})$ and $\mathbf{Z} = L^2(\Omega)$ as well as

$$\mathcal{S}_{\mathbf{X}} = \text{Ball}\left(0, 1; \ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q}\right), \qquad \mathcal{S}_{\mathbf{Y}} = \text{Ball}\left(0, 1; \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q}\right),$$

and $\mathcal{S} = \text{Ball}\left(0, 1; B_{p,q}^{\tau}(\Omega; \mathbb{R})\right)$, and finally $\Phi = Q_{\text{ext}}$, $\Psi = Q_{\text{int}}$, and $\kappa = \gamma$. This theorem then shows $s_{L^2(\Omega)}^*(\mathcal{S}) = \frac{\tau}{d} > 0$ (in particular, $\mathcal{S} \subset L^2(\Omega)$ is totally bounded

---

[1] Precisely, this follows by combining Lemmas 11 and 12 and by taking $Q_{\text{int}} = T_{\text{int}} \circ \iota_{\text{int}}$ and $Q_{\text{ext}} = T_{\text{ext}} \circ \iota_{\text{ext}}$.

and hence compact, since $\mathcal{S} \subset L^2(\Omega)$ is closed by Lemma 14) and that $\mathbb{P} := \mathbb{P}_0 \circ Q_{\mathrm{int}}^{-1}$ is a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$.

Finally, Proposition 2 yields a codec $\mathcal{C}^* = \left((E_R^*, D_R^*) \right)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}_X, \ell^2(J^{\mathrm{ext}})}$ satisfying $\delta_{\mathcal{S}_X, \ell^2(J^{\mathrm{ext}})}(E_R^*, D_R^*) \lesssim R^{-\frac{\tau}{d}}$. Furthermore, Q_{ext} is Lipschitz (with respect to $\| \cdot \|_{\ell^2}$ and $\| \cdot \|_{L^2}$) and satisfies (4.1). Thus, Lemma 8 shows that $\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-\frac{\tau}{d}}$ for some codec $\mathcal{C} = \left((E_R, D_R) \right)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, L^2(\Omega)}$.   □

4.2 Sobolev Spaces on Lipschitz Domains $\Omega \subset \mathbb{R}^d$

Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be an open bounded Lipschitz domain (precisely, we require Ω to satisfy the conditions in [33, Chapter VI, Section 3.3]). We consider the usual *Sobolev spaces* $W^{k,p}(\Omega)$ ($k \in \mathbb{N}$ and $p \in [1, \infty]$) and prove that also for the unit balls of these spaces, the phase transition phenomenon holds. To be completely explicit, we endow the space $W^{k,p}(\Omega)$ with the following norm:

$$\| f \|_{W^{k,p}(\Omega)} := \max_{|\alpha| \leq k} \| \partial^\alpha f \|_{L^p(\Omega)}. \tag{4.2}$$

Our phase-transition result reads as follows:

Theorem 8 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be an open bounded Lipschitz domain. Let $k \in \mathbb{N}$ and $p \in [1, \infty]$, and define $\mathcal{S} := \mathrm{Ball}\left(0, 1; W^{k,p}(\Omega) \right)$. If $k > d \cdot (\frac{1}{p} - \frac{1}{2})_+$, then*

(i) $\mathcal{S} \subset L^2(\Omega)$ *is bounded and Borel measurable and satisfies $s^*_{L^2(\Omega)}(\mathcal{S}) = \frac{k}{d}$;*
(ii) *there is a Borel probability measure $\mathbb{P}$ on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$;*
(iii) *there exists a codec $\mathcal{C} = \left((E_R, D_R) \right)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, L^2(\Omega)}$ satisfying*

$$\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-\frac{k}{d}}.$$

Remark 1. As for the case of Besov spaces, the theorem shows that the critical rate $s = s^* = \frac{k}{d}$ is actually attained by a suitable codec.
2. The condition $k > d \cdot (\frac{1}{p} - \frac{1}{2})_+$ is *equivalent* to $\mathcal{S} \subset L^2(\Omega)$ being relatively compact. The sufficiency is a consequence of the *Rellich–Kondrachov theorem*; see [1, Theorem 6.3]. For the necessity, note that if $k \leq d \cdot (\frac{1}{p} - \frac{1}{2})_+$ then necessarily $p < 2$ (since $k > 0$) and thus, $k \leq \frac{d}{p} - \frac{d}{2}$, which implies $kp \leq d - d\frac{p}{2} < d$ and $2 \geq q := \frac{pd}{d-kp}$. However, [1, Example 6.12] shows that the embedding $W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ is *not* compact. Since $L^2(\Omega) \hookrightarrow L^q(\Omega)$ (because of $2 \geq q$), this shows that $\mathcal{S}$ is not a relatively compact subset of $L^2(\Omega)$, since otherwise $W^{k,p}(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow L^q(\Omega)$ would be a compact embedding.   ◀

Proof of Theorem 8 We present here the proof for the case $p \in (1, \infty)$, where we will see that the claim follows from that for the Besov spaces. For the case $p \in \{1, \infty\}$, the proof is more involved and thus, postponed to Appendix D.

First, the Rellich–Kondrachov compactness theorem (see [1, Theorem 6.3]) shows that $W^{k,p}(\Omega)$ embeds compactly into $L^2(\Omega)$. In particular, the ball $\mathcal{S} = \mathrm{Ball}(0, 1; W^{k,p}(\Omega)) \subset L^2(\Omega)$ is bounded; in fact, $\mathcal{S}$ is also compact (hence Borel measurable) by reflexivity [2] of $W^{k,p}(\Omega)$.

Define $\widetilde{p} := \min\{p, 2\}$ and $\widehat{p} := \max\{p, 2\}$, as well as $\mathcal{S}_s := \mathrm{Ball}\big(0, 1; B^k_{p,\widetilde{p}}(\Omega)\big)$ and $\mathcal{S}_b := \mathrm{Ball}\big(0, 1; B^k_{p,\widehat{p}}(\Omega)\big)$; here, the subscript "$s$" stands for "small," while "b" stands for "big," since $\widetilde{p}$ is the small exponent, while $\widehat{p}$ is the large (or big) exponent. We will prove below that there are constants $C_1, C_2 > 0$ such that

$$C_1^{-1} \cdot \mathcal{S}_s = \mathrm{Ball}\big(0, C_1^{-1}; B^k_{p,\widetilde{p}}(\Omega)\big) \subset \mathcal{S} \subset \mathrm{Ball}\big(0, C_2; B^k_{p,\widehat{p}}(\Omega)\big) = C_2 \cdot \mathcal{S}_b \quad (4.3)$$

Assuming for the moment that Equation (4.3) holds, recall from Theorem 7 that $s^*_{L^2(\Omega)}(\mathcal{S}_b) = s^*_{L^2(\Omega)}(\mathcal{S}_s) = \frac{k}{d}$ and that there exists a Borel probability measure $\mathbb{P}_0$ on $\mathcal{S}_s$ that is critical for $\mathcal{S}_s$ with respect to $L^2(\Omega)$. Define $\mathbf{X} := \mathbf{Y} := \mathbf{Z} := L^2(\Omega)$ and $\mathcal{S}_{\mathbf{X}} := \mathcal{S}_b$, $\mathcal{S}_{\mathbf{Y}} := \mathcal{S}_s$, as well as

$$\Phi : \quad \mathcal{S}_b \to L^2(\Omega), \quad f \mapsto C_2 \cdot f \quad \text{and} \quad \Psi : \quad \mathcal{S}_s \to \mathcal{S}, \quad f \mapsto C_1^{-1} \cdot f.$$

Using (4.3), one easily checks that all assumptions of Theorem 5 are satisfied. An application of that theorem shows that $s^*_{L^2(\Omega)}(\mathcal{S}) = \frac{k}{d}$ and that $\mathbb{P} := \mathbb{P}_0 \circ \Psi^{-1}$ is a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$.

Finally, Part (iii) of Theorem 7 shows that $\delta_{\mathcal{S}_b, L^2(\Omega)}(E^*_R, D^*_R) \lesssim R^{-\frac{k}{d}}$ for a suitable codec $\mathcal{C}^* = \big((E^*_R, D^*_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}_b, L^2(\Omega)}$. Next, since Φ is Lipschitz continuous (with respect to $\|\cdot\|_{L^2}$) with $\mathcal{S} \subset \Phi(\mathcal{S}_b)$, Lemma 8 provides a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, L^2(\Omega)}$ satisfying $\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-\frac{k}{d}}$ as well. This establishes Property (iii) of the current theorem.

It remains to prove (4.3). First, a combination of [36, Theorem in Section 2.5.6] and [36, Proposition 2 in Section 2.3.2] shows for the so-called *Triebel–Lizorkin spaces* [3] $F^k_{p,2}(\mathbb{R}^d)$ that

$$B^k_{p,\widetilde{p}}(\mathbb{R}^d) \hookrightarrow F^k_{p,2}(\mathbb{R}^d) = W^{k,p}(\mathbb{R}^d) \hookrightarrow B^k_{p,\widehat{p}}(\mathbb{R}^d).$$

Hence, there are $C_3, C_4 > 0$ satisfying $\|f\|_{W^{k,p}(\mathbb{R}^d)} \leq C_3 \cdot \|f\|_{B^k_{p,\widetilde{p}}(\mathbb{R}^d)}$ for all $f \in B^k_{p,\widetilde{p}}(\mathbb{R}^d)$, and $\|f\|_{B^k_{p,\widehat{p}}(\mathbb{R}^d)} \leq C_4 \cdot \|f\|_{W^{k,p}(\mathbb{R}^d)}$ for all $f \in W^{k,p}(\mathbb{R}^d)$. Furthermore, since Ω is a Lipschitz domain, [33, Chapter VI, Theorem 5] shows that

---

[2] Indeed, if $(f_n)_{n \in \mathbb{N}} \subset \mathcal{S}$ is arbitrary, then since $W^{k,p}(\Omega)$ is reflexive (see [2, Example 8.11]), the closed unit ball in $W^{k,p}(\Omega)$ is weakly sequentially compact (see [2, Theorem 8.10]), so that there is a subsequence $(f_{n_\ell})_{\ell \in \mathbb{N}}$ satisfying $f_{n_\ell} \xrightarrow{W^{k,p}(\Omega)} f \in \mathcal{S}$ (weak convergence in $W^{k,p}(\Omega)$). Again by compactness of the embedding $W^{k,p}(\Omega) \hookrightarrow L^2(\Omega)$, this implies $f_{n_\ell} \xrightarrow{L^2} f \in \mathcal{S}$ (see, e.g., [6, Chapter VI, Proposition 3.3]), showing that $\mathcal{S} \subset L^2(\Omega)$ is compact.

[3] The precise definition of these spaces is immaterial for us. We merely remark that the identity $F^k_{p,2}(\mathbb{R}^d) = W^{k,p}(\mathbb{R}^d)$ is only valid for $p \in (1, \infty)$.

there is a bounded linear "extension operator" $\mathscr{E} : W^{k,p}(\Omega) \to W^{k,p}(\mathbb{R}^d)$ satisfying $(\mathscr{E} f)|_\Omega = f$ for all $f \in W^{k,p}(\Omega)$.

It is now easy to prove the inclusion (4.3), with $C_1 := C_3$ and $C_2 := C_4 \cdot \|\mathscr{E}\|$. First, if $f \in \mathcal{S}_s$ and $\varepsilon > 0$, then (by the definition of Besov spaces on domains; see Eqs. (C.1) and (C.2)) there exists $g \in B^k_{p,\widetilde{p}}(\mathbb{R}^d)$ satisfying $f = g|_\Omega$ and $\|g\|_{B^k_{p,\widetilde{p}}(\mathbb{R}^d)} \leq 1 + \varepsilon$. Hence,

$$\|f\|_{W^{k,p}(\Omega)} = \|g|_\Omega\|_{W^{k,p}(\Omega)} \leq \|g\|_{W^{k,p}(\mathbb{R}^d)} \leq C_3 (1 + \varepsilon).$$

Since this holds for all $\varepsilon > 0$, we see that $\|C_1^{-1} f\|_{W^{k,p}(\Omega)} \leq 1$; that is, $C_1^{-1} f \in \mathcal{S}$.

Conversely, if $f \in \mathcal{S}$, then $g := \mathscr{E} f \in W^{k,p}(\mathbb{R}^d) \subset B^k_{p,\widehat{p}}(\mathbb{R}^d)$ and $f = g|_\Omega$, which implies

$$\|f\|_{B^k_{p,\widehat{p}}(\Omega)} \leq \|g\|_{B^k_{p,\widehat{p}}(\mathbb{R}^d)} \leq C_4 \|g\|_{W^{k,p}(\mathbb{R}^d)} \leq C_4 \|\mathscr{E}\| \cdot \|f\|_{W^{k,p}(\Omega)} \leq C_2,$$

and hence $f \in C_2 \cdot \mathcal{S}_b$. $\qquad\square$

Declarations

A Transferring Approximation Rates and Measures

In this appendix, we provide the proof of Theorem 5. Along the way, we will show that expansive maps can be used to transfer measures with a certain growth order from one set to another, while Lipschitz maps can be used to transfer estimates for the optimal compression rate from one set to another.

Lemma 7 *Let* $\mathbf{X}$, $\mathbf{Y}$ *be Banach spaces and let* $\mathcal{S} \subset \mathbf{X}$ *and* $\mathcal{S}' \subset \mathbf{Y}$. *Assume that there exists a (not necessarily surjective) function* $\Phi : \mathcal{S} \to \mathcal{S}'$ *which is measurable (with respect to the trace σ-algebras of the Borel σ-algebras) and* expansive, *in the sense that there exists* $\kappa > 0$ *such that*

$$\|\Phi(\mathbf{x}) - \Phi(\mathbf{x}')\|_{\mathbf{Y}} \geq \kappa \cdot \|\mathbf{x} - \mathbf{x}'\|_{\mathbf{X}} \qquad \forall \, \mathbf{x}, \mathbf{x}' \in \mathcal{S}.$$

If $s_0 \geq 0$ and if $\mathbb{P}$ is a Borel probability measure on $\mathcal{S}$ of growth order s_0, then the push-forward measure $\mathbb{P} \circ \Phi^{-1}$ is a Borel probability measure on $\mathcal{S}'$ of growth order s_0 as well.

Proof Since $\Phi : \mathcal{S} \to \mathcal{S}'$ is measurable, $\nu := \mathbb{P} \circ \Phi^{-1}$ is a Borel probability measure on $\mathcal{S}'$.

To prove that ν has growth order s_0, let $s > s_0$ be arbitrary. Since $\mathbb{P}$ is of growth order s_0, there are $\varepsilon_0, c > 0$ such that Eq. (2.1) is satisfied. Define $\varepsilon_0' := \frac{\kappa}{2} \cdot \varepsilon_0$ and $c' := c \cdot (2\kappa^{-1})^{-1/s} = 2^{-1/s} c \cdot \kappa^{1/s}$. We claim that $\nu\big(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y})\big) \leq 2^{-c' \cdot \varepsilon^{-1/s}}$ for all $\mathbf{y} \in \mathbf{Y}$ and all $\varepsilon \in (0, \varepsilon_0')$; this will show that ν has growth order s_0.

The estimate is trivial if $\Phi(\mathcal{S}) \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y}) = \emptyset$, since then $\Phi^{-1}\big(\mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y})\big) = \emptyset$, and hence $\nu\big(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y})\big) = \mathbb{P}\big(\Phi^{-1}(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y}))\big) = \mathbb{P}(\emptyset) = 0$. Therefore, let us assume that $\emptyset \neq \Phi(\mathcal{S}) \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y}) \ni \mathbf{y}'$; say $\mathbf{y}' = \Phi(\mathbf{x}')$ for some $\mathbf{x}' \in \mathcal{S}$. Now, for arbitrary $\mathbf{x} \in \Phi^{-1}\big(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y})\big) \subset \mathcal{S}$, we have

$$\|\mathbf{x} - \mathbf{x}'\|_{\mathbf{X}} \leq \kappa^{-1} \|\Phi(\mathbf{x}) - \Phi(\mathbf{x}')\|_{\mathbf{Y}} \leq \kappa^{-1} \cdot \big(\|\Phi(\mathbf{x}) - \mathbf{y}\|_{\mathbf{Y}} + \|\mathbf{y} - \Phi(\mathbf{x}')\|_{\mathbf{Y}}\big) \leq 2 \cdot \kappa^{-1} \cdot \varepsilon.$$

We have thus shown $\Phi^{-1}(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y})) \subset \mathcal{S} \cap \mathrm{Ball}(\mathbf{x}', \frac{2}{\kappa}\varepsilon; \mathbf{X})$. Since $\frac{2}{\kappa}\varepsilon < \frac{2}{\kappa}\varepsilon_0' = \varepsilon_0$, we see by Property (2.1) as claimed that

$$\begin{aligned}
\nu\big(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon, \mathbf{Y})\big) &= \mathbb{P}\big(\Phi^{-1}(\mathcal{S}' \cap \mathrm{Ball}(\mathbf{y}, \varepsilon; \mathbf{Y}))\big) \\
&\leq \mathbb{P}\Big(\mathcal{S} \cap \mathrm{Ball}\left(\mathbf{x}', \tfrac{2}{\kappa}\varepsilon; \mathbf{X}\right)\Big) \\
&\leq 2^{-c \cdot (2\kappa^{-1}\varepsilon)^{-1/s}} = 2^{-c' \cdot \varepsilon^{-1/s}}.
\end{aligned}$$

$\square$

As a kind of converse of the previous result, we now show that Lipschitz maps can be used to obtain bounds for the optimal compression rate $s_{\mathbf{X}}^*(\mathcal{S})$ of a signal class $\mathcal{S} \subset \mathbf{X}$.

Lemma 8 *Let $\mathbf{X}, \mathbf{Y}$ be Banach spaces, and let $\mathcal{S} \subset \mathbf{X}$ and $\mathcal{S}' \subset \mathbf{Y}$. Assume that $\Phi : \mathcal{S} \to \mathbf{Y}$ is Lipschitz continuous, and that $\Phi(\mathcal{S}) \supset \mathcal{S}'$. Then, $s_{\mathbf{Y}}^*(\mathcal{S}') \geq s_{\mathbf{X}}^*(\mathcal{S})$.*

In fact, given a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}}$ satisfying $\delta_{\mathcal{S}, \mathbf{X}}(E_R, D_R) \lesssim R^{-s}$ for some $s \geq 0$, one can construct a modified codec $\mathcal{C}^* = \big((E_R^*, D_R^*)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}', \mathbf{Y}}$ satisfying $\delta_{\mathcal{S}', \mathbf{Y}}(E_R^*, D_R^*) \lesssim R^{-s}$.

Proof The claim is clear if $s_{\mathbf{X}}^*(\mathcal{S}) = 0$. Thus, let us assume $s_{\mathbf{X}}^*(\mathcal{S}) > 0$, and let $s \in \big[0, s_{\mathbf{X}}^*(\mathcal{S})\big)$ be arbitrary. Then, there is a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, \mathbf{X}}$ and a constant $C > 0$ such that $\delta_{\mathcal{S}, \mathbf{X}}(E_R, D_R) \leq C \cdot R^{-s}$ for all $R \in \mathbb{N}$. Let $L > 0$ denote a Lipschitz constant for Φ.

Now, for $\varepsilon > 0$ and $\mathbf{x} \in \mathbf{X}$, choose $\Psi_\varepsilon(\mathbf{x}) \in \mathcal{S}$ such that $\|\mathbf{x} - \Psi_\varepsilon(\mathbf{x})\|_{\mathbf{X}} \leq \varepsilon + \mathrm{dist}(\mathbf{x}, \mathcal{S})$, and let

$$D_R^* : \quad \{0, 1\}^R \to \mathbf{Y}, \quad c \mapsto \Phi\big(\Psi_{R^{-s}}(D_R(c))\big) \qquad \text{for } R \in \mathbb{N}.$$

Now, if $\mathbf{y} \in \mathcal{S}' \subset \Phi(\mathcal{S})$ is arbitrary, then $\mathbf{y} = \Phi(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{S}$, and hence

$$
\begin{aligned}
\|\mathbf{y} - D_R^*(E_R(\mathbf{x}))\|_{\mathbf{Y}} &= \left\| \Phi(\mathbf{x}) - \Phi\Big(\Psi_{R-s}\big(D_R(E_R(\mathbf{x}))\big)\Big) \right\|_{\mathbf{Y}} \\
&\le L \cdot \left\| \mathbf{x} - \Psi_{R-s}\big(D_R(E_R(\mathbf{x}))\big) \right\|_{\mathbf{X}} \\
&\le L \cdot \left[\|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} + \left\| D_R(E_R(\mathbf{x})) - \Psi_{R-s}\big(D_R(E_R(\mathbf{x}))\big) \right\|_{\mathbf{X}} \right] \\
&\le L \cdot \left[C \cdot R^{-s} + R^{-s} + \mathrm{dist}\big(D_R(E_R(\mathbf{x})), \mathcal{S}\big) \right] \le L \cdot (1 + 2C) \cdot R^{-s},
\end{aligned}
$$

since $\mathrm{dist}\big(D_R(E_R(\mathbf{x})), \mathcal{S}\big) \le \|D_R(E_R(\mathbf{x})) - \mathbf{x}\|_{\mathbf{X}} \le C \cdot R^{-s}$. Therefore, if for each $\mathbf{y} \in \mathcal{S}'$ and $R \in \mathbb{N}$ we choose $c_{\mathbf{y},R} \in \{0,1\}^R$ with $\|\mathbf{y} - D_R^*(c_{\mathbf{y},R})\|_{\mathbf{Y}} = \min_{c \in \{0,1\}^R} \|\mathbf{y} - D_R^*(c)\|_{\mathbf{Y}}$ and define $E_R^* : \mathcal{S}' \to \{0,1\}^R$, $\mathbf{y} \mapsto c_{\mathbf{y},R}$, then $\|\mathbf{y} - D_R^*(E_R^*(\mathbf{y}))\|_{\mathbf{Y}} \le L \cdot (1 + 2C) \cdot R^{-s}$ for all $\mathbf{y} \in \mathcal{S}'$ and $R \in \mathbb{N}$, and hence $s_{\mathbf{Y}}^*(\mathcal{S}') \ge s$. Since $s \in [0, s_{\mathbf{X}}^*(\mathcal{S}))$ was arbitrary, this completes the proof. $\qquad\square$

The following lemma shows that if a signal class $\mathcal{S} \subset \mathbf{X}$ carries a Borel probability measure of growth order s_0 and satisfies $s_{\mathbf{X}}^*(\mathcal{S}) \ge s_0$, then in fact $s_{\mathbf{X}}^*(\mathcal{S}) = s_0$. This is elementary, but will be used quite frequently, so that we prefer to state it as a lemma.

Lemma 9 *Let $s_0 \in [0, \infty)$, let $\mathbf{X}$ be a Banach space, and let $\mathcal{S} \subset \mathbf{X}$. Assume that there exists a Borel probability measure $\mathbb{P}$ of growth order s_0 on $\mathcal{S}$ and that $s_{\mathbf{X}}^*(\mathcal{S}) \ge s_0$. Then, $s_{\mathbf{X}}^*(\mathcal{S}) = s_0$ and $\mathbb{P}$ is critical for $\mathcal{S}$ with respect to $\mathbf{X}$.*

Proof Corollary 2 shows that $s_0 \ge s_{\mathbf{X}}^*(\mathcal{S})$. Since $s_0 \le s_{\mathbf{X}}^*(\mathcal{S})$ by assumption, the claim of the lemma follows. $\qquad\square$

We finally provide the proof of Theorem 5.

Proof of Theorem 5 Since $\Phi : \mathcal{S}_{\mathbf{X}} \to \mathbf{Z}$ is Lipschitz continuous with $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$, Lemma 8 shows that $s_{\mathbf{Z}}^*(\mathcal{S}) \ge s_{\mathbf{X}}^*(\mathcal{S}_{\mathbf{X}}) =: s^*$. Furthermore, since $\Psi : \mathcal{S}_{\mathbf{Y}} \to \mathcal{S}$ is measurable and expansive and $\mathbb{P}$ has growth order $s_{\mathbf{Y}}^*(\mathcal{S}_{\mathbf{Y}}) = s^*$, Lemma 7 shows that $\nu := \mathbb{P} \circ \Psi^{-1}$ is a Borel probability measure on $\mathcal{S}$ of growth order s^* as well. Now, Lemma 9 shows that $s_{\mathbf{Z}}^*(\mathcal{S}) = s^*$ and that ν is critical for $\mathcal{S}$ with respect to $\mathbf{Z}$. $\qquad\square$

B A Lower Bound for the Optimal Compression Rate $s_{\ell^2(\mathcal{I})}^*\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}\big)$

Our goal in this appendix is to show that the optimal compression rate for the class $\mathcal{S}_{\mathscr{P},\alpha}^{p,q}$ satisfies $s_{\ell^2(\mathcal{I})}^*\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}\big) \ge \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$, assuming that $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$. Our proof of this fact relies on an equivalence between the optimal distortion for a set and the so-called *entropy numbers* of that set. By combining this equivalence with known estimates for the entropy numbers of certain embeddings between sequence spaces (taken from [20]), we will obtain the claim.

First, let us describe the equivalence between the optimal achievable distortion and the entropy numbers of a set. Following [5,12], given a (quasi-)Banach space $\mathbf{X}$, a set $M \subset \mathbf{X}$, and $k \in \mathbb{N}$, the *k-th entropy number* $e_k(M) := e_k(M; \mathbf{X})$ of M is defined as

$$
e_k(M; \mathbf{X}) := \inf \left\{ \varepsilon > 0 \;\middle|\; \exists\, \mathbf{x}_1, \ldots, \mathbf{x}_{2^{k-1}} \in \mathbf{X} : M \subset \bigcup_{i=1}^{2^{k-1}} \mathrm{Ball}(\mathbf{x}_i, \varepsilon; \mathbf{X}) \right\} \in [0, \infty],
$$

<div align="right">(B.1)</div>

with the convention that $\inf \emptyset = \infty$. Note that $e_k(M)$ is finite if and only if M is bounded. Furthermore, $e_k(M) \xrightarrow[k \to \infty]{} 0$ if and only if $M \subset \mathbf{X}$ is totally bounded. Finally, if $\mathbf{Y}$ is a further (quasi)-Banach space, and $T : \mathbf{Y} \to \mathbf{X}$ is linear, then the entropy numbers $e_k(T)$ are defined as $e_k(T) := e_k\big(T(\mathrm{Ball}(0, 1; \mathbf{Y})); \mathbf{X}\big)$.

For proving that $s^*_{\ell^2(\mathcal{I})}\big(\mathcal{S}^{p,q}_{\mathcal{P},\alpha}\big) \geq \frac{\alpha}{d} - (\frac{1}{2} - \frac{1}{p})$, we will use the following folklore equivalence between entropy numbers and the optimal achievable distortion for a given set:

Lemma 10 *Let $\mathbf{X}$ be a Banach space and $\emptyset \neq \mathcal{S} \subset \mathbf{X}$. Then,*

$$e_{R+1}(\mathcal{S}; \mathbf{X}) = \inf \big\{\delta_{\mathcal{S}, \mathbf{X}}(E_R, D_R) : (E_R, D_R) \in \mathrm{Enc}^R_{\mathcal{S}, \mathbf{X}}\big\} \text{ for all } R \in \mathbb{N}.$$

Proof The intuition is that a covering of $\mathcal{S}$ by 2^R balls $\mathrm{Ball}(\mathbf{x}_i, \varepsilon; \mathbf{X})$ ($i \in \{1, \dots, 2^R\}$) gives rise to an encoder/decoder pair with R bits achieving distortion ε, by mapping each $\mathbf{u} \in \mathcal{S}$ to an index $i_{\mathbf{u}} \in \{1, \dots, 2^R\} \cong \{0, 1\}^R$ such that $\mathbf{u} \in \mathrm{Ball}(\mathbf{x}_{i_{\mathbf{u}}}, \varepsilon; \mathbf{X})$. The reconstruction map is of the form $i \mapsto \mathbf{x}_i$. Conversely, any encoder/decoder pair using R bits and achieving distortion ε induces a covering of $\mathcal{S}$ by 2^R balls of radius ε. For more details, we refer to Remark 1 and [13, Section IV]. $\square$

In addition to this equivalence between entropy numbers and best achievable distortion, we will use two results from [20] about the asymptotic behavior of the entropy numbers of certain sequence spaces. The following definition introduces the terminology used in [20].

Definition 8 (see [20, Equations (10), (11), and Definition 1])
A sequence $(\beta_j)_{j \in \mathbb{N}_0} \subset (0, \infty)$ is called

- an *admissible sequence* if there exist $d_0, d_1 \in (0, \infty)$ such that $d_0 \beta_j \leq \beta_{j+1} \leq d_1 \beta_j$ for all $j \in \mathbb{N}_0$;
- *almost strongly increasing* if there exists $\kappa \in \mathbb{N}$ such that $2\beta_j \leq \beta_k$ for all $j, k \in \mathbb{N}_0$ with $k \geq j + \kappa$.

Given $p, q \in (0, \infty]$ and sequences $\boldsymbol{\beta} = (\beta_j)_{j \in \mathbb{N}_0} \subset (0, \infty)$ and $\mathbf{N} = (N_j)_{j \in \mathbb{N}_0} \subset \mathbb{N}$, define $J_{\mathbf{N}} := \{(j, \ell) \in \mathbb{N}_0 \times \mathbb{N} : 1 \leq \ell \leq N_j\}$ and

$$\|\mathbf{x}\|_{\ell^q(\beta_j \ell^p_{N_j})} := \left\| \left(\beta_j \left\| (x_{j,\ell})_{\ell \in \{1, \dots, N_j\}} \right\|_{\ell^p} \right)_{j \in \mathbb{N}_0} \right\|_{\ell^q} \in [0, \infty] \text{ for } \mathbf{x} = (x_{j,\ell})_{(j,\ell) \in J_{\mathbf{N}}} \in \mathbb{C}^{J_{\mathbf{N}}},$$

as well as $\ell^q(\beta_j \ell^p_{N_j}) := \big\{\mathbf{x} \in \mathbb{C}^{J_{\mathbf{N}}} : \|\mathbf{x}\|_{\ell^q(\beta_j \ell^p_{N_j})} < \infty\big\}$. For the case $\boldsymbol{\beta} = (1)_{j \in \mathbb{N}_0}$, we simply write $\ell^q(\ell^p_{N_j})$ instead of $\ell^q(\beta_j \ell^p_{N_j})$.

Using these notions, Leopold [20] proved the following results:

Theorem 9 (see [20, Theorems 3 and 4])
Let $p_1, p_2, q_1, q_2 \in (0, \infty]$, and let $\mathbf{N} = (N_j)_{j \in \mathbb{N}_0} \subset \mathbb{N}$ and $\boldsymbol{\beta} = (\beta_j)_{j \in \mathbb{N}_0} \subset (0, \infty)$ both be admissible, almost strongly increasing sequences.
Assume that either

(i) $p_1 \leq p_2$; or

(ii) $p_2 < p_1$ and the sequence $\left(\beta_j \cdot N_j^{\frac{1}{p_1} - \frac{1}{p_2}}\right)_{j \in \mathbb{N}_0}$ is almost strongly increasing.

Then, the embedding $\ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}) \hookrightarrow \ell^{q_2}(\ell_{N_j}^{p_2})$ holds, and there exist $C_1, C_2 > 0$ such that for all $L \in \mathbb{N}$, the entropy numbers $e_n\big(\mathrm{id} : \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}) \to \ell^{q_2}(\ell_{N_j}^{p_2})\big)$ (see the discussion around Eq. (B.1)) satisfy

$$C_1 \cdot \beta_L^{-1} \, N_L^{-\left(\frac{1}{p_1} - \frac{1}{p_2}\right)} \leq e_{2N_L}\big(\mathrm{id} : \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}) \to \ell^{q_2}(\ell_{N_j}^{p_2})\big) \leq C_2 \cdot \beta_L^{-1} \, N_L^{-\left(\frac{1}{p_1} - \frac{1}{p_2}\right)}.$$

Remark For what follows, only the *upper* bound in Theorem 9 will matter. Even though Theorem 9 pertain to spaces of *complex* sequences, this upper bound also holds for the real-valued case. To see this, note that if we denote by Re $\mathbf{x}$ the (componentwise) real part of the sequence $\mathbf{x}$, then clearly $\|\mathrm{Re}\,\mathbf{x}\|_{\ell^q(\beta_j \, \ell_{N_j}^p)} \leq \|\mathbf{x}\|_{\ell^q(\beta_j \, \ell_{N_j}^p)}$. Hence, defining the real-valued version of the space $\ell^q(\beta_j \, \ell_{N_j}^p)$ as

$$\ell^q(\beta_j \, \ell_{N_j}^p; \mathbb{R}) := \big\{\mathbf{x} \in \mathbb{R}^{J_\mathbf{N}} : \|\mathbf{x}\|_{\ell^q(\beta_j \, \ell_{N_j}^p)} < \infty\big\},$$

we see that if Ball $\big(0, 1; \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1})\big) \subset \bigcup_{i=1}^M$ Ball $\big(\mathbf{x}_i, \varepsilon; \ell^{q_2}(\ell_{N_j}^{p_2})\big)$ for $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \ell^{q_2}(\ell_{N_j}^{p_2})$, then Ball $\big(0, 1; \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}; \mathbb{R})\big) \subset \bigcup_{i=1}^M$ Ball $\big(\mathrm{Re}\,\mathbf{x}_i, \varepsilon; \ell^{q_2}(\ell_{N_j}^{p_2}; \mathbb{R})\big)$, and hence

$$e_k\big(\mathrm{id} : \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}; \mathbb{R}) \to \ell^{q_2}(\ell_{N_j}^{p_2}; \mathbb{R})\big) \leq e_k\big(\mathrm{id} : \ell^{q_1}(\beta_j \, \ell_{N_j}^{p_1}) \to \ell^{q_2}(\ell_{N_j}^{p_2})\big) \quad \forall k \in \mathbb{N}.$$
$$\tag{B.2}$$

Proof of Proposition 2 Let $n_m := |\mathcal{I}_m|$ and $N_j := n_{j+1}$ for $m \in \mathbb{N}$ and $j \in \mathbb{N}_0$. Further, set $\kappa := \lceil d^{-1} \cdot (1 + \log_2(A/a)) \rceil$, where we recall from Eq. (3.1) that $a, A > 0$ satisfy $a \cdot 2^{dm} \leq n_m \leq A \cdot 2^{dm}$. Thus, $N_{j+1} = n_{j+2} \asymp 2^{d(j+2)} = 2^d \, 2^{d(j+1)} \asymp n_{j+1} = N_j$, which shows that $\mathbf{N} = (N_j)_{j \in \mathbb{N}_0}$ is admissible. Furthermore, if $k \geq j + \kappa$, then

$$N_k = n_{k+1} \geq a \cdot 2^{d(k+1)} \geq a \cdot 2^{d(j+1)+1+\log_2(A/a)} = 2 \cdot A \cdot 2^{d(j+1)} \geq 2\,n_{j+1} = 2N_j,$$

which shows that $\mathbf{N}$ is almost strongly increasing.

Next, define $\beta_j := 2^{\alpha(j+1)}$ for $j \in \mathbb{N}_0$, noting that $\beta_{j+1} = 2^\alpha \, \beta_j$, which implies that $(\beta_j)_{j \in \mathbb{N}_0}$ is admissible. Furthermore, if $k \geq j + \lceil \alpha^{-1} \rceil$, then $\beta_k \geq 2 \cdot 2^{\alpha(j+1)} = 2\,\beta_j$, so that $(\beta_j)_{j \in \mathbb{N}_0}$ is also almost strongly increasing. Here, we used that $\alpha > 0$.

Finally, for each $m \in \mathbb{N}$ pick a bijection $\iota_m : [N_{m-1}] \to \mathcal{I}_m$ (which is possible since $N_{m-1} = n_m = |\mathcal{I}_m|$) and define

$$\Psi : \quad \mathbb{R}^\mathcal{I} \to \mathbb{R}^{J_\mathbf{N}}, \quad \mathbf{x} = (x_i)_{i \in \mathcal{I}} \mapsto \big(x_{\iota_{j+1}(\ell)}\big)_{(j,\ell) \in J_\mathbf{N}}.$$

It is easy to see that Ψ is a bijection, and that

$$
\begin{aligned}
\|\Psi(\mathbf{x})\|_{\ell^q(\beta_j\,\ell^p_{N_j})} &= \left\|\left(\beta_j\,\big\|(x_{\iota_{j+1}(\ell)})_{\ell\in[N_j]}\big\|_{\ell^p}\right)_{j\in\mathbb{N}_0}\right\|_{\ell^q} \\
&= \left\|\left(2^{\alpha m}\,\|\mathbf{x}_m\|_{\ell^p(\mathcal{I}_m)}\right)_{m\in\mathbb{N}}\right\|_{\ell^q} = \|\mathbf{x}\|_{\ell^{p,q}_{\mathscr{P},\alpha}}
\end{aligned}
$$

for arbitrary $p,q\in(0,\infty]$. Here, $\mathbf{x}_m=(x_i)_{i\in\mathcal{I}_m}$ is as defined in Eq. (3.2). In the same way, we see $\|\Psi(\mathbf{x})\|_{\ell^q(\ell^p_{N_j})} = \|\mathbf{x}\|_{\ell^{p,q}_{\mathscr{P},0}}$ and also $\|\Psi(\mathbf{x})\|_{\ell^2(\ell^2_{N_j})} = \|\mathbf{x}\|_{\ell^{2,2}_{\mathscr{P},0}} = \|\mathbf{x}\|_{\ell^2(\mathcal{I})}$. Using these identities, it is straightforward to see that $\ell^{p,q}_{\mathscr{P},\alpha}\hookrightarrow\ell^2(\mathcal{I})$ holds if and only if $\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})$, and furthermore that

$$
e_k(\mathcal{S}^{p,q}_{\mathscr{P},\alpha};\ell^2(\mathcal{I}))=e_k\big(\ell^{p,q}_{\mathscr{P},\alpha}\hookrightarrow\ell^2(\mathcal{I})\big)=e_k\big(\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})\big)\qquad\forall k\in\mathbb{N}.
$$

There are now two cases. First, if $p\le 2$, then Eq. (B.2) and the first part of Theorem 9 with $p_1=p,q_1=q$ and $p_2=q_2=2$ show that $\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})$, and yield a constant $C_1>0$ such that

$$
e_{2N_L}\big(\mathcal{S}^{p,q}_{\mathscr{P},\alpha};\ell^2(\mathcal{I})\big) = e_{2N_L}\big(\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})\big) \le C_1\cdot\beta_L^{-1}\cdot N_L^{-(\frac{1}{p}-\frac{1}{2})}
$$

for all $L\in\mathbb{N}$.

If otherwise $p>2$, then $\frac{1}{2}-\frac{1}{p}>0$, so that our assumptions concerning α imply that $\alpha>d\cdot(\frac{1}{2}-\frac{1}{p})_+=d\cdot(\frac{1}{2}-\frac{1}{p})$, and hence $\gamma:=\alpha+d\cdot(\frac{1}{p}-\frac{1}{2})>0$. Therefore, the sequence $(K_j)_{j\in\mathbb{N}_0}:=\big(\beta_j\cdot N_j^{\frac{1}{p}-\frac{1}{2}}\big)_{j\in\mathbb{N}_0}$ is almost strongly increasing; indeed, if $k\ge j+\lceil\gamma^{-1}\cdot\big(1+\log_2[(a/A)^{\frac{1}{p}-\frac{1}{2}}]\big)\rceil$, then we see because of $a\cdot 2^{d(j+1)}\le N_j\le A\cdot 2^{d(j+1)}$ and $\frac{1}{p}-\frac{1}{2}<0$ that $N_k^{\frac{1}{p}-\frac{1}{2}}\ge A^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{d(k+1)(\frac{1}{p}-\frac{1}{2})}$ and $N_j^{\frac{1}{p}-\frac{1}{2}}\le a^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{d(j+1)(\frac{1}{p}-\frac{1}{2})}$. Since $\gamma>0$, we thus see

$$
\begin{aligned}
K_k = \beta_k\cdot N_k^{\frac{1}{p}-\frac{1}{2}} &\ge A^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{\alpha(k+1)}\,2^{d(k+1)(\frac{1}{p}-\frac{1}{2})} = 2^\gamma A^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{\gamma k} \\
&\ge 2^\gamma A^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{\gamma j}\cdot 2\cdot(a/A)^{\frac{1}{p}-\frac{1}{2}} = 2a^{\frac{1}{p}-\frac{1}{2}}\cdot 2^{\alpha(j+1)}\,2^{d(j+1)(\frac{1}{p}-\frac{1}{2})} \\
&\ge 2\cdot 2^{\alpha(j+1)}\cdot N_j^{\frac{1}{p}-\frac{1}{2}} = 2\,K_j.
\end{aligned}
$$

Thus, Part (ii) of Theorem 9 and Eq. (B.2) show that $\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})$, and that there is a constant $C_2>0$ such that

$$
e_{2N_L}\big(\mathcal{S}^{p,q}_{\mathscr{P},\alpha};\ell^2(\mathcal{I})\big) = e_{2N_L}\big(\ell^q(\beta_j\,\ell^p_{N_j};\mathbb{R})\hookrightarrow\ell^2(\ell^2_{N_j};\mathbb{R})\big) \le C_2\cdot\beta_L^{-1}\cdot N_L^{-(\frac{1}{p}-\frac{1}{2})}
$$

for all $L\in\mathbb{N}$.

Define $C_3 := \max\{C_1, C_2\}$ and note that the preceding estimates only yield bounds for the entropy numbers $e_k(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I}))$ in case of $k = 2N_L$ for some $L \in \mathbb{N}$, not for general $k \in \mathbb{N}$. This, however, suffices to handle the general case. Indeed, let $R \in \mathbb{N}$ with $R \geq 2N_1$ be arbitrary, and let $L \in \mathbb{N}$ be maximal with $2N_L \leq R + 1$; this is possible since $N_L \to \infty$ as $L \to \infty$. Note $R \leq R + 1 < 2N_{L+1} = 2n_{L+2} \leq 2A\, 2^{d(L+2)} = 2^{2d+1}A\, 2^{dL}$ by maximality. Since the sequence of entropy numbers $\big(e_k(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I}))\big)_{k \in \mathbb{N}}$ is non-increasing, we thus see

$$
e_{R+1}\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I})\big) \leq e_{2N_L}\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I})\big) \leq C_3 \cdot \beta_L^{-1} \cdot N_L^{-(\frac{1}{p}-\frac{1}{2})}
$$

$$
(\text{since } N_L = n_{L+1} \asymp 2^{dL}) \leq C_4 \cdot 2^{-\alpha L} \cdot 2^{(\frac{1}{2}-\frac{1}{p})dL} = C_4 \cdot \big(2^{dL}\big)^{-(\frac{\alpha}{d}+\frac{1}{p}-\frac{1}{2})}
$$

$$
(\text{since } \tfrac{\alpha}{d}+\tfrac{1}{p}-\tfrac{1}{2}>0) \leq C_5 \cdot R^{-(\frac{\alpha}{d}+\frac{1}{p}-\frac{1}{2})},
$$

for all $R \geq 2N_1$ and suitable constants $C_4, C_5 > 0$ which are independent of R.

Now, since $\mathcal{S}_{\mathscr{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$ is bounded (otherwise, all entropy numbers would be infinite), it is easy to see $e_{R+1}(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I})) \leq e_1(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I})) \lesssim 1 \lesssim R^{-(\frac{\alpha}{d}+\frac{1}{p}-\frac{1}{2})}$ for $R \in \mathbb{N}$ with $R < 2N_1$. With this, the claim $s_{\ell^2(\mathcal{I})}^*\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}\big) \geq \frac{\alpha}{d} - \big(\frac{1}{2} - \frac{1}{p}\big)$ and the existence of a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}_{\mathscr{P},\alpha}^{p,q}, \ell^2(\mathcal{I})}$ satisfying (3.3) follow from the relation between entropy numbers and optimal distortion described in Lemma 10.

Finally, since $e_R\big(\mathcal{S}_{\mathscr{P},\alpha}^{p,q}; \ell^2(\mathcal{I})\big) \to 0$ as $R \to \infty$, it follows that $\mathcal{S}_{\mathscr{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$ is totally bounded. Since $\mathcal{S}_{\mathscr{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$ is also easily seen to be closed (this essentially follows from Fatou's lemma), we see that $\mathcal{S}_{\mathscr{P},\alpha}^{p,q} \subset \ell^2(\mathcal{I})$ is compact.　□

C A Review of Besov Spaces

In this subsection, we review the relevant properties of Besov spaces on $\mathbb{R}^d$ and on domains, including the characterization of these spaces in terms of wavelets; see Sect. C.2.

Before we dive into the details, a word of caution is in order. In the literature, there are two common definitions of Besov spaces: A Fourier analytic definition and a definition using moduli of continuity. Here, we only consider the former definition; the reader interested in the latter is referred to [9]. It should be mentioned, however, that the two definitions do *not* agree in general; see for instance [16]. Nevertheless, in the regime that we are interested in, the two definitions coincide, as can be deduced from [36, Theorem in Section 2.5.12]. Since we focus on the Fourier analytic definition only, we omit the details.

C.1 The (Fourier-analytic) Definition of Besov Spaces

Our presentation here follows [36, Section 2.3] and [35, Section 1.3]. In this section, all functions are taken to be complex-valued, unless indicated otherwise. Let $\mathscr{S}(\mathbb{R}^d)$ denote the space of *Schwartz functions* (see, for instance, [14, Section 8.1]), and $\mathscr{S}'(\mathbb{R}^d)$ its topological dual space, the space of *tempered distributions* (see [14, Section 9.2]). We use the *Fourier transform* on $L^1(\mathbb{R}^d)$ with the same normalization as in [36,39]; that is,

$$\widehat{f}(\xi) := \mathcal{F}f(\xi) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\langle x, \xi \rangle} \, dx \quad \text{for} \quad f \in L^1(\mathbb{R}^d) \text{ and } \xi \in \mathbb{R}^d,$$

where $\langle x, \xi \rangle = \sum_{j=1}^d x_j \xi_j$ denotes the standard inner product on $\mathbb{R}^d$. With this normalization, the Fourier transform $\mathcal{F} : L^1(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ extends to a unitary operator $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and also to linear homeomorphisms $\mathcal{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ and $\mathcal{F} : \mathscr{S}'(\mathbb{R}^d) \to \mathscr{S}'(\mathbb{R}^d)$, with the latter defined by $\langle \mathcal{F}f, \varphi \rangle_{\mathscr{S}', \mathscr{S}} := \langle f, \mathcal{F}\varphi \rangle_{\mathscr{S}', \mathscr{S}}$. Here, as in the remainder of the paper, the dual pairing for distributions is taken to be *bilinear*. In any case, the inverse Fourier transform is given by (the extension of) the operator $\mathcal{F}^{-1}f(x) = \mathcal{F}f(-x)$. All of the facts listed here can be found in [30, Chapter 7].

Fix $\varphi_0 \in \mathscr{S}(\mathbb{R}^d)$ satisfying $\varphi_0(\xi) = 1$ for all $\xi \in \mathbb{R}^d$ such that $|\xi| \leq 1$, and $\varphi_0(\xi) = 0$ for all $\xi \in \mathbb{R}^d$ satisfying $|\xi| \geq 3/2$. Define $\varphi_k : \mathbb{R}^d \to \mathbb{C}, \xi \mapsto \varphi_0(2^{-k}\xi) - \varphi_0(2^{-k+1}\xi)$ for $k \in \mathbb{N}$, noting that $\sum_{j=0}^\infty \varphi_j(\xi) \equiv 1$ on $\mathbb{R}^d$.

With this, the (inhomogeneous) *Besov space* $B_{p,q}^\tau(\mathbb{R}^d)$ with smoothness $\tau \in \mathbb{R}$ and integrability exponents $p, q \in (0, \infty]$ is defined (see [35, Section 1.3, Definition 1.2]) as

$$B_{p,q}^\tau(\mathbb{R}^d) := \left\{ f \in \mathscr{S}'(\mathbb{R}^d) : \|f\|_{B_{p,q}^\tau(\mathbb{R}^d)} < \infty \right\}$$

where

$$\|f\|_{B_{p,q}^\tau(\mathbb{R}^d)} := \left\| \left(2^{j\tau} \cdot \|\mathcal{F}^{-1}(\varphi_j \cdot \widehat{f})\|_{L^p} \right)_{j \in \mathbb{N}_0} \right\|_{\ell^q} \in [0, \infty] \quad \text{for} \quad f \in \mathscr{S}'(\mathbb{R}^d).$$

This is well-defined, since $\varphi_j \cdot \widehat{f}$ is a tempered distribution with compact support, so that the Paley–Wiener theorem (see [30, Theorem 7.23]) shows that $\mathcal{F}^{-1}(\varphi_j \cdot \widehat{f})$ is a smooth function of which one can take the L^p norm (which might be infinite). One can show that the definition of $B_{p,q}^\tau(\mathbb{R}^d)$ is independent of the precise choice of the function φ_0, with equivalent quasi-norms for different choices; see [36, Proposition 1 in Section 2.3.2]. Furthermore, the spaces $B_{p,q}^\tau(\mathbb{R}^d)$ are quasi-Banach spaces that satisfy $B_{p,q}^\tau(\mathbb{R}^d) \hookrightarrow \mathscr{S}'(\mathbb{R}^d)$; see [36, Theorem in Section 2.3.3].

Now, let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be a bounded open set, and let $\tau \in \mathbb{R}$ and $p, q \in (0, \infty]$. We will use the space $\mathcal{D}'(\Omega)$ of distributions on Ω; for more details on these spaces, we refer to [30, Chapter 6]. Following [35, Definition 1.95], we then define

$$B_{p,q}^\tau(\Omega) := \left\{ f|_\Omega \ : \ f \in B_{p,q}^\tau(\mathbb{R}^d) \right\} \tag{C.1}$$

and

$$\|f\|_{B_{p,q}^{\tau}(\Omega)} := \inf \left\{ \|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \colon g \in B_{p,q}^{\tau}(\mathbb{R}^d) \text{ and } g|_{\Omega} = f \right\} \quad \text{for} \quad f \in B_{p,q}^{\tau}(\Omega). \tag{C.2}$$

Here, given a tempered distribution $f \in \mathscr{S}'(\mathbb{R}^d)$, we write $f|_{\Omega}$ for the restriction of f to Ω, given by $f|_{\Omega} \colon C_c^{\infty}(\Omega) \to \mathbb{C}, \psi \mapsto f(\psi)$. It is easy to see that $f|_{\Omega} \in \mathcal{D}'(\Omega)$. The spaces $B_{p,q}^{\tau}(\Omega)$ are quasi-Banach spaces that satisfy $B_{p,q}^{\tau}(\Omega) \hookrightarrow \mathcal{D}'(\Omega)$; see [35, Remark 1.96].

C.2 The Wavelet Characterization of Besov Spaces

Wavelets are usually constructed using a so-called *multiresolution analysis of $L^2(\mathbb{R})$*. A multiresolution analysis (see [39, Definition 2.2] or [8, Section 5.1]) of $L^2(\mathbb{R})$ is a sequence $(V_j)_{j \in \mathbb{Z}}$ of closed subspaces $V_j \subset L^2(\mathbb{R})$ with the following properties:

1. $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$;
2. $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$;
3. $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$;
4. for $f \in L^2(\mathbb{R})$, we have $f \in V_j$ if and only if $f(2^{-j} \bullet) \in V_0$;
5. there exists a function $\psi_F \in V_0$ (called the *scaling function* or the *father wavelet*) such that $(\psi_F(\bullet - m))_{m \in \mathbb{Z}}$ is an orthonormal basis of V_0.

To each multiresolution analysis, one can associate a *(mother) wavelet* $\psi_M \in L^2(\mathbb{R})$; see [39, Theorem 2.20]. More precisely, denote by $W_0 \subset L^2(\mathbb{R})$ the orthogonal complement of V_0 as a subset of V_1, and define $W_j := \{f(2^j \bullet) \colon f \in W_0\}$ for $j \in \mathbb{N}$, so that W_j is the orthogonal complement of V_j in V_{j+1}. We then have $L^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$, where the sum is orthogonal.

One can show (see [39, Lemma 2.19]) that there exists $\psi_M \in W_0$ such that the family $(\psi_M(\bullet - k))_{k \in \mathbb{Z}}$ is an orthonormal basis of W_0. In this case, we say that ψ_M is a *mother wavelet* associated with the given multiresolution analysis. For each such ψ_M, one can show (see [35, Proposition 1.51]) that if we define

$$\psi_{j,m} \colon \quad \mathbb{R} \to \mathbb{C}, \quad x \mapsto \begin{cases} \psi_F(x - m), & \text{if } j = 0 \\ 2^{\frac{j-1}{2}} \cdot \psi_M(2^{j-1} x - m), & \text{if } j \in \mathbb{N} \end{cases}$$

for $j \in \mathbb{N}_0$ and $m \in \mathbb{Z}$, then the *inhomogeneous wavelet system* $(\psi_{j,m})_{j \in \mathbb{N}_0, m \in \mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$. Furthermore, the family $\left(2^{j/2} \psi_M(2^j \bullet - k)\right)_{j,k \in \mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$.

For our purposes, we will need sufficiently regular wavelet systems, as provided by the following theorem:

Theorem 10 *For each $k \in \mathbb{N}$, there is a multiresolution analysis $(V_j)_{j \in \mathbb{Z}}$ of $L^2(\mathbb{R})$ with father/mother wavelets $\psi_F, \psi_M \in L^2(\mathbb{R})$ such that the following hold:*

1. *ψ_F, ψ_M are real-valued and have compact support;*

2. $\psi_F, \psi_M \in C^k(\mathbb{R})$;
3. $\widehat{\psi_F}(0) = (2\pi)^{-1/2}$;
4. $\int_{\mathbb{R}} x^\ell \cdot \psi_M(x)\, dx = 0$ for all $\ell \in \{0, \ldots, k\}$ (vanishing moment condition).

Proof The existence of a multi-resolution analysis $(V_j)_{j \in \mathbb{Z}}$ with compactly supported father/mother wavelets $\psi_F, \psi_M \in C^k(\mathbb{R})$ is shown in [39, Theorem 4.7] (while the original proof was given in [7]). It is not stated explicitly, however, that ψ_F, ψ_M are real-valued, but this can be extracted from the proof: The function $\Phi := \psi_F$ is constructed as $\Phi = (2\pi)^{-1/2} \mathcal{F}^{-1}\Theta$, with $\Theta(\xi) = \prod_{j=1}^\infty m(2^{-j}\xi)$ (see [39, Theorem 4.1]), where the trigonometric polynomial $m(\xi) = \sum_{k=0}^T a_k e^{ik\xi}$ is obtained through [39, Lemma 4.6], so that $a_0, \ldots, a_T \in \mathbb{R}$ and $m(0) = 1$. Therefore, [39, Lemma 4.3] shows that Φ is real-valued. Finally, $\Psi := \psi_M$ is obtained from Φ as $\Psi(x) = 2\sum_{k=0}^T \overline{a_k}(-1)^k \Phi(2x + k + 1)$; see [39, Equation (4.5)]. Since $a_0, \ldots, a_T \in \mathbb{R}$, this shows that $\psi_M = \Psi$ is real-valued as well.

The above construction also implies $\widehat{\psi_F}(0) = (2\pi)^{-1/2} \Theta(0) = (2\pi)^{-1/2}$, because of $\Theta(0) = \prod_{j=1}^\infty m(0) = 1$. Finally, the vanishing moment condition is a consequence of [39, Proposition 3.1]. □

Wavelet systems in $\mathbb{R}^d$ can be constructed by taking suitable tensor products of a one-dimensional wavelet system. To describe this, let ψ_F, ψ_M be father/mother wavelets, and let $T_0 := \{F\}^d$ and $T_j := T := \{F, M\}^d \setminus T_0$ for $j \in \mathbb{N}$. Now, for $t = (t_1, \ldots, t_d) \in T$ and $m = (m_1, \ldots, m_d) \in \mathbb{Z}^d$, define $\Psi_m : \mathbb{R}^d \to \mathbb{C}$ and $\Psi_{t,m} : \mathbb{R}^d \to \mathbb{C}$ by

$$\Psi_m(x) := \prod_{j=1}^d \psi_F(x_j - m_j) \quad \text{and} \quad \Psi_{t,m}(x) := \prod_{j=1}^d \psi_{t_j}(x_j - m_j). \quad \text{(C.3)}$$

Finally, set $J := \{(j, t, m) : j \in \mathbb{N}_0, t \in T_j, m \in \mathbb{Z}^d\}$, and

$$\Psi_{j,t,m} : \ \mathbb{R}^d \to \mathbb{C}, \ x \mapsto \begin{cases} \Psi_m(x) & \text{if } j = 0, \ t \in T_0, \ \text{and } m \in \mathbb{Z}^d \\ \\ 2^{(j-1)d/2}\, \Psi_{t,m}(2^{j-1}x) & \text{if } j \in \mathbb{N}, \ t \in T_j, \ \text{and } m \in \mathbb{Z}^d \end{cases} \quad \text{(C.4)}$$

Then (see [35, Proposition 1.53]), the system $(\Psi_{j,t,m})_{(j,t,m)\in J}$ is an orthonormal basis of $L^2(\mathbb{R}^d)$.

Finally, we have the following wavelet characterization of the Besov spaces $B_{p,q}^\tau(\mathbb{R}^d)$.

Theorem 11 (consequence of [35, Theorem 1.64])

Let $d \in \mathbb{N}$, $p, q \in (0, \infty]$ and $\tau \in \mathbb{R}$. For a sequence $\mathbf{c} = (c_{j,t,m})_{(j,t,m)\in J} \in \mathbb{C}^J$ define

$$\|\mathbf{c}\|_{b_{p,q}^\tau} := \|\mathbf{c}\|_{b_{p,q}^\tau(\mathbb{R}^d)} := \left\| \left(2^{j(\tau + d \cdot (\frac{1}{2} - \frac{1}{p}))} \cdot \|(c_{j,t,m})_{m \in \mathbb{Z}^d}\|_{\ell^p} \right)_{j \in \mathbb{N}_0, t \in T_j} \right\|_{\ell^q} \in [0, \infty],$$

and $b_{p,q}^{\tau}(\mathbb{R}^d) := \{ \mathbf{c} \in \mathbb{C}^J : \|\mathbf{c}\|_{b_{p,q}^{\tau}(\mathbb{R}^d)} < \infty \}$.

Let $k \in \mathbb{N}$, and let ψ_F, ψ_M as provided by Theorem 10, for this choice of k. Let the d-dimensional wavelet system $(\Psi_{j,t,m})_{(j,t,m) \in J}$ be as defined in Eq. (C.4). Finally, suppose that $k > \max\{\tau, \frac{2d}{p} + \frac{d}{2} - \tau\}$. Then, the map

$$\Gamma = \Gamma_k : \quad b_{p,q}^{\tau}(\mathbb{R}^d) \to B_{p,q}^{\tau}(\mathbb{R}^d), \quad (c_{j,t,m})_{(j,t,m) \in J} \mapsto \sum_{(j,t,m) \in J} c_{j,t,m} \, \Psi_{j,t,m}$$

is well-defined (with unconditional convergence of the series in $\mathscr{S}'(\mathbb{R}^d)$), and an isomorphism of (quasi)-Banach spaces. The inverse map of Γ will be denoted by

$$\Theta = \Theta_k := \Gamma_k^{-1} : \quad B_{p,q}^{\tau}(\mathbb{R}^d) \to b_{p,q}^{\tau}(\mathbb{R}^d), \quad f \mapsto \big(\theta_{j,t,m}(f)\big)_{(j,t,m) \in J}.$$

We will also use the *real-valued Besov space*

$$B_{p,q}^{\tau}(\mathbb{R}^d; \mathbb{R}) := B_{p,q}^{\tau}(\mathbb{R}^d) \cap \mathscr{S}'(\mathbb{R}^d; \mathbb{R}) \quad \text{equipped with the (quasi)-norm} \quad \| \cdot \|_{B_{p,q}^{\tau}(\mathbb{R}^d)},$$

where we write $\mathscr{S}'(\mathbb{R}^d; \mathbb{R}) := \{ \varphi \in \mathscr{S}'(\mathbb{R}^d) : \forall f \in \mathscr{S}(\mathbb{R}^d; \mathbb{R}) : \langle \varphi, f \rangle_{\mathscr{S}', \mathscr{S}} \in \mathbb{R} \}$ and $\mathscr{S}(\mathbb{R}^d; \mathbb{R}) := \{ f : \mathbb{R}^d \to \mathbb{R} : f \in \mathscr{S}(\mathbb{R}^d) \}$. The spaces $B_{p,q}^{\tau}(\Omega; \mathbb{R})$ are defined similarly. We will also use the space $b_{p,q}^{\tau}(\mathbb{R}^d; \mathbb{R}) := b_{p,q}^{\tau}(\mathbb{R}^d) \cap \mathbb{R}^J$.

C.3 Wavelets and Besov Spaces on Bounded Domains

Note that Theorem 11 only pertains to the Besov spaces $B_{p,q}^{\tau}(\mathbb{R}^d)$. To describe Besov spaces on domains, we will use the sequence spaces $b_{p,q}^{\tau}(\Omega_{\text{int}}; \mathbb{R})$ and $b_{p,q}^{\tau}(\Omega_{\text{ext}}; \mathbb{R})$ that we now define.

Definition 9 Let $p, q \in (0, \infty]$ and $\tau \in \mathbb{R}$, and let $k \in \mathbb{N}$ with $k > \max\{\tau, \frac{2d}{p} + \frac{d}{2} - \tau\}$. Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be a bounded open set. With the father/mother wavelets ψ_F, ψ_M as in Theorem 10 (with the above choice of k) and $\Psi_{j,t,m}$ as in Eq. (C.4), define

$$J^{\text{ext}} := \bigcup_{j \in \mathbb{N}_0} \big(\{j\} \times J_j^{\text{ext}} \big) \text{ where } J_j^{\text{ext}} := \{(t, m) \in T_j \times \mathbb{Z}^d : \Omega \cap \operatorname{supp} \Psi_{j,t,m} \neq \emptyset\},$$

$$\text{and} \quad J^{\text{int}} := \bigcup_{j \in \mathbb{N}_0} \big(\{j\} \times J_j^{\text{int}} \big) \text{ where } J_j^{\text{int}} := \{(t, m) \in T_j \times \mathbb{Z}^d : \operatorname{supp} \Psi_{j,t,m} \subset \Omega\}.$$

Finally, set

$$b_{p,q}^{\tau}(\Omega_{\text{ext}}; \mathbb{R}) := \big\{ (c_{j,t,m})_{(j,t,m) \in J} \in b_{p,q}^{\tau}(\mathbb{R}^d; \mathbb{R}) : c_{j,t,m} = 0 \quad \forall (j, t, m) \in J \setminus J^{\text{ext}} \big\},$$

and define $b_{p,q}^{\tau}(\Omega_{\text{int}}; \mathbb{R})$ similarly. Both of these spaces are considered as subspaces of $b_{p,q}^{\tau}(\mathbb{R}^d; \mathbb{R})$; they are thus equipped with the (quasi)-norm $\| \cdot \|_{b_{p,q}^{\tau}}$.

Remark Strictly speaking, the spaces $b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R})$ and $b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R})$ (and the index sets J^{ext} and J^{int}) depend on the choice of $k \in \mathbb{N}$ and on the precise choice of ψ_F, ψ_M. We will, however, suppress this dependence. ◄

The next lemma describes the relation between these sequence spaces and the Besov spaces $B_{p,q}^\tau(\Omega; \mathbb{R})$.

Lemma 11 *Let $d \in \mathbb{N}$, $\emptyset \neq \Omega \subset \mathbb{R}^d$ open and bounded, $p, q \in (0, \infty]$ and $\tau \in \mathbb{R}$. Let $k \in \mathbb{N}$ with $k > \max\{\tau, \frac{2d}{p} + \frac{d}{2} - \tau\}$. Let $b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R})$ and $b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R})$ be as in Definition 9 (for the given choice of k, ψ_F, ψ_M), and J as defined before Eq. (C.4).*
Then, there exist continuous linear maps

$$T_{\mathrm{int}} : b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R}) \to B_{p,q}^\tau(\Omega; \mathbb{R}) \quad and \quad T_{\mathrm{ext}} : b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R}) \to B_{p,q}^\tau(\Omega; \mathbb{R})$$

with the following properties:

- *There is $\gamma > 0$ such that $\|T_{\mathrm{int}}\mathbf{c}\|_{L^2(\Omega)} = \gamma \cdot \|\mathbf{c}\|_{\ell^2}$ for all $\mathbf{c} \in \ell^2(J) \cap b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R})$, and $\|T_{\mathrm{int}}\mathbf{c}\|_{B_{p,q}^\tau(\Omega)} \leq \|\mathbf{c}\|_{b_{p,q}^\tau}$ for all $\mathbf{c} \in b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R})$.*
- *There exists $\varrho > 0$ such that $\|T_{\mathrm{ext}}\mathbf{c}\|_{L^2(\Omega)} \leq \varrho \cdot \|\mathbf{c}\|_{\ell^2}$ for all $\mathbf{c} \in b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R})$, and we have*

$$\mathrm{Ball}\left(0, 1; B_{p,q}^\tau(\Omega; \mathbb{R})\right) \subset T_{\mathrm{ext}}\left(\mathrm{Ball}\left(0, 1; b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R})\right)\right). \tag{C.5}$$

Proof With the operator Γ as in Theorem 11, let $\gamma := \left(1 + \|\Gamma\|_{b_{p,q}^\tau(\mathbb{R}^d) \to B_{p,q}^\tau(\mathbb{R}^d)}\right)^{-1}$, and define

$$T_{\mathrm{int}} : \quad b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R}) \to B_{p,q}^\tau(\Omega; \mathbb{R}), \quad \mathbf{c} \mapsto \gamma \cdot (\Gamma \mathbf{c})|_\Omega.$$

By definition of the Besov space $B_{p,q}^\tau(\Omega; \mathbb{R})$ and its norm (see Eq. (C.2)), we then see that T_{int} is a well-defined continuous linear map, with

$$\|T_{\mathrm{int}}\, \mathbf{c}\|_{B_{p,q}^\tau(\Omega)} \leq \gamma \cdot \|\Gamma \mathbf{c}\|_{B_{p,q}^\tau(\mathbb{R}^d)} \leq \|\mathbf{c}\|_{b_{p,q}^\tau} \qquad \forall\, \mathbf{c} \in b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R}).$$

Next, let $\mathbf{c} = (c_{j,t,m})_{(j,t,m) \in J} \in \ell^2(J) \cap b_{p,q}^\tau(\Omega_{\mathrm{int}}; \mathbb{R})$ be arbitrary. By orthonormality of the family $(\Psi_{j,t,m})_{(j,t,m) \in J} \subset L^2(\mathbb{R}^d)$, and since $c_{j,t,m} = 0$ for $(j, t, m) \in J \setminus J^{\mathrm{int}}$, while $\mathrm{supp}\, \Psi_{j,t,m} \subset \Omega$ for $(j, t, m) \in J^{\mathrm{int}}$, we see

$$\|T_{\mathrm{int}}\, \mathbf{c}\|_{L^2(\Omega)} = \gamma \cdot \left\|\sum_{(j,t,m) \in J^{\mathrm{int}}} c_{j,t,m}\, \Psi_{j,t,m}\right\|_{L^2(\Omega)} = \gamma \cdot \left\|\sum_{(j,t,m) \in J^{\mathrm{int}}} c_{j,t,m}\, \Psi_{j,t,m}\right\|_{L^2(\mathbb{R}^d)}$$

$$= \gamma \cdot \|(c_{j,t,m})_{(j,t,m) \in J^{\mathrm{int}}}\|_{\ell^2} = \gamma \cdot \|\mathbf{c}\|_{\ell^2}.$$

To construct T_{ext}, let Θ be as in Theorem 11, set $\varrho := 2 \cdot \left(1 + \|\Theta\|_{B_{p,q}^\tau(\mathbb{R}^d) \to b_{p,q}^\tau(\mathbb{R}^d)}\right)$, and define

$$T_{\mathrm{ext}} : \quad b_{p,q}^\tau(\Omega_{\mathrm{ext}}; \mathbb{R}) \to B_{p,q}^\tau(\Omega; \mathbb{R}), \quad \mathbf{c} \mapsto \varrho \cdot (\Gamma \mathbf{c})|_\Omega.$$

Exactly as for T_{int}, we see that T_{ext} is a well-defined continuous linear map. Furthermore, using again that the family $(\Psi_{j,t,m})_{(j,t,m)\in J} \subset L^2(\mathbb{R}^d)$ is an orthonormal system, we see

$$\|T_{\text{ext}}\,\mathbf{c}\|_{L^2(\Omega)} \le \varrho \cdot \|\Gamma\,\mathbf{c}\|_{L^2(\mathbb{R}^d)} \le \varrho \cdot \|\mathbf{c}\|_{\ell^2} \quad \forall \mathbf{c} \in b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R}).$$

It remains to prove the inclusion (C.5). To this end, let $f \in \text{Ball}\left(0, 1; B_{p,q}^\tau(\Omega; \mathbb{R})\right)$ be arbitrary. By definition, this implies $f = g|_\Omega$ for some $g \in B_{p,q}^\tau(\mathbb{R}^d)$ with $\|g\|_{B_{p,q}^\tau} \le 2$. Let $\mathbf{e} := \Theta g \in b_{p,q}^\tau(\mathbb{R}^d)$, and $\mathbf{c} = (c_{j,t,m})_{(j,t,m)\in J}$ where $c_{j,t,m} := \varrho^{-1} \cdot \mathbb{1}_{J^{\text{ext}}}((j,t,m)) \cdot \text{Re}(e_{j,t,m})$. Clearly, $\|\mathbf{c}\|_{b_{p,q}^\tau} \le \varrho^{-1}\|\mathbf{e}\|_{b_{p,q}^\tau} \le 2\|\Theta\|/\varrho \le 1$, which means $\mathbf{c} \in \text{Ball}\left(0, 1; b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R})\right)$.

Finally, for an arbitrary *real-valued* test function $\varphi \in C_c^\infty(\Omega)$, we have $\langle \Psi_{j,t,m}, \varphi \rangle \in \mathbb{R}$ for all $(j,t,m) \in J$ and $\langle \Psi_{j,t,m}, \varphi \rangle = 0$ if $(j,t,m) \notin J^{\text{ext}}$. Therefore,

$$\langle T_{\text{ext}}\,\mathbf{c}, \varphi \rangle = \text{Re} \sum_{(j,t,m)\in J^{\text{ext}}} \left(e_{j,t,m}\,\langle \Psi_{j,t,m}, \varphi \rangle\right) = \text{Re}\left\langle \sum_{(j,t,m)\in J} e_{j,t,m}\,\Psi_{j,t,m},\ \varphi \right\rangle$$

$$= \text{Re}\langle \Gamma\,\mathbf{e}, \varphi \rangle = \text{Re}\langle \Gamma(\Theta g), \varphi \rangle = \text{Re}\langle g, \varphi \rangle = \text{Re}\langle f, \varphi \rangle = \langle f, \varphi \rangle,$$

since $\Theta = \Gamma^{-1}$ and $f = g|_\Omega$ and $\varphi \in C_c^\infty(\Omega)$, and since f is a real-valued distribution. Therefore, $f = T_{\text{ext}}\,\mathbf{c}$, proving (C.5). □

Finally, we show that the sequence spaces $b_{p,q}^\tau(\Omega_{\text{int}}; \mathbb{R})$ and $b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R})$ are quite similar to the sequence spaces $\ell_{\mathscr{P},\alpha}^{p,q}$ introduced in Definition 5. In fact, the following (seemingly) weak property will be enough for our purposes.

Lemma 12 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be open and bounded. Let $p, q \in (0, \infty]$ and $\tau \in \mathbb{R}$, and define $\alpha := \tau + d \cdot (\frac{1}{2} - \frac{1}{p})$. Let $k \in \mathbb{N}$ with $k > \max\{\tau, \frac{2d}{p} + \frac{d}{2} - \tau\}$, and let $b_{p,q}^\tau(\Omega_{\text{int}}; \mathbb{R})$ and $b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R})$ be as in Definition 9 (for the given choice of k, ψ_F, ψ_M).*

Assume that $\alpha > d \cdot (\frac{1}{2} - \frac{1}{p})_+$. Then, the embeddings $b_{p,q}^\tau(\Omega_{\text{int}}; \mathbb{R}) \hookrightarrow \ell^2(J^{\text{int}})$ and $b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R}) \hookrightarrow \ell^2(J^{\text{ext}})$ hold. Furthermore,

(i) *There is a d-regular partition $\mathscr{P}^{\text{int}}$ of J^{int} and some $\gamma > 0$ such that if we define*

$$\iota_{\text{int}} : \quad \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q} \to b_{p,q}^\tau(\Omega_{\text{int}}; \mathbb{R}), \quad \mathbf{c} \mapsto \gamma \cdot \mathbf{c}^\natural$$

where $\mathbf{c}^\natural \in \mathbb{R}^J$ is obtained by extending $\mathbf{c} \in \mathbb{R}^{J^{\text{int}}}$ by zero, then $\|\iota_{\text{int}}\| \le 1$ and $\|\iota_{\text{int}}\,\mathbf{c}\|_{\ell^2} = \gamma\,\|\mathbf{c}\|_{\ell^2}$ for all $\mathbf{c} \in \ell_{\mathscr{P}^{\text{int}},\alpha}^{p,q}$.

(ii) *There is a d-regular partition $\mathscr{P}^{\text{ext}}$ of J^{ext} and some $\varrho > 0$ such that if we define*

$$\iota_{\text{ext}} : \quad \ell_{\mathscr{P}^{\text{ext}},\alpha}^{p,q} \to b_{p,q}^\tau(\Omega_{\text{ext}}; \mathbb{R}), \quad \mathbf{c} \mapsto \varrho \cdot \mathbf{c}^\natural$$

where $\mathbf{c}^\natural \in \mathbb{R}^J$ is obtained by extending $\mathbf{c} \in \mathbb{R}^{J^{\text{ext}}}$ by zero, then $\|\iota_{\text{ext}}\,\mathbf{c}\|_{\ell^2} = \varrho\,\|\mathbf{c}\|_{\ell^2}$ for all $\mathbf{c} \in \ell^{p,q}_{\mathscr{P}_{\text{ext}},\alpha}$, and

$$\text{Ball}\left(0,\, 1;\, b^\tau_{p,q}(\Omega_{\text{ext}};\,\mathbb{R})\right) \subset \iota_{\text{ext}}\left(\text{Ball}(0,\,1;\,\ell^{p,q}_{\mathscr{P}_{\text{ext}},\alpha})\right).$$

Proof The proof is divided into three steps.

Step 1 (*Estimating* $|J^{\text{int}}_j|$ *and* $|J^{\text{ext}}_j|$): We show that there are $j_0 \in \mathbb{N}$ and $a, A > 0$ satisfying

$$|J^{\text{int}}_j| \le |J^{\text{ext}}_j| \le A \cdot 2^{dj} \quad \forall\, j \in \mathbb{N} \quad \text{and} \quad |J^{\text{ext}}_j| \ge |J^{\text{int}}_j| \ge a \cdot 2^{dj} \quad \forall\, j \in \mathbb{N}_{\ge j_0}.$$

First of all, we clearly have $J^{\text{int}}_j \subset J^{\text{ext}}_j$ and thus $|J^{\text{int}}_j| \le |J^{\text{ext}}_j|$. Next, since $\Omega \subset \mathbb{R}^d$ is bounded and ψ_F, ψ_M have compact support, there is $R \in \mathbb{N}$ such that $\Omega \subset [-R, R]^d$ and $\text{supp}\,\psi_F \cup \text{supp}\,\psi_M \subset [-R, R]$. Define $A := (8R)^d$. In view of Eqs. (C.3) and (C.4), this implies $\text{supp}\,\Psi_m \cup \text{supp}\,\Psi_{t,m} \subset m + [-R, R]^d$, and hence $\text{supp}\,\Psi_{0,t,m} \subset m + [-R, R]^d$ for $t \in T_0$ and $m \in \mathbb{Z}^d$, and finally $\text{supp}\,\Psi_{j,t,m} \subset 2^{1-j}(m + [-R, R]^d)$ for $j \in \mathbb{N}, t \in T_j$, and $m \in \mathbb{Z}^d$.

Now, it is not hard to see that if $\emptyset \ne \Omega \cap \text{supp}\,\Psi_{0,t,m} \subset [-R, R]^d \cap (m + [-R, R]^d)$, then $m \in [-2R, 2R]^d \cap \mathbb{Z}^d = \{-2R, \dots, 2R\}^d$, and thus $|J^{\text{ext}}_0| \le (1 + 4R)^d \le A \cdot 2^{d \cdot 0}$.

Furthermore, if $j \in \mathbb{N}$ and $\emptyset \ne \Omega \cap \text{supp}\,\Psi_{j,t,m} \subset [-R, R]^d \cap 2^{1-j}(m + [-R, R]^d)$, then $2^{1-j}(m + x) = y$ for certain $x, y \in [-R, R]^d$, and hence

$$m = 2^{j-1}y - x \in \mathbb{Z}^d \cap [-(R + 2^{j-1}R),\, R + 2^{j-1}R]^d \subset \{-2^j R, \dots, 2^j R\}^d.$$

Because of $|T_j| \le 2^d$, this implies $|J^{\text{ext}}_j| \le |T_j| \cdot (1 + 2^{j+1}R)^d \le (8R)^d 2^{jd} \le A\,2^{dj}$.

Regarding the lower bound, recall that $\Omega \ne \emptyset$ is open, so that there are $x_0 \in \mathbb{R}^d$ and $n \in \mathbb{N}$ satisfying $x_0 + [-r, r]^d \subset \Omega$, where $r := 2^{-n}$. Choose $j_0 \in \mathbb{N}_{\ge n+3}$ such that $2^{j_0-1}r \ge 2R$, and note $2^{j_0-3}r = 2^{j_0-3-n} \in \mathbb{N}$. Let $j \ge j_0$. Choose $m_0 := \lfloor 2^{j-1}x_0 \rfloor \in \mathbb{Z}^d$, with the "floor" operation applied componentwise. We have $\|2^{j-1}x_0 - m_0\|_\infty \le 1$, and hence

$$\|2^{j-1}x_0 - (m + m_0)\|_\infty \le 1 + 2^{j-3}r \le 2^{j-2}r \quad \text{for} \quad m \in \{-2^{j-3}r, \dots, 2^{j-3}r\}^d.$$

Here, one should observe $2^{j-3}r = 2^{j-j_0}2^{j_0-3-n} \in \mathbb{N}$. Because of $R \le 2^{j_0-2}r \le 2^{j-2}r$, the above estimate implies that

$$2^{1-j} \cdot (m + m_0 + [-R, R]^d) \subset 2^{1-j} \cdot (2^{j-1}x_0 + [-(R + 2^{j-2}r), (R + 2^{j-2}r)]^d)$$
$$\subset 2^{1-j} \cdot (2^{j-1}x_0 + [-2^{j-1}r, 2^{j-1}r]^d) = x_0 + [-r, r]^d \subset \Omega$$

for all $m \in \{-2^{j-3}r, \dots, 2^{j-3}r\}^d$. Because of $\text{supp}\,\Psi_{j,t,m+m_0} \subset 2^{1-j}(m + m_0 + [-R, R]^d)$, this implies $|J^{\text{int}}_j| \ge (2^{j-2}r)^d = (r/4)^d \cdot 2^{dj}$, so that we can choose $a = (r/4)^d$.

Step 2 *(Constructing the partitions $\mathscr{P}^{\mathrm{int}}$, $\mathscr{P}^{\mathrm{ext}}$ and showing $\|c^\natural\|_{b_{p,q}^\tau} \asymp \|c\|_{\ell_{\mathscr{P},\alpha}^{p,q}}$):*
Define $\mathcal{I}_1^{\mathrm{int}} := \bigcup_{j=0}^{j_0}(\{j\} \times J_j^{\mathrm{int}})$ and $\mathcal{I}_1^{\mathrm{ext}} := \bigcup_{j=0}^{j_0}(\{j\} \times J_j^{\mathrm{ext}})$, as well as

$$\mathcal{I}_m^{\mathrm{int}} := \{j_0 + m - 1\} \times J_{j_0+m-1}^{\mathrm{int}} \quad \text{and} \quad \mathcal{I}_m^{\mathrm{ext}} := \{j_0 + m - 1\} \times J_{j_0+m-1}^{\mathrm{ext}}$$

for $m \in \mathbb{N}_{\geq 2}$. As shown in Step 1, we have for $m \in \mathbb{N}_{\geq 2}$ that

$$a \cdot 2^{dm} \leq a \cdot 2^{d(j_0+m-1)} \leq |\mathcal{I}_m^{\mathrm{int}}| \leq |\mathcal{I}_m^{\mathrm{ext}}| \leq A \cdot 2^{d(j_0+m-1)} =: A' \cdot 2^{dm}$$

and also $|\mathcal{I}_1^{\mathrm{ext}}| \geq |\mathcal{I}_1^{\mathrm{int}}| \geq |J_{j_0}^{\mathrm{int}}| \geq a \cdot 2^{dj_0} \geq$
$a \cdot 2^d$. Thus, $a \cdot 2^{dm} \leq |\mathcal{I}_m^{\mathrm{int}}| \leq |\mathcal{I}_m^{\mathrm{ext}}| \leq A'' \cdot 2^{dm}$ for all $m \in \mathbb{N}$, where $A'' := \max\{A', |\mathcal{I}_1^{\mathrm{ext}}|\}$. Furthermore, we have $J^{\mathrm{int}} = \biguplus_{m \in \mathbb{N}} \mathcal{I}_m^{\mathrm{int}}$ and $J^{\mathrm{ext}} = \biguplus_{m \in \mathbb{N}} \mathcal{I}_m^{\mathrm{ext}}$, so that $\mathscr{P}^{\mathrm{int}} := \left(\mathcal{I}_m^{\mathrm{int}}\right)_{m \in \mathbb{N}}$ and $\mathscr{P}^{\mathrm{ext}} := \left(\mathcal{I}_m^{\mathrm{ext}}\right)_{m \in \mathbb{N}}$ are d-regular partitions of J^{int} and J^{ext}, respectively.

Now, for $J_0 \subset J$ and $\mathbf{c} \in \mathbb{R}^{J_0}$, let $\mathbf{c}^\natural \in \mathbb{R}^J$ be the sequence $\mathbf{c}$, extended by zero. We claim that there are $C_1, C_2 > 0$ such that

$$C_1 \cdot \|\mathbf{c}\|_{\ell_{\mathscr{P}^{\mathrm{int}},\alpha}^{p,q}} \leq \|\mathbf{c}^\natural\|_{b_{p,q}^\tau} \leq C_2 \cdot \|\mathbf{c}\|_{\ell_{\mathscr{P}^{\mathrm{int}},\alpha}^{p,q}} \qquad \forall \mathbf{c} \in \mathbb{R}^{J^{\mathrm{int}}}, \tag{C.6}$$

and similarly for $\mathscr{P}^{\mathrm{ext}}$ and J^{ext} instead of $\mathscr{P}^{\mathrm{int}}$ and J^{int}. For brevity, we only prove the claim for $\mathscr{P}^{\mathrm{int}}$.

To prove (C.6), let $\mathbf{c} \in \mathbb{R}^{J^{\mathrm{int}}}$. For $m \in \mathbb{N}$ and $j \in \mathbb{N}_0$, define $\zeta_m := 2^{\alpha m} \|(c_\kappa)_{\kappa \in \mathcal{I}_m^{\mathrm{int}}}\|_{\ell^p}$ and $\omega_j := 2^{\alpha j} \left\| \left(\|(c_{j,t,k}^\natural)_{k \in \mathbb{Z}^d}\|_{\ell^p}\right)_{t \in T_j} \right\|_{\ell^q}$, noting that $\|\mathbf{c}\|_{\ell_{\mathscr{P}^{\mathrm{int}},\alpha}^{p,q}} = \|(\zeta_m)_{m \in \mathbb{N}}\|_{\ell^q}$ as well as $\|\mathbf{c}^\natural\|_{b_{p,q}^\tau} = \|(\omega_j)_{j \in \mathbb{N}_0}\|_{\ell^q}$. Since $|T_j| \leq 2^d$ for all $j \in \mathbb{N}_0$, we have $\|\cdot\|_{\ell^p(T_j)} \asymp \|\cdot\|_{\ell^q(T_j)}$ for all $j \in \mathbb{N}_0$, with implied constant only depending on d, p, q.

Now, define $J_{j,t}^{\mathrm{int}} := \{k \in \mathbb{Z}^d : (t, k) \in J_j^{\mathrm{int}}\}$ for $j \in \mathbb{N}_0$ and $t \in T_j$, and note for $m \geq 2$ that $\mathcal{I}_m^{\mathrm{int}} = \biguplus_{t \in T_{j_0+m-1}} \left(\{j_0 + m - 1\} \times \{t\} \times J_{j_0+m-1,t}^{\mathrm{int}}\right)$, which implies

$$\begin{aligned}
\zeta_m &= 2^{\alpha m} \left\| \left(\|(c_{j_0+m-1,t,k})_{k \in J_{j_0+m-1,t}^{\mathrm{int}}}\|_{\ell^p} \right)_{t \in T_{j_0+m-1}} \right\|_{\ell^p} \\
&\asymp 2^{\alpha(j_0+m-1)} \left\| \left(\|(c_{j_0+m-1,t,k}^\natural)_{k \in \mathbb{Z}^d}\|_{\ell^p} \right)_{t \in T_{j_0+m-1}} \right\|_{\ell^q} = \omega_{j_0+m-1},
\end{aligned}$$

with implied constants only depending on d, p, q, j_0, α. With similar arguments, we see that

$$\begin{aligned}
\zeta_1 &= 2^\alpha \left\| \left(\|(c_{j,t,k})_{k \in J_{j,t}^{\mathrm{int}}}\|_{\ell^p} \right)_{j \in \{0,\ldots,j_0\}, t \in T_j} \right\|_{\ell^p} \\
&\asymp \left\| \left(\left(2^{\alpha j} \|(c_{j,t,k}^\natural)_{k \in \mathbb{Z}^d}\|_{\ell^p}\right)_{t \in T_j} \right\|_{\ell^q} \right)_{j \in \{0,\ldots,j_0\}} \right\|_{\ell^q} = \left\| (\omega_j)_{j \in \{0,\ldots,j_0\}} \right\|_{\ell^q}.
\end{aligned}$$

Overall, we obtain that

$$\|\mathbf{c}\|_{\ell^{p,q}_{\mathscr{P}^{\text{int}},\alpha}} = \big\|(\zeta_m)_{m\in\mathbb{N}}\big\|_{\ell^q} \asymp \zeta_1 + \big\|(\zeta_m)_{m\in\mathbb{N}_{\geq 2}}\big\|_{\ell^q}$$

$$\asymp \big\|(\omega_j)_{j\in\{0,\dots,j_0\}}\big\|_{\ell^q} + \big\|(\omega_{m+j_0-1})_{m\in\mathbb{N}_{\geq 2}}\big\|_{\ell^q} \asymp \big\|(\omega_j)_{j\in\mathbb{N}_0}\big\|_{\ell^q} = \|\mathbf{c}^\natural\|_{b^\tau_{p,q}},$$

which proves Eq. (C.6).

Step 3 *(Completing the proof):* Step 2 guarantees the existence of $\gamma > 0$ satisfying $\|\iota_{\text{int}}\,\mathbf{c}\|_{b^\tau_{p,q}} = \gamma \cdot \|\mathbf{c}^\natural\|_{b^\tau_{p,q}} \leq \|\mathbf{c}\|_{\ell^{p,q}_{\mathscr{P}^{\text{int}},\alpha}}$ for all $\mathbf{c} \in \ell^{p,q}_{\mathscr{P}^{\text{int}},\alpha}$. Furthermore, we clearly have $\|\iota_{\text{int}}\,\mathbf{c}\|_{\ell^2} = \gamma\,\|\mathbf{c}\|_{\ell^2}$.

Similarly, Step 2 shows that there is $\varrho > 0$ satisfying $\|\mathbf{c}\|_{\ell^{p,q}_{\mathscr{P}^{\text{ext}},\alpha}} \leq \varrho\,\|\mathbf{c}^\natural\|_{b^\tau_{p,q}}$ for all $\mathbf{c} \in \mathbb{R}^{J^{\text{ext}}}$. Now, given $\mathbf{b} \in \text{Ball}\big(0, 1; b^\tau_{p,q}(\Omega_{\text{ext}}; \mathbb{R})\big)$, note that $\mathbf{b} = (\mathbf{b}|_{J^{\text{ext}}})^\natural$ and furthermore $\|\mathbf{b}|_{J^{\text{ext}}}\|_{\ell^{p,q}_{\mathscr{P}^{\text{ext}},\alpha}} \leq \varrho\,\|(\mathbf{b}|_{J^{\text{ext}}})^\natural\|_{b^\tau_{p,q}} \leq \varrho$, so that $\mathbf{c} := \varrho^{-1} \cdot \mathbf{b}|_{J^{\text{ext}}} \in \text{Ball}\big(0, 1; \ell^{p,q}_{\mathscr{P}^{\text{ext}},\alpha}\big)$ satisfies $\mathbf{b} = \iota_{\text{ext}}\mathbf{c}$. It is clear that $\|\iota_{\text{ext}}\mathbf{c}\|_{\ell^2} = \varrho\,\|\mathbf{c}\|_{\ell^2}$ for all $\mathbf{c} \in \ell^{p,q}_{\mathscr{P}^{\text{ext}},\alpha}$.

Finally, Proposition 2 shows that $\ell^{p,q}_{\mathscr{P}^{\text{int}},\alpha} \hookrightarrow \ell^2(J^{\text{int}})$ and $\ell^{p,q}_{\mathscr{P}^{\text{ext}},\alpha} \hookrightarrow \ell^2(J^{\text{ext}})$. Hence, we see for $\mathbf{c} \in b^\tau_{p,q}(\Omega_{\text{int}}; \mathbb{R})$ that $\|\mathbf{c}\|_{\ell^2} \asymp \|\iota_{\text{int}}\mathbf{c}\|_{\ell^2} \lesssim \|\iota_{\text{int}}\mathbf{c}\|_{\ell^{p,q}_{\mathscr{P}^{\text{int}},\alpha}} \lesssim \|c\|_{b^\tau_{p,q}}$ and hence $b^\tau_{p,q}(\Omega_{\text{int}}; \mathbb{R}) \hookrightarrow \ell^2(J^{\text{int}})$. Similarly, one can show that $b^\tau_{p,q}(\Omega_{\text{ext}}; \mathbb{R}) \hookrightarrow \ell^2(J^{\text{ext}})$. $\qquad\square$

D The Phase Transition for Sobolev Spaces with $p \in \{1, \infty\}$

In this subsection, we provide the missing proof of Theorem 8 for the cases $p = 1$ and $p = \infty$. We begin with the case $p = 1$.

D.1 The case $p = 1$

The proof is crucially based on the following embedding.

Lemma 13 *For arbitrary $k, d \in \mathbb{N}$ and $1 \leq p < \infty$, we have $W^{k,p}(\mathbb{R}^d) \hookrightarrow B^k_{p,\infty}(\mathbb{R}^d)$.*

Proof This follows from [1, Section 7.33]. Here, the definition of Besov spaces used in [1] coincides with our definition, as can be seen by combining [36, Theorem in Section 2.5.12] with [19, Proposition 17.21 and Theorem 17.24]. $\qquad\square$

Using this embedding, we can now prove Theorem 8 for the case $p = 1$

Proof of Theorem 8 for $p = 1$ Let $k \in \mathbb{N}$ with $k > d \cdot (\frac{1}{1} - \frac{1}{2})_+ = \frac{d}{2}$ and define $\mathcal{S} := \text{Ball}(0, 1; W^{k,1}(\Omega))$. Our goal is to apply Theorem 5 for $\mathbf{X} := \mathbf{Y} := \mathbf{Z} := L^2(\Omega)$, $\mathcal{S}_{\mathbf{X}} := \text{Ball}(0, 1; B^k_{1,\infty}(\Omega))$, and $\mathcal{S}_{\mathbf{Y}} := \text{Ball}(0, 1; W^{k,2}(\Omega))$, with suitable choices of $\Phi, \Psi, \mathbb{P}$.

To this end, first note that since $\Omega \subset \mathbb{R}^d$ is bounded, there exists $\kappa > 0$ satisfying $\kappa \|f\|_{W^{k,1}(\Omega)} \leq \|f\|_{W^{k,2}(\Omega)}$ for all $f \in W^{k,2}(\Omega)$. Next, Theorems 7 and 8 (the latter for $p = 2 \in (1, \infty)$) show that $\mathcal{S}_{\mathbf{X}}, \mathcal{S}_{\mathbf{Y}} \subset L^2(\Omega)$ are bounded with

$$s^*_{L^2(\Omega)}(\mathcal{S}_{\mathbf{X}}) = s^*_{L^2(\Omega)}(\mathcal{S}_{\mathbf{Y}}) = \frac{k}{d}$$

and that there exists a Borel measure $\mathbb{P}_0$ on $\mathcal{S}_{\mathbf{Y}}$ that is critical for $\mathcal{S}_{\mathbf{Y}}$ with respect to $L^2(\Omega)$.

Next, we claim that there exists $C > 0$ satisfying $\|f\|_{B^k_{1,\infty}(\Omega)} \leq C \|f\|_{W^{k,1}(\Omega)}$ for all $f \in W^{k,1}(\Omega)$. Indeed, [33, Theorem 5 in Chapter VI] shows that there exists a bounded linear *extension operator* $\mathscr{E} : W^{k,1}(\Omega) \to W^{k,1}(\mathbb{R}^d)$ satisfying $(\mathscr{E}f)|_\Omega = f$ for all $f \in W^{k,1}(\Omega)$. Then, Lemma 13 yields $C_1 > 0$ satisfying

$$\|f\|_{B^k_{1,\infty}(\Omega)} = \|(\mathscr{E}f)|_\Omega\|_{B^k_{1,\infty}(\Omega)} \leq \|\mathscr{E}f\|_{B^k_{1,\infty}(\mathbb{R}^d)} \leq C_1 \cdot \|\mathscr{E}f\|_{W^{k,1}(\mathbb{R}^d)} \leq C_1 \|\mathscr{E}\| \|f\|_{W^{k,1}(\Omega)},$$

so that we can choose $C = C_1 \cdot \|\mathscr{E}\|$. In particular, this implies that $\mathcal{S} \subset C \cdot \mathcal{S}_{\mathbf{X}} \subset L^2(\Omega)$ is bounded, so that Lemma 16 shows that $\mathcal{S} = \mathcal{S} \cap L^2(\Omega) \subset L^2(\Omega)$ is measurable.

Overall, we see that if we choose

$$\Phi : \ \mathcal{S}_{\mathbf{X}} \to L^2(\Omega), \quad f \mapsto C \cdot f \quad \text{and} \quad \Psi : \ \mathcal{S}_{\mathbf{Y}} \to \mathcal{S}, \quad f \mapsto \kappa \cdot f,$$

then Φ, Ψ are well-defined and satisfy all assumptions of Theorem 5. This theorem then shows that $s^*_{L^2(\Omega)}(\mathcal{S}) = \frac{k}{d}$ and that $\mathbb{P} := \mathbb{P}_0 \circ \Psi^{-1}$ is a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$.

Finally, Part (iii) of Theorem 7 yields a codec $\mathcal{C}^* = \left((E^*_R, D^*_R) \right)_{R \in \mathbb{N}} \in$ Codecs$_{\mathcal{S}_{\mathbf{X}}, L^2(\Omega)}$ satisfying $\delta_{\mathcal{S}_{\mathbf{X}}, L^2(\Omega)}(E^*_R, D^*_R) \lesssim R^{-k/d}$. Since Φ is Lipschitz with $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$, Lemma 8 shows that there exists a codec $\mathcal{C} = \left((E_R, D_R) \right)_{R \in \mathbb{N}} \in$ Codecs$_{\mathcal{S}, L^2(\Omega)}$ that satisfies the estimate $\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-k/d}$, claimed in Part (iii) of Theorem 8. □

D.2 The Case $p = \infty$

Let $k \in \mathbb{N}$ with $k > d \cdot (\frac{1}{\infty} - \frac{1}{2})_+ = 0$ and define $\mathcal{S} := \text{Ball}(0, 1; W^{k,\infty}(\Omega))$. Note that trivially $\mathcal{S} \subset L^\infty(\Omega) \subset L^2(\Omega)$ is bounded, so that Lemma 16 implies that $\mathcal{S} \subset L^2(\Omega)$ is Borel measurable. Our goal is to apply Theorem 5 for $\mathbf{X} := \mathbf{Y} := \mathbf{Z} := L^2(\Omega)$, $\mathcal{S}_{\mathbf{X}} := \text{Ball}(0, 1; W^{k,2}(\Omega))$, and $\mathcal{S}_{\mathbf{Y}} := \text{Ball}(0, 1; B^k_{\infty,1}(\Omega))$, for suitable choices of Φ, Ψ and $\mathbb{P}$.

To this end, first note that since $\Omega \subset \mathbb{R}^d$ is bounded, there exists $C > 0$ satisfying $\|f\|_{W^{k,2}(\Omega)} \leq C \|f\|_{W^{k,\infty}(\Omega)}$ for all $f \in W^{k,\infty}(\Omega)$.

Next, it is well-known (see for instance [38, Example 7.2]) that there is $\kappa > 0$ such that $\kappa \|f\|_{W^{k,\infty}(\mathbb{R}^d)} \leq \|f\|_{B^k_{\infty,1}(\mathbb{R}^d)}$ for all $f \in B^k_{\infty,1}(\mathbb{R}^d)$. Now, for $f \in B^k_{\infty,1}(\Omega)$ and $\varepsilon > 0$, by definition of the norm on $B^k_{\infty,1}(\Omega)$ (see Eq. (C.2)) there

is some $g \in B^k_{\infty,1}(\mathbb{R}^d)$ with $\|g\|_{B^k_{\infty,1}(\mathbb{R}^d)} \leq (1+\varepsilon)\|f\|_{B^k_{\infty,1}(\Omega)}$ and $f = g|_\Omega$. Since $g \in B^k_{\infty,1}(\mathbb{R}^d) \subset W^{k,\infty}(\mathbb{R}^d)$, we see $f \in W^{k,\infty}(\Omega)$ and $\kappa\|f\|_{W^{k,\infty}(\Omega)} \leq \kappa\|g\|_{W^{k,\infty}(\mathbb{R}^d)} \leq \|g\|_{B^k_{\infty,1}(\mathbb{R}^d)} \leq (1+\varepsilon)\|f\|_{B^k_{\infty,1}(\Omega)}$. We have thus shown

$$\kappa\|f\|_{W^{k,\infty}(\Omega)} \leq \|f\|_{B^k_{\infty,1}(\Omega)} \qquad \forall f \in B^k_{\infty,1}(\Omega).$$

Finally, Theorems 7 and 8 (the latter applied with $p = 2 \in (1,\infty)$) show that the respective optimal compression rates are given by $s^*_{L^2(\Omega)}(\mathcal{S}_{\mathbf{X}}) = s^*_{L^2(\Omega)}(\mathcal{S}_{\mathbf{Y}}) = \frac{k}{d}$ and that there exists a Borel probability measure $\mathbb{P}_0$ on $\mathcal{S}_{\mathbf{Y}}$ that is critical for $\mathcal{S}_{\mathbf{Y}}$ with respect to $L^2(\Omega)$.

Combining these observations, it is not hard to see that all assumptions of Theorem 5 are satisfied for

$$\Phi : \quad \mathcal{S}_{\mathbf{X}} \to L^2(\Omega), \quad f \mapsto C \cdot f \quad \text{and} \quad \Psi : \quad \mathcal{S}_{\mathbf{Y}} \to \mathcal{S}, \quad f \mapsto \kappa \cdot f.$$

This theorem thus shows that $s^*_{L^2(\Omega)}(\mathcal{S}) = \frac{k}{d}$ and that $\mathbb{P} := \mathbb{P}_0 \circ \Psi^{-1}$ is a Borel probability measure on $\mathcal{S}$ that is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$.

Finally, Theorem 8 shows that there exists $\mathcal{C}^* = \left((E^*_R, D^*_R)\right)_{R\in\mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}_{\mathbf{X}}, L^2(\Omega)}$ satisfying $\delta_{\mathcal{S}_{\mathbf{X}}, L^2(\Omega)}(E^*_R, D^*_R) \lesssim R^{-k/d}$. Since Φ is Lipschitz with $\Phi(\mathcal{S}_{\mathbf{X}}) \supset \mathcal{S}$, Lemma 8 shows that there exists a codec $\mathcal{C} = \left((E_R, D_R)\right)_{R\in\mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S}, L^2(\Omega)}$ satisfying the estimate $\delta_{\mathcal{S}, L^2(\Omega)}(E_R, D_R) \lesssim R^{-k/d}$, as claimed in Part (iii) of Theorem 8. $\qquad\square$

E Measurability of Besov and Sobolev Balls

In this subsection, we show for the range of parameters considered in Theorems 7 and 8 that the balls $\mathrm{Ball}\left(0, R; B^\tau_{p,q}(\Omega)\right)$ and $\mathrm{Ball}\left(0, R; W^{k,p}(\Omega)\right)$ are measurable subsets of $L^2(\Omega)$. We remark that for the case where $p, q \in (1,\infty)$, easier proofs than the ones given here are possible. Yet, since the proofs for the cases where $p \in \{1,\infty\}$ or $q \in \{1,\infty\}$ apply verbatim for a whole range of exponents, we prefer to state and prove the more general results.

We begin with the case of Besov spaces, for which the balls are in fact closed.

Lemma 14 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be open and bounded and let $p, q \in (0,\infty]$ and $\tau \in \mathbb{R}$ with $\tau > d \cdot (\frac{1}{p} - \frac{1}{2})_+$. Then, $B^\tau_{p,q}(\Omega) \hookrightarrow L^2(\Omega)$, and the balls $\mathrm{Ball}(0, R; B^\tau_{p,q}(\Omega)) \subset L^2(\Omega)$ are closed for all $R > 0$.*

Proof Let $p_0 := \max\{p, 2\}$. Then, [38, Example 7.2] shows that $B^\tau_{p,q}(\mathbb{R}^d) \hookrightarrow L^{p_0}(\mathbb{R}^d)$, since $p \leq p_0$ and since $\tau > d \cdot (\frac{1}{p} - \frac{1}{p_0})$ by our assumptions on τ. This implies $B^\tau_{p,q}(\Omega) \hookrightarrow L^2(\Omega)$, since if $f \in B^\tau_{p,q}(\Omega)$, then by definition of this space (see Eqs. (C.1) and (C.2)) there exists some $g \in B^\tau_{p,q}(\mathbb{R}^d)$ satisfying

$\|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le 2\|f\|_{B_{p,q}^{\tau}(\Omega)}$ and $f = g|_{\Omega}$, and hence

$$\|f\|_{L^2(\Omega)} \lesssim \|f\|_{L^{p_0}(\Omega)} = \|g|_{\Omega}\|_{L^{p_0}(\Omega)} \le \|g\|_{L^{p_0}(\mathbb{R}^d)} \lesssim \|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le 2\|f\|_{B_{p,q}^{\tau}(\Omega)}.$$

It remains to show that $\mathrm{Ball}(0, R; B_{p,q}^{\tau}(\Omega)) \subset L^2(\Omega)$ is closed. To see this, first note that if $(g_n)_{n \in \mathbb{N}} \subset B_{p,q}^{\tau}(\mathbb{R}^d)$ satisfies $g_n \to g \in \mathscr{S}'(\mathbb{R}^d)$ with convergence in $\mathscr{S}'(\mathbb{R}^d)$, then $\|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le \liminf_{n \to \infty} \|g_n\|_{B_{p,q}^{\tau}(\mathbb{R}^d)}$. Indeed, with the family $(\varphi_j)_{j \in \mathbb{N}_0} \subset \mathscr{S}(\mathbb{R}^d)$ used in the definition of Besov spaces (see Sect. C.1), we have for $f \in \mathscr{S}'(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ that $\mathcal{F}^{-1}(\varphi_j \cdot \widehat{f})(x) = (2\pi)^{-d/2} \cdot \langle \widehat{f}, e^{i\langle x, \bullet \rangle} \varphi_j \rangle_{\mathscr{S}', \mathscr{S}}$; see for instance [30, Theorem 7.23]. From this, we easily see that $\mathcal{F}^{-1}(\varphi_j \cdot \widehat{g_n}) \to \mathcal{F}^{-1}(\varphi_j \cdot \widehat{g})$, with pointwise convergence as $n \to \infty$. Therefore, Fatou's lemma shows that $\|\mathcal{F}^{-1}(\varphi_j \cdot \widehat{g})\|_{L^p} \le \liminf_{n \to \infty} \|\mathcal{F}^{-1}(\varphi_j \cdot \widehat{g_n})\|_{L^p}$. By another application of Fatou's lemma, we therefore see

$$\|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} = \left\| \left(2^{\tau j} \|\mathcal{F}^{-1}(\varphi_j \cdot \widehat{g})\|_{L^p} \right)_{j \in \mathbb{N}_0} \right\|_{\ell^q} \le \liminf_{n \to \infty} \left\| \left(2^{\tau j} \|\mathcal{F}^{-1}(\varphi_j \cdot \widehat{g_n})\|_{L^p} \right)_{j \in \mathbb{N}_0} \right\|_{\ell^q}$$
$$= \liminf_{n \to \infty} \|g_n\|_{B_{p,q}^{\tau}(\mathbb{R}^d)},$$

as claimed.

Now, we prove the claimed closedness. Let $(f_n)_{n \in \mathbb{N}} \subset \mathrm{Ball}(0, R; B_{p,q}^{\tau}(\Omega)) \subset L^2(\Omega)$ such that $f_n \to f \in L^2(\Omega)$ with convergence in $L^2(\Omega)$. By definition of $B_{p,q}^{\tau}(\Omega)$ (see Eqs. (C.1) and (C.2)), for each $n \in \mathbb{N}$ there exists $g_n \in B_{p,q}^{\tau}(\mathbb{R}^d)$ satisfying $f_n = g_n|_{\Omega}$ and

$$\|g_n\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le \left(1 + \tfrac{1}{n} \right) \cdot \|f_n\|_{B_{p,q}^{\tau}(\Omega)} \le \left(1 + \tfrac{1}{n} \right) R \le 2R.$$

As seen above, $B_{p,q}^{\tau}(\mathbb{R}^d) \hookrightarrow L^{p_0}(\mathbb{R}^d)$, so that $(g_n)_{n \in \mathbb{N}} \subset L^{p_0}(\mathbb{R}^d) = (L^{p_0'}(\mathbb{R}^d))'$ is bounded, where $p_0' \le 2 < \infty$, so that $L^{p_0'}(\mathbb{R}^d)$ is separable. Thus, [2, Theorem 8.5] shows that there is a subsequence $(g_{n_k})_{k \in \mathbb{N}}$ and some $g \in L^{p_0}(\mathbb{R}^d)$ such that $g_{n_k} \to g$ in the weak-$*$-sense in $L^{p_0}(\mathbb{R}^d) = (L^{p_0'}(\mathbb{R}^d))'$. In particular, $g_{n_k} \to g$ in $\mathscr{S}'(\mathbb{R}^d)$. By what we showed above, this implies $\|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le \liminf_{k \to \infty} \|g_{n_k}\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le R$. Finally, we have for any $\varphi \in C_c^{\infty}(\Omega)$ that $\langle g, \varphi \rangle = \lim_{k \to \infty} \langle g_{n_k}, \varphi \rangle = \lim_{k \to \infty} \langle f_{n_k}, \varphi \rangle = \langle f, \varphi \rangle$, since $f_{n_k} = g_{n_k}|_{\Omega}$ and $f_{n_k} \to f$ in $L^2(\Omega)$. Overall, we thus see that $f = g|_{\Omega} \in B_{p,q}^{\tau}(\Omega)$ and $\|f\|_{B_{p,q}^{\tau}(\Omega)} \le \|g\|_{B_{p,q}^{\tau}(\mathbb{R}^d)} \le R$.   □

For the Sobolev spaces $W^{k,p}(\Omega)$ with $p = 1$, the set $\mathrm{Ball}\left(0, R; W^{k,1}(\Omega)\right)$ is not closed in $L^2(\Omega)$. In order to show that this ball is nonetheless Borel measurable, we begin with the following result on $\mathbb{R}^d$.

Lemma 15 *Let $d, k \in \mathbb{N}$ and $p \in [1, 2]$. Then, $L^2(\mathbb{R}^d) \cap W^{k,p}(\mathbb{R}^d)$ is a Borel-measurable subset of $L^2(\mathbb{R}^d)$.*

Proof Let $\varphi \in C_c^{\infty}(\mathbb{R}^d)$ with $\varphi \ge 0$ and $\int_{\mathbb{R}^d} \varphi(x) \, dx = 1$ and define $\varphi_n(x) := n^d \cdot \varphi(nx)$. It follows from [2, Section 4.13] that if $f \in L^2(\mathbb{R}^d)$, then $\varphi_n * f \in L^2(\mathbb{R}^d) \cap C^{\infty}(\mathbb{R}^d)$ with $\partial^{\alpha}(\varphi_n * f) = (\partial^{\alpha} \varphi_n) * f$.

Step 1: Define $\mathcal{S} := L^2(\mathbb{R}^d) \cap W^{k,p}(\mathbb{R}^d)$. In this step, we show that

$$\mathcal{S} = \left\{ f \in L^2(\mathbb{R}^d) \quad : \quad \forall |\alpha| \leq k : \left((\partial^\alpha \varphi_n) * f \right)_{n \in \mathbb{N}} \text{ is Cauchy in } L^p(\mathbb{R}^d) \right\}.$$

For "$\subset$", note that if $f \in \mathcal{S}$, then from the definition of the weak derivative we see

$$[(\partial^\alpha \varphi_n) * f](x) = \int_{\mathbb{R}^d} f(y) \cdot (\partial^\alpha \varphi_n)(x - y)\, dy = (-1)^{|\alpha|} \int_{\mathbb{R}^d} f(y) \cdot \partial_y^\alpha[\varphi_n(x - y)]\, dy$$

$$= \int_{\mathbb{R}^d} \partial^\alpha f(y) \cdot \varphi_n(x - y)\, dy = [\varphi_n * (\partial^\alpha f)](x),$$

so that [2, Theorem 4.15] shows that $(\partial^\alpha \varphi_n) * f \xrightarrow[n \to \infty]{} \partial^\alpha f$, with convergence in $L^p(\mathbb{R}^d)$. This proves "$\subset$".

For "$\supset$", let $f \in L^2(\mathbb{R}^d)$ such that $\left((\partial^\alpha \varphi_n) * f \right)_{n \in \mathbb{N}}$ is Cauchy in $L^p(\mathbb{R}^d)$ for each $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$. Define $g_\alpha := \lim_{n \to \infty}[(\partial^\alpha \varphi_n) * f] \in L^p(\mathbb{R}^d)$ for $|\alpha| \leq k$. Since [2, Theorem 4.15] shows that $\varphi_n * f \to f$ with convergence in L^2, we get $f = g_0 \in L^p(\mathbb{R}^d)$. Furthermore, as seen above, we have $\varphi_n * f \in C^\infty(\mathbb{R}^d)$ with $\partial^\alpha(\varphi_n * f) = (\partial^\alpha \varphi_n) * f$. Therefore, we see for arbitrary $\psi \in C_c^\infty(\mathbb{R}^d)$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$ that

$$\int_{\mathbb{R}^d} f \cdot \partial^\alpha \psi \, dx = \lim_{n \to \infty} \int_{\mathbb{R}^d} (\varphi_n * f) \cdot \partial^\alpha \psi \, dx = \lim_{n \to \infty} (-1)^{|\alpha|} \int_{\mathbb{R}^d} [(\partial^\alpha \varphi_n) * f] \cdot \psi \, dx$$

$$(\text{since } \psi \in C_c^\infty \subset L^{p'}) = (-1)^{|\alpha|} \int_{\mathbb{R}^d} g_\alpha \cdot \psi \, dx,$$

which shows that g_α is the α-th weak derivative of f; that is, $\partial^\alpha f = g_\alpha \in L^p(\mathbb{R}^d)$. Since this holds for all $|\alpha| \leq k$, we see that $f \in W^{k,p}(\mathbb{R}^d)$ and thus $f \in \mathcal{S}$.

Step 2: For $n, m, M \in \mathbb{N}$, define

$$\Gamma_{n,m,M} : \quad L^2(\mathbb{R}^d) \to [0, \infty), \quad f \mapsto \left\| [\partial^\alpha(\varphi_n - \varphi_m) * f] \cdot \mathbb{1}_{[-M,M]^d} \right\|_{L^p}.$$

Since $p \leq 2$, it is easy to see that $\Gamma_{n,m,M}$ is well-defined and continuous. Furthermore, $\|[(\partial^\alpha \varphi_n) * f] - [(\partial^\alpha \varphi_m) * f]\|_{L^p} = \sup_{M \in \mathbb{N}} \Gamma_{n,m,M}(f)$, which—together with the result from Step 1—implies that

$$\mathcal{S} = \bigcap_{\ell=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n,m=N}^{\infty} \bigcap_{M=1}^{\infty} \left\{ f \in L^2(\mathbb{R}^d) : \Gamma_{n,m,M}(f) \leq 1/\ell \right\}$$

is a Borel-measurable subset of $L^2(\mathbb{R}^d)$. $\qquad \square$

We can now prove a similar result on bounded domains. For the convenience of the reader, we recall that $\|f\|_{W^{k,p}} = \max_{|\alpha| \leq k} \|\partial^\alpha f\|_{L^p}$; see Eq. (4.2).

Lemma 16 *Let $p \in [1, \infty]$, $k \in \mathbb{N}$, $R \in (0, \infty)$, and let $\Omega \subset \mathbb{R}^d$ be open and bounded. In case of $p = 1$, assume additionally that Ω is a Lipschitz domain.*
Then, $L^2(\Omega) \cap \text{Ball}\left(0, R; W^{k,p}(\Omega)\right)$ is a Borel-measurable subset of $L^2(\Omega)$.

Proof **Step 1:** The space $C^k(\overline{\Omega})$ (with the norm $\|f\|_{C^k(\overline{\Omega})} = \max_{|\alpha| \le k} \|\partial^\alpha f\|_{\sup}$) is separable; see [2, Section 4.18]. Since subsets of separable spaces are separable, there exists a sequence $(\varphi_n)_{n \in \mathbb{N}} \subset C_c^\infty(\Omega) \setminus \{0\}$ that is dense in $C_c^\infty(\Omega) \setminus \{0\}$ with respect to $\| \bullet \|_{C^k(\overline{\Omega})}$. For $n \in \mathbb{N}$, define

$$\gamma_n : \quad L^2(\Omega) \to [0, \infty), \quad f \mapsto \max_{|\alpha| \le k} \left| \int_\Omega f \cdot \partial^\alpha \varphi_n \, dx \right| \bigg/ \|\varphi_n\|_{L^{p'}},$$

where $p' \in [1, \infty]$ is the conjugate exponent to p. Since $\partial^\alpha \varphi_n \in C_c^\infty(\Omega) \subset L^2(\Omega)$, we see that γ_n is continuous, so that $\gamma : L^2(\Omega) \to [0, \infty]$, $f \mapsto \sup_{n \in \mathbb{N}} \gamma_n(f)$ is Borel measurable.

Step 2: We claim that $\left| \int_\Omega f \cdot \partial^\alpha \varphi \, dx \right| \le \gamma(f) \cdot \|\varphi\|_{L^{p'}}$ for all $f \in L^2(\Omega)$, $\varphi \in C_c^\infty(\Omega)$, and $|\alpha| \le k$. Clearly, we can assume without loss of generality that $\gamma(f) < \infty$ and $\varphi \ne 0$. Thus, there is a sequence $(n_\ell)_{\ell \in \mathbb{N}} \subset \mathbb{N}$ such that $\|\varphi - \varphi_{n_\ell}\|_{C^k(\overline{\Omega})} \to 0$, which easily implies $\|\varphi_{n_\ell}\|_{L^{p'}} \to \|\varphi\|_{L^{p'}}$ and $\partial^\alpha \varphi_{n_\ell} \to \partial^\alpha \varphi$ with convergence in $L^2(\Omega)$ for all $|\alpha| \le k$. Hence, $\left| \int_\Omega f \cdot \partial^\alpha \varphi \, dx \right| = \lim_{\ell \to \infty} \left| \int_\Omega f \cdot \partial^\alpha \varphi_{n_\ell} \, dx \right| \le \lim_{\ell \to \infty} \gamma(f) \cdot \|\varphi_{n_\ell}\|_{L^{p'}} = \gamma(f) \cdot \|\varphi\|_{L^{p'}}$, as claimed.

Step 3: In this step, we prove for $p > 1$ that $\mathcal{S} := L^2(\Omega) \cap \mathrm{Ball}(0, R; W^{k,p}(\Omega))$ satisfies $\mathcal{S} = \{f \in L^2(\Omega) : \gamma(f) \le R\}$, which then implies that $\mathcal{S}$ is a Borel measurable subset of $L^2(\Omega)$.

First, if $f \in \mathcal{S}$, then $\left| \int_\Omega f \, \partial^\alpha \varphi_n \, dx \right| = \left| \int_\Omega \varphi_n \, \partial^\alpha f \, dx \right| \le \|\partial^\alpha f\|_{L^p} \cdot \|\varphi_n\|_{L^{p'}} \le R \cdot \|\varphi_n\|_{L^{p'}}$ for all $|\alpha| \le k$ and $n \in \mathbb{N}$, so that $\gamma(f) \le R$.

Conversely, if $\gamma(f) \le R$, then Step 2 shows for arbitrary $|\alpha| \le k$ and $\varphi \in C_c^\infty(\Omega)$ that $\left| \int_\Omega f \cdot \partial^\alpha \varphi \, dx \right| \le \gamma(f) \cdot \|\varphi\|_{L^{p'}} \le R \cdot \|\varphi\|_{L^{p'}}$, so that [2, Section E6.7] implies that $f \in W^{k,p}(\Omega)$; this uses our assumption $p > 1$. Finally, for $\varphi \in C_c^\infty(\Omega)$ and $|\alpha| \le k$, we have $\left| \int_\Omega \varphi \cdot \partial^\alpha f \, dx \right| = \left| \int_\Omega f \cdot \partial^\alpha \varphi \, dx \right| \le R \cdot \|\varphi\|_{L^{p'}}$. Therefore, [2, Corollary 6.13] shows $\|\partial^\alpha f\|_{L^p} \le R$ for all $|\alpha| \le k$. By our definition of $\| \bullet \|_{W^{k,p}(\Omega)}$ (see Eq. (4.2)), this implies $f \in \mathcal{S}$.

Step 4: We prove the claim for the case $p = 1$. Since Ω is a Lipschitz domain, [33, Theorem 5 in Chapter VI] yields a linear extension operator $E : L^1(\Omega) \to L^1(\mathbb{R}^d)$ satisfying $(Ef)|_\Omega = f$ for all $f \in L^1(\Omega)$, and such that for arbitrary $\ell \in \mathbb{N}_0$ and $q \in [1, \infty]$ the restriction $E : W^{\ell,q}(\Omega) \to W^{\ell,q}(\mathbb{R}^d)$ is well-defined and bounded.

In particular, $E : L^2(\Omega) \to L^2(\mathbb{R}^d)$ is continuous and hence measurable. By Lemma 15, this means that $\Theta := \{f \in L^2(\Omega) : Ef \in W^{k,1}(\mathbb{R}^d)\} \subset L^2(\Omega)$ is measurable. We claim that

$$(\mathcal{S} := L^2(\Omega) \cap \mathrm{Ball}(0, R; W^{k,1}(\Omega)) = \Theta \cap \{f \in L^2(\Omega) : \gamma(f) \le R\},)$$

which then implies that $\mathcal{S} \subset L^2(\Omega)$ is measurable.

For "$\subset$", we see as in Step 3 that $\gamma(f) \le R$ if $f \in \mathcal{S}$. Furthermore, by the properties of the extension operator E, we also have $f \in \Theta$ if $f \in \mathcal{S}$. For "$\supset$", let $f \in \Theta$ satisfy $\gamma(f) \le R$. Since $f \in \Theta$, we have $f = (Ef)|_\Omega \in W^{k,1}(\Omega)$. One can then argue as at the end of Step 3 (using [2, Corollary 6.13]) to see that $f \in \mathrm{Ball}(0, R; W^{k,1}(\Omega))$ and thus $f \in \mathcal{S}$. $\qquad\square$

F Proof of the Lower Bounds for Neural Network Approximation

In this appendix, we establish a connection between rate distortion theory and approximation by neural networks. This is based on the observation from [4,29] that one can use the existence of approximating networks to construct a codec for a function class. This in turn relies on sharp estimates for the number of functions in the class $\mathcal{NN}_{d,W}^{\sigma,\varrho}$ appearing in Theorem 3.

The first ingredient for such a bound is the following "compression lemma," showing that for each neural network Φ, one can find a "compressed network" Ψ with the same realization $R_\varrho \Psi = R_\varrho \Phi$ and such that the number of weights, layers, and neurons of Ψ is controlled by the number of weights of Φ. We remark that a similar result already appears in [29, Lemma G.1]. However, the proof in [29] proceeds by removing "dead neurons;" that is, neurons that do not receive input from any other neurons. The proof then relies on the assumption $\varrho(0) = 0$ to ensure that a "dead neuron" has trivial output. In contrast, our proof relies on removing "ignored neurons," whose output is not used in the subsequent layer; this allows us to omit the assumption $\varrho(0) = 0$.

Lemma 17 *Let $d, k \in \mathbb{N}$ and $\Theta \subset \mathbb{R}$ with $0 \in \Theta$. Let Φ be a neural network with all weights/biases contained in Θ and with $d_{\text{in}}(\Phi) = d$ and $d_{\text{out}}(\Phi) = k$. Then, there exists a "compressed" network Ψ with the following properties:*

1. *$d_{\text{in}}(\Psi) = d$ and $d_{\text{out}}(\Psi) = k$,*
2. *all weights/biases of Ψ are contained in Θ,*
3. *for every $\varrho : \mathbb{R} \to \mathbb{R}$, we have $R_\varrho \Phi = R_\varrho \Psi$,*
4. *$W(\Psi) \leq W(\Phi)$,*
5. *$L(\Psi) \leq W(\Psi) + 1$,*
6. *$N(\Psi) \leq k + d + W(\Psi)$.*

Proof For an arbitrary neural network Φ, define $\mathcal{C}(\Phi) := L(\Phi) + N(\Phi) \in \mathbb{N}$. Furthermore, define the set of "bad" networks as

$$\mathcal{B} := \left\{ \Phi : \begin{array}{l} \Phi \text{ NN with } d_{\text{in}}(\Phi) = d, d_{\text{out}}(\Phi) = k, \text{ all weights/biases of } \Phi \text{ belong to } \Theta, \\ \text{and } \nexists \text{ NN } \Psi \text{ satisfying properties 1-6 from the lemma} \end{array} \right\}.$$

Toward a contradiction, assume that the lemma is false. Then, there exists $\Phi^* \in \mathcal{B}$ satisfying $\mathcal{C}(\Phi^*) = \min_{\Phi \in \mathcal{B}} \mathcal{C}(\Phi)$. Write $\Phi^* = ((A_1, b_1), \dots, (A_L, b_L))$ with $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b_\ell \in \mathbb{R}^{N_\ell}$ where $N_0 = d$ and $N_L = k$. Note that if $L = 1$, then $L(\Phi^*) = 1 \leq W(\Phi^*) + 1$ and $N(\Phi^*) \leq k + d \leq k + d + W(\Phi^*)$, so that properties 1–6 of the lemma are satisfied for $\Psi = \Phi^*$ (and Φ^* instead of Φ), in contradiction to $\Phi^* \in \mathcal{B}$. Hence, $L \geq 2$.

Below, we will show that

$$\neg\left(\exists \ell \in \{2, \dots, L\} \ \exists j \in [N_{\ell-1}] \ \forall i \in [N_\ell] : \quad (A_\ell)_{i,j} = 0 \right). \tag{F.1}$$

Thus, for every $\ell \in \{2, \dots, L\}$ and $j \in [N_{\ell-1}]$, we have $\sum_{i=1}^{N_\ell} \mathbb{1}_{(A_\ell)_{i,j} \neq 0} \geq 1$ and hence $W(\Phi^*) \geq \sum_{\ell=2}^{L} \sum_{j=1}^{N_{\ell-1}} \sum_{i=1}^{N_\ell} \mathbb{1}_{(A_\ell)_{i,j} \neq 0} \geq \sum_{\ell=2}^{L} N_{\ell-1}$. On the one hand, this

implies $W(\Phi^*) \geq L - 1$ and on the other hand

$$W(\Phi^*) \geq \sum_{\ell=2}^{L} N_{\ell-1} = N(\Phi^*) - N_L - N_0 = N(\Phi^*) - k - d.$$

Thus, choosing $\Psi = \Phi^*$, we see that Ψ satisfies properties 1–6 of the lemma (with Φ replaced by Φ^*), in contradiction to $\Phi^* \in \mathcal{B}$.

It remains to prove Eq. (F.1). Assume toward a contradiction that it is false. Then, there exist $\ell^* \in \{2, \ldots, L\}$ and $j^* \in [N_{\ell^*-1}]$ satisfying $(A_{\ell^*})_{-,j^*} = 0$. We distinguish two cases:

Case 1 ($N_{\ell^*-1} \geq 2$): In this case, define a new network

$$\Phi := \big((A_1, b_1), \ldots, (A_{\ell^*-2}, b_{\ell^*-2}), (A, b), (B, c), (A_{\ell^*+1}, b_{\ell^*+1}), \ldots, (A_L, b_L)\big),$$

where $A \in \mathbb{R}^{(N_{\ell^*-1}-1) \times N_{\ell^*-2}}$ and $b \in \mathbb{R}^{N_{\ell^*-1}-1}$ are obtained from A_{ℓ^*-1} and b_{ℓ^*-1}, respectively, by dropping the j^*-th row. Further, we set $c := b_{\ell^*} \in \mathbb{R}^{N_{\ell^*}}$ and we choose $B \in \mathbb{R}^{N_{\ell^*} \times (N_{\ell^*-1}-1)}$ as the result of dropping the j^*-th column of A_{ℓ^*}.

Since $(A_{\ell^*})_{-,j^*} = 0$, it is easy to see $R_\varrho \Phi = R_\varrho \Phi^*$ for every $\varrho : \mathbb{R} \to \mathbb{R}$. Note that $\mathcal{C}(\Phi) = \mathcal{C}(\Phi^*) - 1$, so that $\Phi \notin \mathcal{B}$ by "minimality" of Φ^*. Thus, there exists a network Ψ satisfying properties 1–6 from the statement of the lemma. Note in particular that $R_\varrho \Phi^* = R_\varrho \Phi = R_\varrho \Psi$ and that $W(\Psi) \leq W(\Phi) \leq W(\Phi^*)$; hence, Ψ also satisfies properties 1–6 of the lemma for Φ^* instead of Φ, in contradiction to $\Phi^* \in \mathcal{B}$.

Case 2 ($N_{\ell^*-1} = 1$): Since $j^* \in [N_{\ell^*-1}]$, this implies $j^* = 1$. Furthermore, since $A_{\ell^*} \in \mathbb{R}^{N_{\ell^*} \times N_{\ell^*-1}} = \mathbb{R}^{N_{\ell^*} \times 1}$ and $(A_{\ell^*})_{-,1} = (A_{\ell^*})_{-,j^*} = 0$, this implies $A_{\ell^*} = 0$. Define $\Phi := \big((0_{N_{\ell^*} \times d}, b_{\ell^*}), (A_{\ell^*+1}, b_{\ell^*+1}), \ldots, (A_L, b_L)\big)$ and note for any $\varrho : \mathbb{R} \to \mathbb{R}$ and $T_\ell x := A_\ell x + b_\ell$ that $T_{\ell^*} x = b_{\ell^*}$ for all $x \in \mathbb{R}^{N_{\ell^*-1}}$, which implies that

$$
\begin{aligned}
&R_\varrho \Phi^*(x) \\
&= \underbrace{\big[T_L \circ (\varrho \circ T_{L-1}) \circ \cdots \circ (\varrho \circ T_{\ell^*+1}) \circ \varrho\big]}_{\text{omit this if } \ell^*=L} \big(T_{\ell^*}\big((\varrho \circ T_{\ell^*-1}) \circ \cdots \circ (\varrho \circ T_1)(x)\big)\big) \\
&= \underbrace{\big[T_L \circ (\varrho \circ T_{L-1}) \circ \cdots \circ (\varrho \circ T_{\ell^*+1}) \circ \varrho\big]}_{\text{omit this if } \ell^*=L} (b_{\ell^*}) = R_\varrho \Phi(x) \quad \forall x \in \mathbb{R}^d.
\end{aligned}
$$

Furthermore, note that all weights/biases of Φ belong to $\Theta \cup \{0\} = \Theta$ and because of $\ell^* \geq 2$ that $L(\Phi) = L - \ell^* + 1 \leq L - 1 < L(\Phi^*)$ and $N(\Phi) \leq N(\Phi^*)$, so that $\mathcal{C}(\Phi) < \mathcal{C}(\Phi^*)$. Now, since also $W(\Phi) \leq W(\Phi^*)$, one obtains a contradiction exactly as in Case 1. $\qquad \square$

Using the preceding compression result, we can now derive a sharp bound on the cardinality of the set $\mathcal{NN}_{d,W}^{\sigma,\varrho}$ of neural network functions that appears in Theorem 3. We mention that this result is similar to [29, Lemma B.4], but without the assumption $\varrho(0) = 0$.

Lemma 18 *Let $\varrho : \mathbb{R} \to \mathbb{R}$ and $d, \sigma \in \mathbb{N}$. With $\mathcal{NN}_{d,W}^{\sigma,\varrho}$ as defined in Theorem 3, there exists a constant $C_0 = C_0(d, \sigma) \in \mathbb{N}$ satisfying*

$$\left| \mathcal{NN}_{d,W}^{\sigma,\varrho} \right| \leq 2^{C_0 \cdot W \cdot \lceil \log_2(1+W) \rceil^2} \quad \forall W \in \mathbb{N}.$$

In fact, one can choose $C_0 = 4 + 4\lceil \log_2(4ed) \rceil + 8\sigma$.

Proof Fix $W \in \mathbb{N}$.

Step 1: Recall the definition of $G_{\sigma,W}$ (Eq. (1.7)) and note $|\mathbb{Z} \cap [a, b]| \leq 1 + (b-a)$ for $a \leq b$. Thus, noting that $\log_2(1 + W) \geq \log_2(2) = 1$, we easily see

$$
\begin{aligned}
|G_{\sigma,W}| &= \left| \mathbb{Z} \cap 2^{\sigma \lceil \log_2 W \rceil^2} \left[-W^{\sigma \lceil \log_2 W \rceil}, W^{\sigma \lceil \log_2 W \rceil} \right] \right| \\
&\leq 1 + 2 \cdot W^{\sigma \lceil \log_2 W \rceil} 2^{\sigma \lceil \log_2 W \rceil^2} \leq 1 + 2 \cdot 2^{2\sigma \lceil \log_2 W \rceil^2} \leq 3 \cdot 2^{2\sigma \lceil \log_2(1+W) \rceil^2} \\
&\leq 2^{4\sigma \lceil \log_2(1+W) \rceil^2}.
\end{aligned}
$$

Step 2: Given $L \in \mathbb{N}$, define $N_0 := d$ and $N_L := 1$ and define the set of all architectures with L layers and at most W "hidden neurons" as

$$\mathscr{A}_L := \left\{ N = (N_1, \ldots, N_{L-1}) \in \mathbb{N}^{L-1} : \sum_{\ell=1}^{L-1} N_\ell \leq W \right\},$$

with the understanding that $\mathscr{A}_1$ contains a single element, the empty tuple. Next, for fixed architecture $N \in \mathscr{A}_L$, define

$$H_N := \bigcup_{\ell=1}^{L} \left(\{\ell\} \times [N_\ell] \times [N_{\ell-1}] \right) \quad \text{and} \quad R_N := \bigcup_{\ell=1}^{L} \left(\{\ell\} \times [N_\ell] \right).$$

Finally, given subsets $I \subset H_N$ and $J \subset R_N$, define

$$\Omega_{\ell,i,j}^I := \begin{cases} \{0\}, & \text{if } (\ell, i, j) \notin I, \\ G_{\sigma,W}, & \text{otherwise} \end{cases} \quad \text{and} \quad \Theta_{\ell,k}^J := \begin{cases} \{0\}, & \text{if } (\ell, k) \notin J, \\ G_{\sigma,W}, & \text{otherwise} \end{cases}$$

for $\ell, i, j, k \in \mathbb{N}$. Overall, recalling the definition of $\mathcal{NN}_{d,W}^{\sigma,\varrho}$ and using Lemma 17, we see

$$\mathcal{NN}_{d,W}^{\sigma,\varrho} \subset \bigcup_{L=1}^{W+1} \bigcup_{N \in \mathscr{A}_L} \bigcup_{\substack{I \subset H_N \\ |I| \leq W}} \bigcup_{\substack{J \subset R_N \\ |J| \leq W}} \left\{ R_\varrho \Psi : \Psi \in \bigtimes_{\ell=1}^{L} \left(\bigtimes_{(i,j) \in [N_\ell] \times [N_{\ell-1}]} \Omega_{\ell,i,j}^I \times \bigtimes_{k \in [N_\ell]} \Theta_{\ell,k}^J \right) \right\}.$$

Finally, note that $|R_N| = \sum_{\ell=1}^{L} N_\ell \leq W + 1 \leq 4d \cdot W^2$ and

$$|H_N| = \sum_{\ell=1}^{L} N_\ell N_{\ell-1} \leq \left(\max_{0 \leq i \leq L-1} N_i \right) \sum_{\ell=1}^{L} N_\ell \leq (d + W) \cdot (W + 1) \leq 4d \cdot W^2.$$

Step 3: We complete the proof by estimating $|\mathcal{NN}_{d,W}^{\sigma,\varrho}|$ using the inclusion from Step 2. First, note that if Λ is any non-empty set with $|\Lambda| \leq 4d \cdot W^2$, then a standard estimate for (sums of) binomial coefficients (see, e.g., [37, Exercise 0.0.5]) shows

$$|\{I \subset \Lambda : |I| \leq W\}| = \sum_{k=0}^{\min\{W,|\Lambda|\}} \binom{|\Lambda|}{k} \leq \left(e|\Lambda|/\min\{W,|\Lambda|\}\right)^{\min\{W,|\Lambda|\}} \leq (4edW)^W,$$

where the overall estimate also holds if $\Lambda = \emptyset$. Furthermore, using the estimate from Step 1, we see for $I \subset H_N$ and $J \subset R_N$ with $|I|, |J| \leq W$ that

$$\left| \bigtimes_{\ell=1}^{L} \left(\bigtimes_{(i,j)\in[N_\ell]\times[N_{\ell-1}]} \Omega_{\ell,i,j}^{I} \times \bigtimes_{k\in[N_\ell]} \Theta_{\ell,k}^{J} \right) \right| = \prod_{\ell=1}^{L} \left(\prod_{(i,j)\in[N_\ell]\times[N_{\ell-1}]} |\Omega_{\ell,i,j}^{I}| \prod_{k\in[N_\ell]} |\Theta_{\ell,k}^{J}| \right)$$

$$\leq |G_{\sigma,W}|^{|I|+|J|} \leq |G_{\sigma,W}|^{2W}$$

$$\leq 2^{8\sigma W \lceil \log_2(1+W) \rceil^2}.$$

Moreover, each $N = (N_1, \ldots, N_{L-1}) \in \mathscr{A}_L$ satisfies $N_\ell \leq W$, which implies for $L \leq W+1$ that $|\mathscr{A}_L| \leq |[W]^{L-1}| \leq W^W$. Overall, combining the preceding observations with the inclusion from Step 2, we conclude as claimed that

$$|\mathcal{NN}_{d,W}^{\sigma,\varrho}| \leq \sum_{L=1}^{W+1} \sum_{N\in\mathscr{A}_L} |\{I \subset H_N : |I| \leq W\}| \cdot |\{J \subset R_N : |J| \leq W\}| \cdot 2^{8\sigma W \lceil \log_2(1+W) \rceil^2}$$

$$\leq (W+1) \cdot W^W \cdot (4edW)^W \cdot (4edW)^W \cdot 2^{8\sigma W \lceil \log_2(1+W) \rceil^2}$$

$$\leq (4edW)^{4W} \cdot 2^{8\sigma W \lceil \log_2(1+W) \rceil^2} \leq 2^{(4+4\log_2(4ed)+8\sigma)\cdot W\cdot \lceil \log_2(1+W) \rceil^2}.$$

$\square$

Finally, using the preceding estimate for the cardinality of $\mathcal{NN}_{d,W}^{\sigma,\varrho}$, we can now formulate the precise connection between neural network approximation and rate distortion theory.

Lemma 19 *Let $\emptyset \neq \Omega \subset \mathbb{R}^d$ be measurable, let $\varrho : \mathbb{R} \to \mathbb{R}$ be measurable, and let $\sigma \in \mathbb{N}$. For $f \in L^2(\Omega)$ and $\varepsilon \in (0,1)$, let $W_{\sigma,\varrho}(f; \varepsilon) \in \mathbb{N} \cup \{\infty\}$ be as defined in Theorem 3. For $\tau > 0$ define*

$$\mathcal{A}_{\mathcal{NN},\sigma,\varrho}^{\tau} := \left\{ f \in L^2(\Omega) \ : \ \exists C > 0 \ \forall \varepsilon \in (0,1) : \ W_{\sigma,\varrho}(f; \varepsilon) \leq C \cdot \varepsilon^{-\tau} \right\}.$$

Then, there exists a codec $\mathcal{C} = \mathcal{C}(\sigma, \varrho, \Omega) \in \mathrm{Codecs}_{L^2(\Omega),L^2(\Omega)}$ such that

$$\mathcal{A}_{\mathcal{NN},\sigma,\varrho}^{\tau} \subset \mathcal{A}_{L^2(\Omega),L^2(\Omega)}^{\frac{1}{\tau}-\delta}(\mathcal{C}) \quad \forall \ \tau > 0 \text{ and } \delta \in (0, \tfrac{1}{\tau}).$$

Proof **Step 1:** (*Constructing the codec $\mathcal{C}$*): Let $C_0 = C_0(d,\sigma) \in \mathbb{N}$ as in Lemma 18. For $R \in \mathbb{N}_{\geq C_0}$, let $W_R \in \mathbb{N}$ be maximal with $C_0 \cdot W_R \cdot \lceil \log_2(1 + W_R) \rceil^2 \leq R$. By

Lemma 18, there exists a surjection $D_R : \{0, 1\}^R \to \mathcal{N}\mathcal{N}_{d, W_R}^{\sigma, \varrho}$. For each $f \in L^2(\Omega)$, choose $\mathbf{c}(R, f) \in \{0, 1\}^R$ such that

$$\|f - D_R(\mathbf{c}(R, f))\|_{L^2(\Omega)} = \min_{\mathbf{c} \in \{0,1\}^R} \|f - D_R(\mathbf{c})\|_{L^2(\Omega)} = \min_{g \in \mathcal{N}\mathcal{N}_{d, W_R}^{\sigma, \varrho}} \|f - g\|_{L^2(\Omega)},$$

and define $E_R : L^2(\Omega) \to \{0, 1\}^R$, $f \mapsto \mathbf{c}(R, f)$.

Finally, for $R \in \mathbb{N}$ with $R < C_0$, define

$$E_R : \ L^2(\Omega) \to \{0, 1\}^R, \ \ f \mapsto (0, \dots, 0) \quad \text{and} \quad D_R : \ \{0, 1\}^R \to L^2(\Omega), \ \ \mathbf{c} \mapsto 0.$$

Step 2: (*Completing the proof*): Let $\tau > 0$, $\delta \in (0, \frac{1}{\tau})$, and $f \in \mathcal{A}_{\mathcal{N}\mathcal{N}, \sigma, \varrho}^{\tau}$, so that there is $C = C(f) > 0$ satisfying $W_{\sigma, \varrho}(f; \varepsilon) \leq C \cdot \varepsilon^{-\tau}$ for all $\varepsilon \in (0, 1)$.

Since the logarithm grows slower than any positive power and since the maximality of W_R implies that $R \leq C_0 \cdot (W_R + 1) \cdot \lceil \log_2(W_R + 2) \rceil^2$, it is easy to see that there exists $C_1 = C_1(\tau, \delta, d, \sigma) > 0$ such that $R \leq C_1 \cdot W_R^{1/(1 - \tau \delta)}$ for all $R \in \mathbb{N}_{\geq C_0}$. Note that if R is large enough, then $\varepsilon := C^{1/\tau} \cdot C_1^{(1 - \delta \tau)/\tau} \cdot R^{-(\frac{1}{\tau} - \delta)}$ satisfies $\varepsilon \in (0, 1)$. For these R, we thus get

$$W_{\sigma, \varrho}(f; \varepsilon) \leq C \cdot \varepsilon^{-\tau} \leq C_1^{-(1 - \delta \tau)} \cdot R^{1 - \delta \tau} \leq W_R.$$

By definition of $W_{\sigma, \varrho}(f; \varepsilon)$ and by choice of D_R, E_R, we therefore see for all sufficiently large $R \in \mathbb{N}$ that

$$
\begin{aligned}
\left\| f - D_R(E_R(f)) \right\|_{L^2(\Omega)} &= \min_{g \in \mathcal{N}\mathcal{N}_{d, W_R}^{\sigma, \varrho}} \|f - g\|_{L^2(\Omega)} \\
&\leq \min_{g \in \mathcal{N}\mathcal{N}_{d, W_{\sigma, \varrho}(f; \varepsilon)}^{\sigma, \varrho}} \|f - g\|_{L^2(\Omega)} \\
&\leq \varepsilon = C^{1/\tau} \cdot C_1^{(1 - \delta \tau)/\tau} \cdot R^{-(\frac{1}{\tau} - \delta)},
\end{aligned}
$$

which easily implies that $f \in \mathcal{A}_{L^2(\Omega), L^2(\Omega)}^{\frac{1}{\tau} - \delta}(\mathcal{C})$. Since $\tau > 0$, $\delta \in (0, \frac{1}{\tau})$, and $f \in \mathcal{A}_{\mathcal{N}\mathcal{N}, \sigma, \varrho}^{\tau}$ were arbitrary, we are done. □

Proof of Theorem 3 **Part 1:** Let $s > s^*$. As shown in Theorems 7 and 8, the measure $\mathbb{P}$ from Theorem 1 is critical for $\mathcal{S}$ with respect to $L^2(\Omega)$. Thus (see Eq. (2.1)), there exist $c, \varepsilon_0 > 0$ such that $\mathbb{P}(\mathcal{S} \cap \text{Ball}(f, \varepsilon; L^2(\Omega))) \leq 2^{-c \cdot \varepsilon^{-1/s}}$ for all $f \in L^2(\Omega)$ and $\varepsilon \in (0, \varepsilon_0)$. Lemma 18 shows that $|\mathcal{N}\mathcal{N}_{d, W}^{\sigma, \varrho}| \leq 2^{C_0 W \lceil \log_2(1 + W) \rceil^2}$ for all $W \in \mathbb{N}$ and a suitable $C_0 = C_0(d, \sigma)$; in fact, one can take $C_0 = 4 + 4 \lceil \log_2(4ed) \rceil + 8\sigma$. Thus, setting $C := C_0$, we see that

$$
\Pr\left(\min_{g \in \mathcal{N}\mathcal{N}_{d, W}^{\sigma, \varrho}} \|f - g\|_{L^2(\Omega)} \leq \varepsilon \right) = \mathbb{P}\left(\bigcup_{g \in \mathcal{N}\mathcal{N}_{d, W}^{\sigma, \varrho}} \text{Ball}(g, \varepsilon; L^2(\Omega)) \right)
$$

$$
\leq 2^{C W \lceil \log_2(1 + W) \rceil^2} 2^{-c \cdot \varepsilon^{-1/s}}
$$

for all $\varepsilon \in (0, \varepsilon_0)$. This is precisely what is claimed in Part 1 of Theorem 3.

Part 2: For $\ell, \sigma \in \mathbb{N}$, define

$$\mathcal{A}_{\ell,\sigma} := \left\{ f \in \mathcal{S} \ : \ \exists\, C > 0 \ \ \forall\, \varepsilon \in (0,1): \ \ W_{\sigma,\varrho}(f;\varepsilon) \le C \cdot \varepsilon^{-(1-\frac{1}{2\ell})/s^*} \right\}.$$

It is not hard to see that $\mathcal{A}^*_{\mathcal{NN},\varrho} \subset \bigcup_{\sigma \in \mathbb{N}} \bigcup_{\ell \in \mathbb{N}} \mathcal{A}_{\ell,\sigma}$, so that it suffices to show $\mathbb{P}^*(\mathcal{A}_{\ell,\sigma}) = 0$ for all $\sigma, \ell \in \mathbb{N}$. To see this, let $\mathcal{C} \in \mathrm{Codecs}_{L^2(\Omega), L^2(\Omega)}$ as in Lemma 19, and note with the notation of that lemma and with $\delta := \frac{s^*/2}{2\ell-1}$ and $\tau = \left(1 - \frac{1}{2\ell}\right)/s^*$ that

$$\mathcal{A}_{\ell,\sigma} = \mathcal{S} \cap \mathcal{A}^{(1-\frac{1}{2\ell})/s^*}_{\mathcal{NN},\sigma,\varrho} \subset \mathcal{S} \cap \mathcal{A}^{\frac{s^*}{1-1/(2\ell)}-\delta}_{L^2(\Omega),L^2(\Omega)}(\mathcal{C}) = \mathcal{A}^{s^* \cdot \frac{2\ell-1/2}{2\ell-1}}_{\mathcal{S},L^2(\Omega)}(\mathcal{C}),$$

where Theorem 1 shows that $\mathbb{P}^*\left(\mathcal{A}^{s^* \cdot \frac{2\ell-1/2}{2\ell-1}}_{\mathcal{S},L^2(\Omega)}(\mathcal{C})\right) = 0$. $\qquad\square$

We close this section by proving the claim at the end of Remark 3.

Proof of Remark 3 Case 1 (Besov spaces): Here, $\mathcal{S} = \mathrm{Ball}(0, 1; B^\tau_{p,q}(\Omega; \mathbb{R}))$ and $s^* = \tau/d$, where $\tau > d \cdot (\frac{1}{p} - \frac{1}{2})_+$. By definition of the space $B^\tau_{p,q}(\Omega; \mathbb{R})$ (see Eqs. (C.1) and (C.2)), each $f \in \mathcal{S}$ extends to a function $\tilde{f} \in B^\tau_{p,q}(\mathbb{R}^d; \mathbb{R})$ satisfying $\|\tilde{f}\|_{B^\tau_{p,q}(\mathbb{R}^d;\mathbb{R})} \le 2$. Thanks to [36, Theorem in Section 2.5.12], this implies for a suitable $C_1 = C_1(d, p, q, \tau) > 0$ that $\|f\|_{B^{\tau,*}_{p,q}(\Omega)} \le C_1$, where $\| \cdot \|_{B^{\tau,*}_{p,q}(\Omega)}$ is the norm on the Besov space used in [34].

Now, [34, Proposition 1] yields $C_2, C_3, N_0, \theta \in \mathbb{N}_{\ge 2}$ (all depending only on d, p, q, τ) such that for every $f \in \mathcal{S}$ and $N \ge N_0$, there exists a network $\Phi = \Phi(f, N)$ satisfying $\|f - R_\varrho(\Phi)\|_{L^2(\Omega)} \le \frac{C_2}{2} \cdot N^{-s^*}$ as well as $L(\Phi) \le C_3 \cdot \log_2 N$ and $W(\Phi) \le C_3 \cdot N \cdot \log_2 N$, and such that all weights of Φ have absolute value at most $C_3 \cdot N^\theta$. This almost implies the desired estimate; the main issue is that the weights are merely bounded, but not necessarily quantized. To fix this, we will use [13, Lemma VI.8].

To make this formal, let us assume in what follows that $\varepsilon \in (0, \frac{1}{2}) \cap (0, C_2 \cdot N_0^{-s^*})$; it is easy to see that this implies the claim of Remark 3 for general $\varepsilon \in (0, \frac{1}{2})$. Let $N \in \mathbb{N}$ be minimal with $C_2 \cdot N^{-s^*} \le \varepsilon$, noting that this entails $N \ge N_0 \ge 2$ as well as $\varepsilon \le C_2 \cdot (N-1)^{-s^*} \le 2^{s^*} C_2 \cdot N^{-s^*}$, and therefore $N \le 2\, C_2^{1/s^*} \varepsilon^{-1/s^*}$. Now, given $f \in \mathcal{S}$, choose $\Phi = \Phi(f, N)$ as above and note that $\|f - R_\varrho(\Phi)\|_{L^2(\Omega)} \le \frac{\varepsilon}{2}$.

Define $W := \lceil C_3 \cdot N \cdot \log_2 N \rceil \ge N$ and choose $k = k(p, q, d, \tau) \in \mathbb{N}$ with $k \ge \frac{\theta}{s^*}$ so large that $C_3 \cdot \left(2C_2^{1/s^*}\right)^2 \le C_3 \cdot \left(2C_2^{1/s^*}\right)^\theta \le 2^k$.

Since $\log_2 N \le N$, we then see that $W \le C_3 N^2 \le C_3 \cdot \left(2\, C_2^{1/s^*} \varepsilon^{-1/s^*}\right)^2 \le (\frac{\varepsilon}{2})^{-k}$ and $C_3 \cdot N^\theta \le C_3 \cdot \left(2\, C_2^{1/s^*} \varepsilon^{-1/s^*}\right)^\theta \le (\frac{\varepsilon}{2})^{-k}$. Therefore, [13, Lemma VI.8] produces a network Φ' satisfying $W(\Phi') \le W(\Phi) \le W$ and $\|R_\varrho(\Phi) - R_\varrho(\Phi')\|_{L^\infty([0,1]^d)} \le \frac{\varepsilon}{2}$ and such that

all weights of Φ' belong to $[-(\frac{\varepsilon}{2})^{-\sigma_0}, (\frac{\varepsilon}{2})^{-\sigma_0}] \cap 2^{-\sigma_0 \lceil \log_2(2/\varepsilon) \rceil} \mathbb{Z}$, where $\sigma_0 := 3k L(\Phi)$.

To see that this implies the claim, first note that $\|f - R_\varrho(\Phi')\|_{L^2(\Omega)} \le \varepsilon$ and that

$$W(\Phi') \le W \le 2C_3\, N \log_2 N \le 4C_2^{\frac{1}{s^*}} C_3 \cdot \varepsilon^{-\frac{1}{s^*}} \cdot \log_2\big(2C_2^{\frac{1}{s^*}} \varepsilon^{-\frac{1}{s^*}}\big) \le C_4 \cdot \varepsilon^{-\frac{1}{s^*}} \log_2(1/\varepsilon)$$

for a suitable constant $C_4 = C_4(d, p, q, \tau)$. Regarding the quantization, first define $\sigma_1 := 3kC_3$, so that $\sigma_0 = 3kL(\Phi) \le 3kC_3 \cdot \log_2 N \le \sigma_1 \lceil \log_2 W \rceil$. Next, note that $\frac{2}{\varepsilon} \le \frac{C_2}{\varepsilon} \le N^{s^*} \le W^{s^*}$, meaning $(\varepsilon/2)^{-\sigma_0} \le (W^{s^*})^{\sigma_0} \le W^{\sigma \lceil \log_2 W \rceil}$ with $\sigma := \sigma_1 \lceil s^* \rceil$. Furthermore, we have $\log_2(2/\varepsilon) \le \log_2(W^{s^*}) \le \lceil s^* \rceil \lceil \log_2 W \rceil$, which easily implies that

$$2^{-\sigma_0 \lceil \log_2(2/\varepsilon) \rceil}\mathbb{Z} \subset 2^{-\sigma_0 \lceil s^* \rceil \lceil \log_2 W \rceil}\mathbb{Z} \subset 2^{-\sigma \lceil \log_2 W \rceil^2}\mathbb{Z}.$$

Overall, this shows that Φ' is (σ, W)-quantized, so that $R_\varrho \Phi' \in \mathcal{NN}_{d,W}^{\sigma,\varrho}$. Because of $\|f - R_\varrho(\Phi')\|_{L^2(\Omega)} \le \varepsilon$, this implies that $W_{\sigma,\varrho}(f; \varepsilon) \le W \le C_4 \cdot \varepsilon^{-1/s^*} \log_2(1/\varepsilon)$, which is what we wanted to show.

Case 2 (Sobolev spaces): Set $p_1 := \min\{p, 2\}$ and note $\mathcal{S} \subset \mathcal{S}' := \mathrm{Ball}(0, 1; W^{k,p_1}(\Omega))$. Since $\Omega = [0, 1]^d$ is a Lipschitz domain, [33, Chapter VI, Theorem 5] shows that each $f \in \mathcal{S}'$ extends to a function $\tilde{f} \in W^{k,p_1}(\mathbb{R}^d)$ with $\|\tilde{f}\|_{W^{k,p_1}(\mathbb{R}^d)} \le C_1$, where $C_1 = C_1(d, p, k)$. Now, Lemma 13 shows that $\tilde{f} \in B_{p_1,\infty}^k(\mathbb{R}^d)$ with $\|\tilde{f}\|_{B_{p_1,\infty}^k(\mathbb{R}^d)} \le C_2$ where $C_2 = C_2(d, p, k)$. Overall, this easily implies $\mathcal{S} \subset \mathrm{Ball}(0, C_2; B_{p_1,\infty}^k(\Omega; \mathbb{R}))$, so that the claim follows from that for the Besov spaces. Here, we implicitly used that the condition $k > d \cdot (\frac{1}{p} - \frac{1}{2})_+$ holds if and only if $k > d \cdot (\frac{1}{p_1} - \frac{1}{2})_+$, since $k \in \mathbb{N}$. $\qquad\square$

G Optimal Compression Rate and Baire Category

In Sect. 1.3.3, it was claimed that if $\mathcal{S} \subset \mathbf{X}$ satisfies some mild assumptions and $s^* = s_{\mathbf{X}}^*(\mathcal{S})$, then the set $\bigcup_{s > s^*} \mathcal{A}_{\mathcal{S},\mathbf{X}}^s(\mathcal{C})$ is of first category in $\mathcal{S}$, for every codec $\mathcal{C}$. In this appendix, we provide a proof of this fact. In particular, we will see (thanks to the Baire category theorem) that this implies existence of a "badly encodable" $\mathbf{x} = \mathbf{x}(\mathcal{C}) \in \mathcal{S} \setminus \bigcup_{s > s^*} \mathcal{A}_{\mathcal{S},\mathbf{X}}^s(\mathcal{C})$, meaning that $\mathbf{x}$ is not encoded at any rate $s > s_{\mathbf{X}}^*(\mathcal{S})$ by the codec $\mathcal{C}$.

Proposition 3 *Let $\mathbf{X}$ be a Banach space and let $\emptyset \ne \mathcal{S} \subset \mathbf{X}$. Assume that at least one of the following two conditions is satisfied:*

1. *$\mathcal{S}$ is closed, bounded, and convex; or*
2. *$\mathcal{S} = \{\mathbf{x} \in \mathbf{X} : \|\mathbf{x}\|_* \le r\}$ for some $r \in (0, \infty)$ and a map $\|\cdot\|_* : \mathbf{X} \to [0, \infty]$ with the following properties:*

   (a) *$\|\cdot\|_*$ is a quasi-norm; that is, there exists $\kappa \ge 1$ such that $\|\alpha\,\mathbf{x}\|_* = |\alpha| \cdot \|\mathbf{x}\|_*$ and $\|\mathbf{x} + \mathbf{y}\|_* \le \kappa \cdot (\|\mathbf{x}\|_* + \|\mathbf{y}\|_*)$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbf{X}$;*
   (b) *there exists $C \ge 1$ satisfying $\|\mathbf{x}\|_{\mathbf{X}} \le C \cdot \|\mathbf{x}\|_*$ for all $\mathbf{x} \in \mathbf{X}$;*
   (c) *$\mathcal{S} \subset \mathbf{X}$ is closed;*

(d) $\| \cdot \|_*$ is "continuous with respect to itself," meaning that $\|\mathbf{x}\|_* \to \|\mathbf{x}_0\|_*$ whenever $\|\mathbf{x} - \mathbf{x}_0\|_* \to 0$.

Set $s^* := s^*_{\mathbf{X}}(\mathcal{S})$ and assume $s^* < \infty$. Then, for any codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in$ Codecs$_{\mathcal{S}, \mathbf{X}}$, the set $\bigcup_{s > s^*} \mathcal{A}^s_{\mathcal{S}, \mathbf{X}}(\mathcal{C})$ is of first category in $\mathcal{S}$. Moreover, there exists $\mathbf{x} = \mathbf{x}(\mathcal{C}) \in \mathcal{S}$ such that for each $\ell \in \mathbb{N}$, we have

$$\big\| \mathbf{x} - D_R(E_R(\mathbf{x})) \big\|_{\mathbf{X}} \geq R^{-(s^* + \frac{1}{\ell})} \quad \text{for infinitely many } R \in \mathbb{N}. \tag{G.1}$$

Remark (1) It is not hard to see that if $\mathbf{x}$ satisfies (G.1), then $\mathbf{x} \in \mathcal{S} \setminus \bigcup_{s > s^*} \mathcal{A}^s_{\mathcal{S}, \mathbf{X}}(\mathcal{C})$.
(2) The assumptions on the quasi-norm $\| \cdot \|_*$ might appear quite technical, but they are usually satisfied. Indeed, the condition $\|\mathbf{x}\|_{\mathbf{X}} \leq C \cdot \|\mathbf{x}\|_*$ is equivalent to $\mathcal{S} \subset \mathbf{X}$ being bounded, which is necessary for having $s^*_{\mathbf{X}}(\mathcal{S}) > 0$. Next, most naturally appearing quasi-norms are q-norms for some $q \in (0, 1]$, meaning that $\|\mathbf{x} + \mathbf{y}\|_*^q \leq \|\mathbf{x}\|_*^q + \|\mathbf{y}\|_*^q$. In this case, it is not hard to see $\big| \|\mathbf{x}\|_*^q - \|\mathbf{y}\|_*^q \big| \leq \|\mathbf{x} - \mathbf{y}\|_*^q$, which implies that $\| \cdot \|_*$ is "continuous with respect to itself." Finally, most natural quasi-norms satisfy the *Fatou property*, meaning that if $\mathbf{x}_n \to \mathbf{x}$ in $\mathbf{X}$, then $\|\mathbf{x}\|_* \leq \liminf_{n \to \infty} \|\mathbf{x}_n\|_*$. If this is the case, then $\mathcal{S} \subset \mathbf{X}$ is closed.

Proof For $R \in \mathbb{N}$, define $M_R := \text{range}(D_R) \subset \mathbf{X}$. Furthermore, for $N, \ell \in \mathbb{N}$, let

$$\mathcal{G}_{N, \ell} := \big\{ \mathbf{x} \in \mathcal{S} : \forall R \in \mathbb{N} : \text{dist}_{\mathbf{X}}(\mathbf{x}, M_R) \leq N \cdot R^{-(s^* + \frac{1}{\ell})} \big\}.$$

Since $\text{dist}_{\mathbf{X}}(\cdot, M_R)$ is continuous, it is not hard to see that each set $\mathcal{G}_{N, \ell} \subset \mathcal{S}$ is closed. Denote by $\mathcal{G}^{\circ}_{N, \ell}$ the (relative) interior of $\mathcal{G}_{N, \ell}$ in $\mathcal{S}$.

Step 1: (*If $\mathcal{G}^{\circ}_{N, \ell} \neq \emptyset$, then there exist $\mathbf{x}'_0 \in \mathbf{X}$ and $t > 0$ such that $\mathbf{x}'_0 + t\mathcal{S} \subset \mathcal{G}_{N, \ell}$*). Choose $\mathbf{x}_0 \in \mathcal{G}^{\circ}_{N, \ell} \subset \mathcal{S}$ and note that $\mathcal{S} \cap \text{Ball}(\mathbf{x}_0, \varepsilon; \mathbf{X}) \subset \mathcal{G}_{N, \ell}$ for a suitable $\varepsilon \in (0, 1)$. We distinguish the two cases regarding the assumptions on $\mathcal{S}$.

Case 1: $\mathcal{S}$ is convex. Note that $\mathcal{S} \subset \text{Ball}(\mathbf{0}, C; \mathbf{X})$ for a suitable $C \geq 1$, since $\mathcal{S}$ is bounded. Define $t := \frac{\varepsilon}{2C} \in (0, 1)$ and $\mathbf{x}'_0 := (1 - t)\mathbf{x}_0 \in \mathbf{X}$. With these choices, we see for arbitrary $\mathbf{x} \in \mathcal{S}$ that $\mathbf{x}'_0 + t\mathbf{x} \in \mathcal{S}$ by convexity, and furthermore

$$\big\| \mathbf{x}_0 - (\mathbf{x}'_0 + t\,\mathbf{x}) \big\|_{\mathbf{X}} = t \cdot \|\mathbf{x}_0 - \mathbf{x}\|_{\mathbf{X}} \leq 2tC = \varepsilon.$$

Thus, $\mathbf{x}'_0 + t\mathcal{S} \subset \mathcal{S} \cap \text{Ball}(\mathbf{x}_0, \varepsilon; \mathbf{X}) \subset \mathcal{G}_{N, \ell}$.

Case 2: $\mathcal{S} = \{\mathbf{x} \in \mathbf{X} : \|\mathbf{x}\|_* \leq r\}$ for some $r \in (0, \infty)$.

With $C, \kappa \geq 1$ as in Part 2 of the assumptions of the proposition, let $0 < \sigma < \frac{\varepsilon}{2\kappa C(1 + r)} < \varepsilon < 1$ and define $\mathbf{x}'_0 := (1 - \sigma)\mathbf{x}_0$, noting that $\|\mathbf{x}'_0\|_* < r$. By continuity of $\| \cdot \|_*$ with respect to itself, we can choose $0 < \delta < \frac{\varepsilon \cdot \min\{1, r\}}{2\kappa C}$ such that $\|\mathbf{x}'_0 + \mathbf{y}\|_* < r$ for all $\mathbf{y} \in \mathbf{X}$ satisfying $\|\mathbf{y}\|_* \leq \delta$. Define $t := \frac{\delta}{r}$. For arbitrary $\mathbf{y} \in \mathcal{S}$, we then have $\|t\,\mathbf{y}\|_* \leq tr = \delta$, and hence, $\mathbf{x}'_0 + t\,\mathbf{y} \in \mathcal{S}$. Furthermore, $\|(\mathbf{x}'_0 + t\,\mathbf{y}) - \mathbf{x}_0\|_{\mathbf{X}} \leq C \cdot \| - \sigma\mathbf{x}_0 + t\,\mathbf{y}\|_* \leq \kappa\sigma Cr + \kappa C\delta \leq \varepsilon$. Overall, we have shown that $\mathbf{x}'_0 + t\mathcal{S} \subset \mathcal{S} \cap \text{Ball}(\mathbf{x}_0, \varepsilon; \mathbf{X}) \subset \mathcal{G}_{N, \ell}$, as desired.

Step 2: (*We have $\mathcal{G}^{\circ}_{N, \ell} = \emptyset$ for all $N, \ell \in \mathbb{N}$*). Assume toward a contradiction that $\mathcal{G}^{\circ}_{N, \ell} \neq \emptyset$. By Step 1, there then exist $\mathbf{x}'_0 \in \mathbf{X}$ and $t > 0$ satisfying $\mathbf{x}'_0 + t\mathcal{S} \subset \mathcal{G}_{N, \ell}$.

Thus, for each $\mathbf{x} \in \mathcal{S}$, we have $\mathbf{x}'_0 + t\,\mathbf{x} \in \mathcal{G}_{N,\ell}$, and therefore $\mathrm{dist}_{\mathbf{X}}\left(\mathbf{x}'_0 + t\,\mathbf{x}, M_R\right) \leq N \cdot R^{-(s^*+\frac{1}{\ell})}$ for all $R \in \mathbb{N}$. Because of $M_R = \mathrm{range}(D_R)$, this implies that there exists $c_{\mathbf{x},R} \in \{0,1\}^R$ satisfying $\|(\mathbf{x}'_0 + t\mathbf{x}) - D_R(c_{\mathbf{x},R})\|_{\mathbf{X}} \leq N \cdot R^{-(s^*+\frac{1}{\ell})}$.

Now, we define a new codec $\widetilde{\mathcal{C}} = \left((\widetilde{E}_R, \widetilde{D}_R)\right)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathbf{X}}$ by

$$\widetilde{E}_R : \mathcal{S} \to \{0,1\}^R, \mathbf{x} \mapsto c_{\mathbf{x},R} \quad \text{and} \quad \widetilde{D}_R : \{0,1\}^R \to \mathbf{X}, c \mapsto t^{-1} \cdot \left(D_R(c) - \mathbf{x}'_0\right).$$

For arbitrary $\mathbf{x} \in \mathcal{S}$, we then see

$$\begin{aligned}
\left\|\mathbf{x} - \widetilde{D}_R\big(\widetilde{E}_R(\mathbf{x})\big)\right\|_{\mathbf{X}} &= t^{-1} \cdot \left\|t\,\mathbf{x} - \left(D_R(c_{\mathbf{x},R}) - \mathbf{x}'_0\right)\right\|_{\mathbf{X}} \\
&= t^{-1} \cdot \left\|(\mathbf{x}'_0 + t\,\mathbf{x}) - D_R(c_{\mathbf{x},R})\right\|_{\mathbf{X}} \leq \frac{N}{t} \cdot R^{-(s^*+\frac{1}{\ell})}
\end{aligned}$$

for all $R \in \mathbb{N}$. By definition of the optimal exponent, this implies $s^* = s^*_{\mathbf{X}}(\mathcal{S}) \geq s^* + \frac{1}{\ell}$, which is the desired contradiction. Hence, $\mathcal{G}^\circ_{N,\ell} = \emptyset$ for all $N, \ell \in \mathbb{N}$.

Step 3: (*Completing the proof*). It is easy to see $\bigcup_{s>s^*} \mathcal{A}^s_{\mathcal{S},\mathbf{X}}(\mathcal{C}) \subset \bigcup_{N,\ell \in \mathbb{N}} \mathcal{G}_{N,\ell}$; see Eq. (1.2). We saw in the beginning of the proof that each $\mathcal{G}_{N,\ell}$ is closed, and in Step 2 that each $\mathcal{G}_{N,\ell}$ has empty interior (in $\mathcal{S}$) and is hence nowhere dense in $\mathcal{S}$. By definition, this shows that $\bigcup_{N,\ell \in \mathbb{N}} \mathcal{G}_{N,\ell}$— and hence also $\bigcup_{s>s^*} \mathcal{A}^s_{\mathcal{S},\mathbf{X}}$—is of first category in $\mathcal{S}$.

Finally, we prove the existence of $\mathbf{x} \in \mathcal{S}$ satisfying Eq. (G.1). Assume toward a contradiction that no such $\mathbf{x}$ exists. Then for each $\mathbf{x} \in \mathcal{S}$ there exist $n_{\mathbf{x}}, \ell_{\mathbf{x}} \in \mathbb{N}$ such that for every $R \geq n_{\mathbf{x}}$, we have $\|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} < R^{-(s^*+\frac{1}{\ell_{\mathbf{x}}})}$. Thus, it is not hard to see that

$$\mathrm{dist}_{\mathbf{X}}(\mathbf{x}, M_R) \leq \|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \leq N_{\mathbf{x}} \cdot R^{-(s^*+\frac{1}{\ell_{\mathbf{x}}})} \quad \forall\, R \in \mathbb{N},$$

where we defined $N_{\mathbf{x}} := 1 + \max\left\{k^{s^*+\frac{1}{\ell_{\mathbf{x}}}} \cdot \|\mathbf{x} - D_k(E_k(\mathbf{x}))\|_{\mathbf{X}} : 1 \leq k \leq n_{\mathbf{x}}\right\}$.

Since $\mathbf{x} \in \mathcal{S}$ was arbitrary, this easily implies $\mathcal{S} = \bigcup_{N,\ell \in \mathbb{N}} \mathcal{G}_{N,\ell}$. Because $\mathcal{S} \subset \mathbf{X}$ is a closed set, and hence a complete metric space (equipped with the metric induced by $\|\cdot\|_{\mathbf{X}}$), the Baire category theorem ([14, Theorem 5.9]) shows that there are certain $N, \ell \in \mathbb{N}$ such that $\mathcal{G}^\circ_{N,\ell} \neq \emptyset$, in contradiction to Step 2. $\qquad\square$

As the second result in this appendix, we show that the preceding property does *not* hold for general compact sets $\mathcal{S} \subset \mathbf{X}$, even if $\mathbf{X} = \mathcal{H}$ is a Hilbert space. In other words, some additional regularity assumption beyond compactness—like convexity—is necessary to ensure the property stated in Proposition 3.

Example 1 We consider the Hilbert space $\mathcal{H} := \ell^2(\mathbb{N})$, where we denote the standard orthonormal basis of this space by $(\mathbf{e}_n)_{n \in \mathbb{N}}$. Fix $s > 0$, define $\mathbf{x}_0 := \mathbf{0} \in \ell^2(\mathbb{N})$ and $\mathbf{x}_n := (\log_2(n+1))^{-s} \cdot \mathbf{e}_n \in \ell^2(\mathbb{N})$ for $n \in \mathbb{N}$, and finally set

$$\mathcal{S} := \{\mathbf{x}_n : n \in \mathbb{N}_0\}.$$

We claim that $s_{\mathcal{H}}^*(\mathcal{S}) = s$, but that there is a codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}} \in \mathrm{Codecs}_{\mathcal{S},\mathcal{H}}$ such that $\mathcal{A}_{\mathcal{S},\mathcal{H}}^{\sigma}(\mathcal{C}) = \mathcal{S}$ for every $\sigma > 0$; that is, *every* element $\mathbf{x} \in \mathcal{S}$ is compressed by $\mathcal{C}$ with *arbitrary* rate $\sigma > 0$.

To prove $s_{\mathcal{H}}^*(\mathcal{S}) \leq s$, let $R \in \mathbb{N}$ and $(E, D) \in \mathrm{Enc}_{\mathcal{S},\mathcal{H}}^R$. By the pigeonhole-principle, there are $n, m \in \{1, \dots, 2^R + 1\}$ satisfying $n \neq m$ but $E(\mathbf{x}_n) = E(\mathbf{x}_m)$. By symmetry, we can assume that $n < m$, so that $n + 1 \leq 2^R + 1 \leq 2^{R+1}$. Therefore,

$$2^{-s} \cdot R^{-s} \leq (R+1)^{-s} \leq (\log_2(n+1))^{-s} \leq \|\mathbf{x}_n - \mathbf{x}_m\|_{\ell^2}$$
$$\leq \|\mathbf{x}_n - D(E(\mathbf{x}_n))\|_{\ell^2} + \|D(E(\mathbf{x}_m)) - \mathbf{x}_m\|_{\ell^2} \leq 2\,\delta_{\mathcal{S},\mathcal{H}}(E, D).$$

Since this holds for any encoder/decoder pair $(E, D) \in \mathrm{Enc}_{\mathcal{S},\mathcal{H}}^R$ and arbitrary $R \in \mathbb{N}$, we see $s_{\mathcal{H}}^*(\mathcal{S}) \leq s$.

Next, we construct the codec $\mathcal{C}$ mentioned above. To do so, for each $n \in \mathbb{N}$, fix a bijection $\kappa_n : \{0, \dots, 2^n - 1\} \to \{0, 1\}^n$ and define

$$E_n : \quad \mathcal{S} \to \{0,1\}^n, \qquad\qquad \mathbf{x}_m \mapsto \begin{cases} \kappa_n(m), & \text{if } m \leq 2^n - 1, \\ \kappa_n(0), & \text{otherwise,} \end{cases}$$

$$D_n : \quad \{0,1\}^n \to \mathcal{S}, \qquad\qquad \theta \mapsto \mathbf{x}_{\kappa_n^{-1}(\theta)}.$$

For $m \in \mathbb{N}_0$ with $m \leq 2^n - 1$, we then have $D_n(E_n(\mathbf{x}_m)) = \mathbf{x}_{\kappa_n^{-1}(\kappa_n(m))} = \mathbf{x}_m$, while if $m \geq 2^n$, then $D_n(E_n(\mathbf{x}_m)) = \mathbf{x}_{\kappa_n^{-1}(\kappa_n(0))} = \mathbf{x}_0 = \mathbf{0}$, and hence

$$\|\mathbf{x}_m - D_n(E_n(\mathbf{x}_m))\|_{\ell^2} = \|\mathbf{x}_m\|_{\ell^2} = \big(\log_2(m+1)\big)^{-s} \leq n^{-s}.$$

Therefore, $\|\mathbf{x} - D_n(E_n(\mathbf{x}))\|_{\ell^2} \leq n^{-s}$ for all $\mathbf{x} \in \mathcal{S}$, and thus $\delta_{\mathcal{S},\mathcal{H}}(E_n, D_n) \leq n^{-s}$, so that $s_{\mathcal{H}}^*(\mathcal{S}) \geq s$.

We have now proved that $s_{\mathcal{H}}^*(\mathcal{S}) = s$. Finally, it is easy to see that given *arbitrary* $\sigma > 0$, the codec $\mathcal{C} = \big((E_R, D_R)\big)_{R \in \mathbb{N}}$ constructed above approximates each *fixed* $\mathbf{x} \in \mathcal{S}$ with rate σ. Indeed, for $m \in \mathbb{N}$ and $\mathbf{x} = \mathbf{x}_m$, we have

$$\|\mathbf{x} - D_n(E_n(\mathbf{x}))\|_{\ell^2} = \begin{cases} \big(\log_2(m+1)\big)^{-s}, & \text{if } n < \log_2(m+1), \\ 0, & \text{if } n \geq \log_2(m+1) \end{cases}$$
$$\leq \big(\log_2(m+1)\big)^{\sigma} \cdot n^{-\sigma} =: C_{\mathbf{x},\sigma} \cdot n^{-\sigma}$$

for all $n \in \mathbb{N}$, while for $\mathbf{x} = \mathbf{x}_0$ we have $\|\mathbf{x} - D_n(E_n(\mathbf{x}))\|_{\ell^2} = 0$ for all $n \in \mathbb{N}$. ◄

Finally, we show (under suitable assumptions on $\mathcal{S}$) that the set of probability measures satisfying the phase transition (1.4) is meager in the set of all atom-free probability measures; this was claimed in Sect. 1.3.3.

Proposition 4 *Let $\mathbf{X}$ be a Banach space and let $\emptyset \neq \mathcal{S} \subset \mathbf{X}$ with $s^* := s_{\mathbf{X}}^*(\mathcal{S}) < \infty$. Assume that there exist $\mathbf{x}_0, \mathbf{y}_0 \in \mathbf{X}$ satisfying $\mathbf{x}_0 \neq \mathbf{y}_0$ and*

$$[\mathbf{x}_0, \mathbf{y}_0] := \big\{(1-t)\mathbf{x}_0 + t\,\mathbf{y}_0 : t \in [0,1]\big\} \subset \mathcal{S}.$$

Denote by $\mathcal{P}_{\mathrm{af}}$ the set of all atom-free Borel probability measures $\mathbb{P}$ on $\mathcal{S}$ (i.e., satisfying $\mathbb{P}(\{\mathbf{x}\}) = 0$ for all $\mathbf{x} \in \mathcal{S}$). Furthermore, denote by $\mathcal{G}$ the set of all $\mathbb{P} \in \mathcal{P}_{\mathrm{af}}$ satisfying the phase transition (1.4). Then, equipping $\mathcal{P}_{\mathrm{af}}$ with the metric d_{TV} given by $d_{TV}(\mu, \nu) := \sup_{A \subset \mathcal{S} \text{ Borel set}} |\mu(A) - \nu(A)|$, the set $\mathcal{G} \subset \mathcal{P}_{\mathrm{af}}$ is closed and has empty interior; in particular, $\mathcal{G}$ is meager in $\mathcal{P}_{\mathrm{af}}$.

Proof $\mathcal{G}$ **is closed:** Let $(\mathbb{P}_n)_{n \in \mathbb{N}} \subset \mathcal{G}$ satisfy $d_{TV}(\mathbb{P}_n, \mathbb{P}) \to 0$. Let $s > s^*$ and $\mathcal{C} \in \text{Codecs}_{\mathcal{S}, \mathbf{X}}$ be arbitrary. Then, we have $\mathbb{P}_n^*(\mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})) = 0$ for each $n \in \mathbb{N}$, so that there exists a Borel set $N_n \supset \mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$ satisfying $\mathbb{P}_n(N_n) = 0$. Define $N := \bigcap_{n=1}^{\infty} N_n$, noting that N is a Borel set satisfying $N \supset \mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$.

We then see $0 \le \mathbb{P}(N) = \lim_{m \to \infty} \mathbb{P}_m(N) \le \lim_{m \to \infty} \mathbb{P}_m(N_m) = 0$. Overall, this shows $\mathbb{P} \in \mathcal{G}$, so that $\mathcal{G}$ is closed.

$\mathcal{G}$ **has empty interior in** $\mathcal{P}_{\mathrm{af}}$: Assume toward a contradiction that there exists $\mathbb{P} \in \mathcal{G}^{\circ}$; thus, there exists $\varepsilon > 0$ such that $\widetilde{\mathbb{P}} \in \mathcal{G}$ whenever $\widetilde{\mathbb{P}} \in \mathcal{P}_{\mathrm{af}}$ with $d_{TV}(\mathbb{P}, \widetilde{\mathbb{P}}) \le \varepsilon$. Let $M := [\mathbf{x}_0, \mathbf{y}_0) = \{(1 - t)\mathbf{x}_0 + t\,\mathbf{y}_0 : t \in [0, 1)\}$. It is easy to see that

$$\mathbb{P}^{\natural}(A) := \int_0^1 \mathbb{1}_A\big(\mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0)\big)\, dt, \qquad A \in \mathscr{B}_{\mathcal{S}}$$

defines an atom-free Borel probability measure on $\mathcal{S}$ satisfying $\mathbb{P}^{\natural}(M) = 1$. This easily implies $\widetilde{\mathbb{P}} := (1 - \varepsilon)\mathbb{P} + \varepsilon\,\mathbb{P}^{\natural} \in \mathcal{P}_{\mathrm{af}}$ with $d_{TV}(\mathbb{P}, \widetilde{\mathbb{P}}) \le \varepsilon$. Therefore, by choice of $\mathbb{P}$ and ε, we get $\widetilde{\mathbb{P}} \in \mathcal{G}$. Finally, it is easy to see [4] that there exists a codec $\mathcal{C} \in \text{Codecs}_{\mathcal{S}, \mathbf{X}}$ satisfying $M \subset \bigcap_{s > 0} \mathcal{A}_{\mathcal{S}, \mathbf{X}}^s(\mathcal{C})$. By definition of $\mathcal{G}$ and since $s^* < \infty$, this implies $\widetilde{\mathbb{P}}(M) = 0$. However, $\widetilde{\mathbb{P}}(M) \ge \varepsilon\,\mathbb{P}^{\natural}(M) = \varepsilon > 0$, which is the desired contradiction. $\square$

H Technical Results Concerning Sequence Spaces

Proof of Lemma 2 Let $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$. Set $u_m := m^{\theta + \vartheta} \cdot 2^{\alpha m} \cdot \|\mathbf{x}_m\|_{\ell^p(\mathcal{I}_m)}$ and $v_m := m^{-\vartheta}$, and observe that $\|(u_m)_{m \in \mathbb{N}}\|_{\ell^q} = \|\mathbf{x}\|_{\ell_{\mathscr{P}, \alpha, \theta + \vartheta}^{p,q}}$ and $\|(v_m \cdot u_m)_{m \in \mathbb{N}}\|_{\ell^r} = \|\mathbf{x}\|_{\ell_{\mathscr{P}, \alpha, \theta}^{p,r}}$.

Let us first consider the case $q < \infty$. In this case, $\frac{q}{r} \in (1, \infty)$ and $\frac{q}{q-r} \in (1, \infty)$ are conjugate exponents, so that Hölder's inequality shows

$$\big\|(v_m \cdot u_m)_{m \in \mathbb{N}}\big\|_{\ell^r} = \big\|(v_m^r \cdot u_m^r)_{m \in \mathbb{N}}\big\|_{\ell^1}^{1/r} \le \Big(\big\|(v_m^r)_{m \in \mathbb{N}}\big\|_{\ell^{q/(q-r)}} \cdot \big\|(u_m^r)_{m \in \mathbb{N}}\big\|_{\ell^{q/r}}\Big)^{1/r}$$

$$= \big\|(v_m)_{m \in \mathbb{N}}\big\|_{\ell^{rq/(q-r)}} \cdot \big\|(u_m)_{m \in \mathbb{N}}\big\|_{\ell^q}.$$

---

[4] To prove this, fix a bijection $\iota_R : [2^R] \to \{0, 1\}^R$ and define $E_R^{(0)} : [0, 1) \to \{0, 1\}^R$ by $E_R^{(0)}(t) := \iota_R(k)$, where $k = k(t)$ is the unique $k \in [2^R]$ with $t \in \big[\frac{k-1}{2^R}, \frac{k}{2^R}\big)$. Furthermore, define $E_R : \mathcal{S} \to \{0, 1\}^R$ by setting $E_R(\mathbf{x}) := E_R^{(0)}(t)$ if $\mathbf{x} \in \mathcal{S}$ is of the form $\mathbf{x} = \mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0)$ for some $t \in [0, 1)$, and $E_R(\mathbf{x}) := E_R^{(0)}(0)$ otherwise. Finally, define $D_R : \{0, 1\}^R \to \mathcal{S}, \mathbf{c} \mapsto \mathbf{x}_0 + 2^{-R} \cdot \iota_R^{-1}(\mathbf{c}) \cdot (\mathbf{y}_0 - \mathbf{x}_0)$. It is then easy to see $\|\mathbf{x} - D_R(E_R(\mathbf{x}))\|_{\mathbf{X}} \le 2^{-R}\|\mathbf{x}_0 - \mathbf{y}_0\|_{\mathbf{X}} \lesssim_s R^{-s}$ for arbitrary $s > 0$ and all $\mathbf{x} \in M = [\mathbf{x}_0, \mathbf{y}_0)$.

Here, we note $\vartheta \cdot \frac{rq}{q-r} = \frac{\vartheta}{\frac{1}{r} - \frac{1}{q}} > 1$ and $v_m = m^{-\vartheta}$, so that $\kappa := \max\{1, \|(v_m)_{m \in \mathbb{N}}\|_{\ell^{rq/(q-r)}}\}$ is finite.

Finally, in case of $q = \infty$, simply note that

$$\big\| (v_m\, u_m)_{m \in \mathbb{N}} \big\|_{\ell^r} \leq \big\| (m^{-\vartheta})_{m \in \mathbb{N}} \big\|_{\ell^r} \cdot \big\| (u_m)_{m \in \mathbb{N}} \big\|_{\ell^q},$$

where now $\kappa := \max\big\{1, \|(m^{-\vartheta})_{m \in \mathbb{N}}\|_{\ell^r}\big\}$ is finite, since $\vartheta > \frac{1}{r} - \frac{1}{q} = \frac{1}{r}$. $\qquad \square$

Proof of Lemma 4 By definition of the product σ-algebra, each of the *finite-dimensional* projections $\pi_m : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}_m}$, $\mathbf{x} \mapsto \mathbf{x}_m$ is measurable. Since $\|\cdot\|_{\ell^p(\mathcal{I}_m)}$ is continuous on $\mathbb{R}^{\mathcal{I}_m}$ and hence Borel measurable, $q_m : \mathbb{R}^{\mathcal{I}} \to [0, \infty)$, $\mathbf{x} \mapsto 2^{\alpha m}\, m^{\theta}\, \|\mathbf{x}_m\|_{\ell^p(\mathcal{I}_m)}$ is $\mathscr{B}_{\mathcal{I}}$-measurable for each $m \in \mathbb{N}$.

In case of $q < \infty$, this implies that the map

$$\mathbb{R}^{\mathcal{I}} \to [0, \infty], \quad \mathbf{x} \mapsto \|\mathbf{x}\|_{\ell^{p,q}_{\mathscr{P},\alpha,\theta}}^q = \sum_{m=1}^{\infty} [q_m(\mathbf{x})]^q$$

is $\mathscr{B}_{\mathcal{I}}$-measurable as a countable series of measurable, non-negative functions, and hence so is $\mathbf{x} \mapsto \|\mathbf{x}\|_{\ell^{p,q}_{\mathscr{P},\alpha,\theta}}$. If $q = \infty$, the (quasi) norm $\|\cdot\|_{\ell^{p,\infty}_{\mathscr{P},\alpha,\theta}} = \sup_{m \in \mathbb{N}} q_m$ is $\mathscr{B}_{\mathcal{I}}$-measurable as a countable supremum of $\mathscr{B}_{\mathcal{I}}$-measurable, non-negative functions.

For proving the final claim, let us write $\mathcal{T} := \ell^2(\mathcal{I}) \cap \mathscr{B}_{\mathcal{I}}$ for brevity. By the first part of the lemma, $\|\cdot\|_{\ell^2} = \|\cdot\|_{\ell^{2,2}_{\mathscr{P},0,0}} : \mathbb{R}^{\mathcal{I}} \to [0, \infty]$ is $\mathscr{B}_{\mathcal{I}}$-measurable. Furthermore, for arbitrary $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$ the translation $\mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$, $\mathbf{y} \mapsto \mathbf{y} + \mathbf{x}$ is $\mathscr{B}_{\mathcal{I}}$-measurable. These two observations imply that the norm $\|\cdot\|_{\ell^2} : \ell^2(\mathcal{I}) \to [0, \infty)$ and the translation operator $\ell^2(\mathcal{I}) \to \ell^2(\mathcal{I})$, $\mathbf{y} \mapsto \mathbf{y} + \mathbf{x}$ are $\mathcal{T}$-measurable for any $\mathbf{x} \in \ell^2(\mathcal{I})$. This implies that the open ball $\{\mathbf{y} \in \ell^2(\mathcal{I}) : \|\mathbf{y} + (-\mathbf{x})\|_{\ell^2} < r\}$ is $\mathcal{T}$-measurable. But $\ell^2(\mathcal{I})$ is separable, so that every open set is a countable union of open balls; therefore, it follows that $\mathscr{B}_{\ell^2} \subset \mathcal{T}$. Conversely, $\mathcal{T}$ is generated by sets of the form $\{\mathbf{x} \in \ell^2(\mathcal{I}) : p_i(\mathbf{x}) \in M\}$, where $M \subset \mathbb{R}$ is a Borel set and $p_i : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}$, $(x_j)_{j \in \mathcal{I}} \mapsto x_i$. Since $p_i|_{\ell^2(\mathcal{I})} : \ell^2(\mathcal{I}) \to \mathbb{R}$ is continuous with respect to $\|\cdot\|_{\ell^2(\mathcal{I})}$, we see that each generating set of $\mathcal{T}$ also belongs to $\mathscr{B}_{\ell^2}$, which completes the proof. $\qquad \square$

References

1. Adams, R., Fournier, J.: Sobolev spaces, *Pure and Applied Mathematics (Amsterdam)*, vol. 140, second edn. Elsevier/Academic Press, Amsterdam (2003)
2. Alt, H.W.: Linear functional analysis. Universitext. Springer-Verlag London, Ltd., London (2016). https://doi.org/10.1007/978-1-4471-7280-2.
3. Berger, T.: Rate-distortion theory. Wiley Encyclopedia of Telecommunications (2003)
4. Bölcskei, H., Grohs, P., Kutyniok, G., Petersen, P.C.: Optimal approximation with sparsely connected deep neural networks. SIAM J. Math. Data Sci. **1**, 8–45 (2019)
5. Carl, B., Stephani, I.: Entropy, compactness and the approximation of operators, *Cambridge Tracts in Mathematics*, vol. 98. Cambridge University Press, Cambridge (1990). https://doi.org/10.1017/CBO9780511897467.
6. Conway, J.B.: A course in functional analysis, *Graduate Texts in Mathematics*, vol. 96, second edn. Springer-Verlag, New York (1990)

7. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Comm. Pure Appl. Math. **41**(7), 909–996 (1988). https://doi.org/10.1002/cpa.3160410705.
8. Daubechies, I.: Ten lectures on wavelets, *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1992). https://doi.org/10.1137/1.9781611970104.
9. DeVore, R.A.: Nonlinear approximation. In: Acta numerica, 1998, *Acta Numer.*, vol. 7, pp. 51–150. Cambridge Univ. Press, Cambridge (1998). https://doi.org/10.1017/S0962492900002816.
10. DeVore, R.A., Lorentz, G.G.: Constructive approximation, *Grundlehren der Mathematischen Wissenschaften*, vol. 303. Springer-Verlag, Berlin (1993). https://doi.org/10.1007/978-3-662-02888-9.
11. Dudley, R.M.: Real analysis and probability, *Cambridge Studies in Advanced Mathematics*, vol. 74. Cambridge University Press, Cambridge (2002). https://doi.org/10.1017/CBO9780511755347.
12. Edmunds, D.E., Triebel, H.: Function spaces, entropy numbers, differential operators, *Cambridge Tracts in Mathematics*, vol. 120. Cambridge University Press, Cambridge (1996). https://doi.org/10.1017/CBO9780511662201.
13. Elbrächter, D., Perekrestenko, D., Grohs, P., Bölcskei, H.: Deep Neural Network Approximation Theory. IEEE Transactions on Information Theory **67**(5), 2581–2623 (2021). https://doi.org/10.1109/TIT.2021.3062161
14. Folland, G.: Real analysis, second edn. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York (1999)
15. Grohs, P.: Optimally sparse data representations. In: Harmonic and Applied Analysis, pp. 199–248. Springer (2015)
16. Haroske, D.D., Schneider, C.: Besov spaces with positive smoothness on $\mathbb{R}^n$, embeddings and growth envelopes. J. Approx. Theory **161**(2), 723–747 (2009). https://doi.org/10.1016/j.jat.2008.12.004.
17. Hunt, B., Sauer, T., Yorke, J.: Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces. Bull. Amer. Math. Soc. (N.S.) **27**(2), 217–238 (1992). https://doi.org/10.1090/S0273-0979-1992-00328-2.
18. Kossaczká, M., Vybíral, J.: Entropy numbers of finite-dimensional embeddings. ArXiv preprint arXiv:1802.00572 (2018)
19. Leoni, G.: A first course in Sobolev spaces, *Graduate Studies in Mathematics*, vol. 181, second edn. American Mathematical Society, Providence, RI (2017)
20. Leopold, H.: Embeddings and entropy numbers for general weighted sequence spaces: the non-limiting case. Georgian Math. J. **7**(4), 731–743 (2000)
21. Li, W.V., Linde, W.: Approximation, metric entropy and small ball estimates for Gaussian measures. Ann. Probab. **27**(3), 1556–1578 (1999). https://doi.org/10.1214/aop/1022677459.
22. Li, W.V., Shao, Q.M.: Gaussian processes: inequalities, small ball probabilities and applications. In: Stochastic processes: theory and methods, *Handbook of Statist.*, vol. 19, pp. 533–597. North-Holland, Amsterdam (2001). https://doi.org/10.1016/S0169-7161(01)19019-X.
23. Lin, S.: Limitations of shallow nets approximation. Neural Netw. **94**, 96 – 102 (2017). https://doi.org/10.1016/j.neunet.2017.06.016. URL http://www.sciencedirect.com/science/article/pii/S0893608017301521
24. Maiorov, V., Meir, R., Ratsaby, J.: On the approximation of functional classes equipped with a uniform measure using ridge functions. J. Approx. Theory **99**(1), 95–111 (1999). https://doi.org/10.1006/jath.1998.3305.
25. Maiorov, V., Pinkus, A.: Lower bounds for approximation by MLP neural networks. Neurocomputing **25**(1), 81 – 91 (1999). https://doi.org/10.1016/S0925-2312(98)00111-8. URL http://www.sciencedirect.com/science/article/pii/S0925231298001118
26. Mallat, S.: A wavelet tour of signal processing, third edn. Elsevier/Academic Press, Amsterdam (2009)
27. Mhaskar, H.N.: Neural networks for optimal approximation of smooth and analytic functions. Neural Computation **8**(1), 164–177 (1996). https://doi.org/10.1162/neco.1996.8.1.164
28. Oxtoby, J.C.: Measure and category, *Graduate Texts in Mathematics*, vol. 2, second edn. Springer-Verlag, New York-Berlin (1980)
29. Petersen, P., Voigtlaender, F.: Optimal approximation of piecewise smooth functions using deep ReLU neural networks. Neural Netw. **108**, 296–330 (2018)
30. Rudin, W.: Functional analysis, second edn. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York (1991)
31. Safran, I., Shamir, O.: Depth-width tradeoffs in approximating natural functions with neural networks. arXiv preprint arXiv:1610.09887 (2016)

32. Safran, I., Shamir, O.: Depth-width tradeoffs in approximating natural functions with neural networks. In: International Conference on Machine Learning, pp. 2979–2987. PMLR (2017)
33. Stein, E.: Singular integrals and differentiability properties of functions. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J. (1970)
34. Suzuki, T.: Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal rate and curse of dimensionality. In: International Conference on Learning Representations (2019). https://openreview.net/forum?id=H1ebTsActm
35. Triebel, H.: Theory of function spaces. III, *Monographs in Mathematics*, vol. 100. Birkhäuser Verlag, Basel (2006)
36. Triebel, H.: Theory of function spaces. Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel (2010)
37. Vershynin, R.: High-dimensional probability, *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 47. Cambridge University Press, Cambridge (2018). https://doi.org/10.1017/9781108231596.
38. Voigtlaender, F.: Embeddings of Decomposition Spaces into Sobolev and BV Spaces. arXiv preprints arXiv:1601.02201 (2016)
39. Wojtaszczyk, P.: A mathematical introduction to wavelets, *London Mathematical Society Student Texts*, vol. 37. Cambridge University Press, Cambridge (1997). https://doi.org/10.1017/CBO9780511623790.
40. Yarotsky, D.: Elementary superexpressive activations. arXiv preprint arXiv:2102.10911 (2021)
41. Yarotsky, D., Zhevnerchuk, A.: The phase diagram of approximation rates for deep neural networks. arXiv preprint arXiv:1906.09477 (2019)