Entropy Numbers of Finite Dimensional Mixed-Norm Balls and Function Space Embeddings with Small Mixed Smoothness

Sebastian Mayer¹ · Tino Ullrich²

Received: 9 April 2019 / Revised: 28 February 2020 / Accepted: 7 April 2020 / Published online: 3 August 2020
© The Author(s) 2020, corrected publication 2021

Abstract
We study the embedding $\text{id} : \ell^b_p(\ell^d_q) \to \ell^b_r(\ell^d_u)$ and prove matching bounds for the entropy numbers $e_k(\text{id})$ provided that $0 < p < r \leq \infty$ and $0 < q \leq u \leq \infty$. Based on this finding, we establish optimal dimension-free asymptotic rates for the entropy numbers of embeddings of Besov and Triebel–Lizorkin spaces of small dominating mixed smoothness, which gives a complete answer to an open problem mentioned in the recent monograph by Dăng, Temlyakov, and Ullrich. Both results rely on a novel covering construction recently found by Edmunds and Netrusov.

Keywords Entropy numbers · Schütt’s theorem · Finite-dimensional (quasi-)normed spaces · Function spaces with small mixed smoothness · Compact embeddings.

Mathematics Subject Classification 41A46 · 42B35 · 42C40 · 46E35

1 Introduction
Entropy numbers quantify the degree of compactness of a set, i.e., how well the set can be approximated by a finite set. Given a compact set $K$ in a quasi-Banach space $Y$, the $k$-th entropy number $e_k(K, Y)$ is defined to be the smallest radius $\varepsilon > 0$ such that $K$ can be covered with $2^{k-1}$ copies of the ball $\varepsilon B_Y$, i.e.,

$$e_k(K, Y) := \inf \left\{ \varepsilon > 0 : \exists y_1, \ldots, y_{2^{k-1}} \text{ such that } K \subseteq \bigcup_{\ell=1}^{2^{k-1}} y_{\ell} + \varepsilon B_Y \right\}, \quad k \in \mathbb{N}.$$
The concept of entropy numbers can be easily extended to operators. Given a compact operator \( T : X \to Y \), where \( X \) and \( Y \) are quasi-Banach spaces, the \( k \)-th entropy number of the operator \( T \) is defined to be

\[
e_k(T : X \to Y) := e_k(T(B_X), Y).
\]

If the spaces \( X, Y \) are clear from the context, we will abbreviate \( e_k(T : X \to Y) \) by \( e_k(T) \).

Entropy numbers (or the inverse concept of metric entropy) belong to the fundamental concepts of approximation theory. They appear in various approximation problems, e.g., in the estimation of the decay of operator eigenvalues \([4, 11, 20]\); in the estimation of learning rates for machine learning problems \([39, 43]\); or in bounding \( s \)-numbers such as approximation, Gelfand, or Kolmogorov numbers from below \([4, 16]\). We note that Gelfand numbers find application in the recent field of compressive sensing \([6, 13, 16]\) and Information Based Complexity in general. Entropy numbers are also closely connected to small ball problems in probability theory \([21, 24]\). For further applications and basic properties, we refer to the monographs \([5, 28]\), and the recent survey \([8, \text{Chapter 6}]\).

The goal of this paper is to improve estimates for entropy numbers of embeddings between function spaces of dominating mixed smoothness

\[
\text{Id} : S_{p_0, q_0}^{r_0} A(\Omega) \to S_{p_1, q_1}^{r_1} A^\dagger(\Omega), \quad A, A^\dagger \in \{B, F\}, \quad (1)
\]

where \( \Omega \subset \mathbb{R}^n \) is a bounded domain, \( 0 < p_0, p_1, q_0, q_1 \leq \infty \), and \( r_0 - r_1 > (1/p_0 - 1/p_1)_+ \). The case \( A = B \) stands for the scale of Besov spaces of dominating mixed smoothness, while \( A = F \) refers to the scale of Triebel–Lizorkin spaces, which includes classical \( L_p \) and Sobolev spaces of mixed smoothness. That is why (1) also includes the classical embeddings

\[
\text{Id} : S_{p_0, q_0}^r A(\Omega) \to L_{p_1}(\Omega), \quad A \in \{B, F\},
\]

if \( r > 1/p_0 - 1/p_1 \). Function space embeddings of this type play a crucial role in hyperbolic cross approximation \([8]\). Entropy numbers of such embeddings have been the subject of intense study, see \([42]\), \([8, \text{Chapt. 6}]\), and the recent papers by A.S. Romanyuk \([33–36]\) and V.N. Temlyakov \([40]\). Note that there are a number of deep open problems connected to the case \( p_1 = \infty \), which reach out to probability and discrepancy theory \([8, 2.6, 6.4]\).

Typically, one observes asymptotic decays of the form

\[
e_m(\text{Id}) \asymp m^{-(r_0-r_1) (\log m)^{(n-1)} \eta},
\]

where \( \eta > 0 \). This behavior is also well-known for \( s \)-numbers of these embeddings such as approximation, Gelfand, or Kolmogorov numbers, see \([8]\) and the references therein. Although the main rate is the same as in the univariate case, the dimension still appears in the logarithmic term. We show that the logarithmic term completely
disappears in regimes of small smoothness

\[
1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1.
\]

That is, we establish sharp purely polynomial asymptotic bounds of the form

\[
e_m(\text{Id}) \asymp_n m^{-(r_0-r_1)}, \quad m \in \mathbb{N},
\]

which depends on the underlying dimension \( n \) only in the constant. This settles several open questions stated in the literature \([8,42]\), see Sect. 5, and makes the framework highly relevant for high-dimensional approximation.

A key ingredient in the proof of (2) is a counterpart of Schütt’s theorem for the entropy numbers of the embedding

\[
\text{id} : \ell^b_p(\ell^d_q) \rightarrow \ell^b_r(\ell^d_u),
\]

where \( 0 < p < r \leq \infty \) and \( 0 < q \leq u \leq \infty \). We prove matching bounds for all parameter constellations. A particularly relevant case for the purpose of this paper is the situation where \( b \leq d \) and

\[
1/p - 1/r > 1/q - 1/u \geq 0.
\]

Here, we have the surprising behavior

\[
e_k(\text{id}) \asymp \begin{cases} 
1 & : 1 \leq k \leq \log(bd), \\
\left( \frac{\log(ed/k)}{k} \right)^{1/q-1/u} & : \log(bd) \leq k \leq d, \\
\left( \frac{d}{k} \right)^{1/p-1/r} d^{-(1/q-1/u)} & : d \leq k \leq bd, \\
b^{-(1/p-1/r)} d^{-(1/q-1/u)} 2^{-(k-1)/p} & : k \geq bd.
\end{cases}
\]

Note that this relation is not a trivial extension of the classical Schütt result \([38]\), which reads as

\[
e_k(\text{id} : \ell^b_p \rightarrow \ell^b_r) \sim \begin{cases} 
1 : 1 \leq k \leq \log(b), \\
\left( \frac{\log(ed/k)}{k} \right)^{1/p-1/r} & : \log(b) \leq k \leq b, \\
2^{-(k-1)/p} b^{-(1/p-1/r)} & : k \geq b,
\end{cases}
\]

for the norm-1-embedding \( \text{id} : \ell^b_p \rightarrow \ell^b_r \), where \( 0 < p \leq r \leq \infty \). In fact, using trivial embeddings would give an additional log-term in the third case of (3). The absence of this log-term makes (3) interesting and useful as we will see below.

For \( 1 \leq k \leq \log(db) \) and \( k \geq bd \), it requires only trivial and standard volumetric arguments to establish matching bounds for the entropy numbers \( e_k(\text{id} : \ell^b_p(\ell^d_q) \rightarrow \ell^b_r(\ell^d_u)) \). The middle range \( \log(bd) \leq k \leq bd \) is much more involved. In general, it
is far from straightforward to generalize the proof ideas from \( d = 1 \) (Schütt) to \( d > 1 \). Fortunately, the crucial work has already been done in a recent work by Edmunds and Netrusov [10]. They prove a general abstract version of Schütt’s theorem for operators between vector-valued sequence spaces. It remains for us to turn these general, abstract bounds into explicit estimates for the entropy numbers \( e_k(\text{id} : \ell^b_p(\ell^d_q) \to \ell^b_p(\ell^d_q)) \). Unfortunately, the paper [10] is written very concisely, which makes it difficult to follow the arguments at several points. Hence, we decided to provide some additional, explanatory material. We hope that Sect. 3 helps a broader readership to appreciate the powerful ideas in [10], in particular, a novel covering construction based on dyadic grids.

**Outline** The paper is organized as follows. In Sect. 2, we recapitulate basic definitions and results including entropy numbers and Schütt’s theorem. Afterwards, in Sect. 3, we discuss the generalization of Schütt’s theorem by [10]. In Sect. 4, we show consequences of this result, including matching bounds for the entropy numbers \( e_k(\text{id} : \ell^b_p(\ell^d_q) \to \ell^b_p(\ell^d_q)) \). Finally, we improve upper bounds for the entropy numbers of Besov and Triebel–Lizorkin embeddings in regimes of small smoothness in Sect. 5.

**Notation** As usual \( \mathbb{N} \) denotes the natural numbers, \( \mathbb{N}_0 := \mathbb{N} \cup \{0\} \), \( \mathbb{Z} \) denotes the integers, \( \mathbb{R} \) the real numbers, \( \mathbb{R}_+ \) the positive real numbers, and \( \mathbb{C} \) the complex numbers. For \( a \in \mathbb{R} \) we denote \( a_+ := \max\{a, 0\} \). We write \( \log \) for the natural logarithm. \( \mathbb{R}^{m \times n} \) denotes the set of all \( m \times n \)-matrices with real entries and \( \mathbb{R}^n \) denotes the Euclidean space. Vectors are usually denoted with \( x, y \in \mathbb{R}^n \). For \( 0 < p \leq \infty \) and \( x \in \mathbb{R}^n \), we use the quasi-norm \( \|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p} \) with the usual modification in the case \( p = \infty \). If \( X \) is a (quasi-)normed space, then \( B_X \) denotes its unit ball and the (quasi-)norm of an element \( x \in X \) is denoted by \( \|x\|_X \). If \( X \) is a Banach space, then we denote its dual by \( X^* \). We will frequently use the quasi-norm constant, i.e., the smallest constant \( \alpha_X \) satisfying

\[
\|x + y\|_X \leq \alpha_X (\|x\|_X + \|y\|_X), \quad \text{for all } x, y \in X.
\]

For a given \( 0 < p \leq 1 \) we say that \( \| \cdot \|_X \) is a \( p \)-norm if

\[
\|x + y\|_X^p \leq \|x\|_X^p + \|y\|_X^p, \quad \text{for all } x, y \in X.
\]

As is well known, any quasi-normed space can be equipped with an equivalent \( p \)-norm (for a certain \( 0 < p \leq 1 \), see [2,32]). If \( T : X \to Y \) is a continuous operator we write \( T \in \mathcal{L}(X, Y) \) and \( \|T\| \) for its operator (quasi-)norm. The notation \( X \hookrightarrow Y \) indicates that the identity operator \( \text{id} : X \to Y \) is continuous. For two non-negative sequences \( (a_n)_{n=1}^{\infty}, (b_n)_{n=1}^{\infty} \subset \mathbb{R} \) we write \( a_n \lesssim b_n \) if there exists a constant \( c > 0 \) such that \( a_n \leq c b_n \) for all \( n \). We will write \( a_n \simeq b_n \) if \( a_n \lesssim b_n \) and \( b_n \lesssim a_n \). If \( \alpha \) is a set of parameters, then we write \( a_n \lesssim_{\alpha} b_n \) if there exists a constant \( c_{\alpha} > 0 \) depending only on \( \alpha \) such that \( a_n \leq c_{\alpha} b_n \) for all \( n \).
Let $b, d \in \mathbb{N}$. For $0 < p, q \leq \infty$, the $bd$-dimensional mixed space $\ell^b_p(\ell^d_q)$ is defined as the space of all matrices $x \in \mathbb{R}^{b \times d}$ equipped with the mixed (quasi-)norm

$$\|x\|_{p,q} := \left( \sum_{i=1}^{b} \left( \sum_{j=1}^{d} |x_{ij}|^q \right)^{p/q} \right)^{1/p}, \quad 0 < p < \infty, 0 < q < \infty,$$

with the usual modification that the corresponding sum is replaced by a maximum in the case that either $p = \infty$ or $q = \infty$. We always refer to the $\ell^p_p$-space supported on $[b] := \{1, \ldots, b\}$ as the outer space and to the $\ell^q_q$-space supported on $[d]$ as the inner space. For any $S \subset [b] \times [d]$ and $x \in \mathbb{R}^{b \times d}$ we define $x_S$ as the matrix $(x_S)_{ij} = x_{ij}$ for $(i, j) \in S$, $(x_S)_{ij} = 0$ for $(i, j) \in S^c$.

### 2 Entropy Numbers and Schütt’s Theorem

Let us recall basic notions and properties concerning entropy numbers. Let $K$ be a subset of a quasi-Banach space $Y$. Given $\varepsilon > 0$, an $\varepsilon$-covering is a set of points $x_1, \ldots, x_n \in K$ such that $K \subset \bigcup_{i=1}^{n} \left( x_i + \varepsilon B_Y \right)$.

An $\varepsilon$-packing is a set of points $x_1, \ldots, x_m \in K$ such that $\|x_i - x_j\|_Y > \varepsilon$ for pairwise different $i, j$. The covering number $N_\varepsilon(K, Y)$ is the smallest $n$ such that there exists an $\varepsilon$-covering of $K$, while the packing number $M_\varepsilon(K, Y)$ is the largest $m$ such that there exists an $\varepsilon$-packing of $K$. It is easy to see that

$$M_{2\varepsilon}(K, Y) \leq N_{\varepsilon}(K, Y) \leq M_{\varepsilon}(K, Y).$$

The metric entropy is defined to be

$$H_\varepsilon(K, Y) = \log_2 N_\varepsilon(K, Y);$$

see Remark 4 for the relation of metric entropy to other notions of entropy.

The $k$-th entropy number $e_k(K, Y)$ can be redefined as

$$e_k(K, Y) := \inf \{ \varepsilon > 0 : H_\varepsilon(K, Y) \leq k - 1 \}.$$

It is easy to see that the sequence of entropy numbers is decaying, i.e., $e_1 \geq e_2 \geq \cdots \geq 0$. Moreover, the set $K$ is compact in $X$ if and only if $\lim_{k \to \infty} e_k(K, Y) = 0$.

Let $T$ denote an operator mapping between two quasi-Banach spaces $X$ and $Y$. Recall from the introduction that the operator’s entropy numbers are given by

$$e_k(T : X \to Y) = e_k(T(B_X), Y), \quad k \in \mathbb{N}.$$
Clearly, we have
\[ \|T\| = e_1(T) \geq e_2(T) \geq e_3(T) \geq \cdots \geq e_k(T) \geq 0. \]

If \( T_1, T_2 \) are both operators from \( X \) to \( Y \), and \( Y \) is a \( \vartheta \)-normed space, then the entropy numbers of the sum can be estimated as follows
\[ e_{k_1+k_2-1}(T_1 + T_2)^\vartheta \leq e_{k_1}(T_1)^\vartheta + e_{k_2}(T_2)^\vartheta. \]

Moreover, if \( S \in \mathcal{L}(X, Y) \) and \( R \in \mathcal{L}(Y, Z) \) then
\[ e_{k_1+k_2-1}(R \circ S) \preceq e_k(R) e_{k_2}(S). \]

In particular, this gives
\[ e_k(R \circ S) \leq e_k(R) \|S\|. \]

For further general properties of entropy numbers and basic estimates, we refer the reader to the monographs \([5,25,29]\). For remarks on the history of entropy number research, see \([5,43]\).

In the concrete situation where \( X = \ell^b_p \) and \( Y = \ell^b_q \) for \( 0 < p \leq q \leq \infty \), the entropy numbers of the embedding \( \text{id} : \ell^b_p \to \ell^b_q \) are completely understood in terms of their decay in \( k \) and \( b \). This central result is often referred to as Schütt’s theorem. For its history and references, see Remark 3. We only state the interesting case \( 0 < p < q \leq \infty \) here.

**Theorem 1** (Schütt’s theorem) For \( 0 < p \leq q \leq \infty \) and \( k, b \in \mathbb{N} \), we have
\[ e_k(\text{id} : \ell^b_p \to \ell^b_q) \simeq \begin{cases} 1 : 1 \leq k \leq \log(b), \\ \left( \frac{\log(1+b/k)}{k} \right)^{1/p-1/q} : \log(b) \leq k \leq b, \\ 2^{-k/b}b^{1/q-1/p} : k \geq b. \end{cases} \]

The constants in the estimates do neither depend on \( k \) nor on \( b \).

**Remark 2** Note that \( e_k(\text{id} : \ell^b_\infty \to \ell^b_\infty) = 1 \) as long as \( k \leq b \) because \( \|x-y\|_\infty = 2 \) for different \( x, y \in \{-1, 1\}^b \).

**Remark 3** In 1984, Schütt [38] gave a proof for the general case of symmetric Banach spaces, which implies Theorem 1 if \( 1 \leq p \leq q \leq \infty \). In the range \( 1 \leq k \leq b \), the upper bound was first proved for all \( 0 < p \leq q \leq \infty \) by Edmunds and Triebel [11] in 1996 by covering the unit ball using suitable sparse vectors. Edmunds and Netrusov [9, Thm. 2] generalized this covering construction in 1998 to arbitrary quasi-Banach spaces. In the same paper, Edmunds and Netrusov also proved matching lower bounds for general quasi-Banach spaces [9, Thm.2]. Kühn [22] also proved the lower bound for \( e_k(\text{id} : \ell^b_p \to \ell^b_q) \) with \( 0 < p \leq q \leq \infty \) in 2001. Both [9, Thm. 2] and [22] rely on the very same idea to pack the unit ball with sparse vectors and use the fundamental combinatorial fact discussed in Remark 12 (ii) below. In 2000, Guédon and Litvak [15, Thm.6] provided an alternative proof of Theorem 1 that relies completely on interpolation arguments and improved the constants in the upper bound.
Remark 4 The concept of metric entropy for compact sets has been introduced independently by Kolmogorov [18] and Pontrjagin and Schnirelmann [31]. It should not be confused with the metric entropy of a dynamical system, which also has been introduced by Kolmogorov [19]. The latter entropy is also called Kolmogorov-Sinai entropy or measure-theoretic entropy. However, these two notions of metric entropy are related [1]. There is also a deep connection between Kolmogorov-Sinai entropy and the notions of information entropy and thermodynamic entropy [3].

3 Edmunds–Netrusov Revisited

In addition to Schütt’s theorem, the main tool that we employ in this work is a powerful result by Edmunds and Netrusov [10]. They prove a generalization of Schütt’s theorem for vector-valued sequence spaces. Let us restate the part of their result that is relevant for us.

Theorem 5 (Theorems 3.1 and 3.2 in [10]) Let $b \in \mathbb{N}$ such that $b \geq 2$, $0 < p \leq r \leq \infty$ and let $X$ and $Y$ be $\gamma$-normed quasi-Banach spaces. For $k, m \in \mathbb{N}$ such that $m \leq k$, let

$$D(m, k) = \max_{\ell \in \mathbb{N}, m \leq \ell \leq k} (\ell/k)^{1/p-1/r} e_\ell (\text{id} : X \to Y),$$

and

$$A(k, b) = \max \left\{ \|\text{id} : X \to Y\| \left( \frac{\log(eb/k)}{k} \right)^{1/p-1/r}, D(1, k) \right\}.$$

For $k \geq \log_2(b)$, we have the following.

(i) If $k \leq b$, then $e_k (\text{id} : \ell^b_p(X) \to \ell^b_r(Y)) \asymp A(k, b)$.

(ii) If $k \geq b$, then there are absolute constants $c_1, c_2$ such that $D(c_1 k/b, k) \lesssim e_k (\text{id} : \ell^b_p(X) \to \ell^b_r(Y)) \lesssim D(c_2 k/b, k)$.

Theorem 5 gives abstract lower and upper bounds that are “matching” in the sense that both have the same functional form. At first glance, this functional form is neither obvious nor easy to interpret. In addition, we found it difficult to follow the arguments in [10] at several points due to its succinct style of presentation. We thus believe that it is of value to review their key arguments and to provide some additional material that makes Theorem 5 more comprehensible. This is the subject of the remainder of this section. The reader who is only interested in applications of Theorem 5 may proceed directly to Sect. 4.
Remark 6  Theorems 3.1 and 3.2 in [10] are only stated for \(0 < p < r \leq \infty\). However, these theorems also hold true for \(p = r\). First note that in the latter case, we have

\[
D(m, k) = e_m(\text{id} : X \to Y), \quad A(k, b) = \|\text{id} : X \to Y\|.
\]

Now for \(k \geq b\), Theorem 5 has been proved in [27, Thm. 4.3]. For \(k \leq b\), the lower bound in Theorem 5 is a consequence of [27, Thm. 4.3] in combination with arguments analogous to Remark 12; the upper bound is trivial.

3.1 A Special Case to Begin with

If \(p = r = \infty\) it is clear that one simply has to take \(b\)-fold Cartesian products of the optimal covering and packing of \(B_X\) in \(Y\) to obtain the bounds

\[
\frac{1}{2} e_{k+1}(\text{id} : X \to Y) \leq e_{k}b(b(\ell^b_\infty(X) \to \ell^b_\infty(Y)) \leq e_k(\text{id} : X \to Y), \quad k \in \mathbb{N}.
\]

In any other case, simple Cartesian products will not be good enough.

The special case of equal inner spaces \(X = Y\) also allows for a rather straightforward solution if the dimension of the inner space is finite. For an easier understanding of the contribution in [10], see Theorem 5 above, we find it instructive to give a direct proof of this special case and point out its limitations. Indeed, a straightforward generalization of the well-known Edmunds–Triebel covering construction [11] based on volume arguments will do the job to establish the optimal upper bound. Recall that the essence of this covering construction is a result from best \(s\)-term approximation, sometimes referred to as Stechkin’s inequality, see [8, Sect. 7.4], which yields a \(s^{-1/p+1/r}\)-covering of \(B_{\ell^b_p}\) in \(\ell^b_r\) using only \(s\)-sparse vectors. We simply have to extend this approach to row-sparse matrices. To improve readability, we will omit some technical details in the following proof.

Proposition 7  Let \(0 < p \leq r \leq \infty\) and \(X \in \mathbb{R}^d\) (quasi-)normed with \(\|\cdot\|_X\). Further let \(b, d \in \mathbb{N}\) and \(d > 5\). Then, for \(1 \leq k \leq bd\),

\[
e_k(\ell^b_p(X) \to \ell^b_r(X)) \lesssim_{p, r} \begin{cases} 1 & : 1 \leq k \leq \max\{\log(b), d\}, \\ \left[\log(eb/k)+d\right]_k^{1/p-1/r} & : \max\{\log(b), d\} \leq k \leq bd, \\ b^{-(1/p-1/r)}2^{-(k-1)/(bd)} & : k \geq bd. \end{cases}
\]

Proof  The first case is trivial. The last case follows from volumetric arguments using the recent findings in [17, Sect. 3.2]. By these we know that

\[
\text{vol}(B_{\ell^b_p(X)})^{1/(bd)} = \frac{\Gamma(1+d/p)^{1/d}}{\Gamma(1+db/p)^{1/(db)}} \cdot \text{vol}(B_X)^{1/d} \quad (7)
\]
and similarly for \( \text{vol}(B_{\ell^d_p}(X))^{1/(bd)} \). For \( k > bd \) we use the standard volume argument to obtain

\[
e_k \precsim \left[ \frac{\text{vol}(B_{\ell^d_p}(X))}{\text{vol}(B_{\ell^d_p}(X))} \right]^{1/(db)} 2^{-(k-1)/bd} \precsim_{p,r} b^{-(1/p-1/r)} 2^{-(k-1)/(bd)}. \tag{8}
\]

For the second case let \( s \in [b] \). Clearly, we have that

\[
B_{\ell_p^s}(X) \subseteq \bigcup_{I \subseteq [b]; \#I = s} B_I,
\]

where \( B_I := \{ x \in B_{\ell^p_p}(X) : \| x_i \|_X \geq \| x_k \|_X \text{ for } i \in I, k \in [b] \setminus I \} \). When we replace the \( s \) rows with the largest \( \| \cdot \|_X \)-(quasi-)norm by \( 0 \) in \( x \in B_I \), then the resulting matrix has a \( \ell^s_r(X) \)-(quasi-)norm of at most \( s^{-(1/p-1/r)} \), which follows from a well-known relation for best \( s \)-term approximation in \( \ell^r_r \). Hence, if we wish to cover the set \( B_I \) by balls of radius \( \epsilon \simeq s^{-(1/p-1/r)} \), it suffices to take care of the \( s \) largest components of the vectors in \( B_I \). That is, we take a suitable covering of \( B_{\ell^s_p}(X) \) in \( \ell^s_r(X) \) and append \( b - s \) zero rows to every matrix of the covering. A similar volumetric argument as above in (7) and (8) tells us that

\[
e_{1+c_p,q,sd}(\text{id}: \ell^s_p(X) \rightarrow \ell^s_r(X)) \precsim_{p,r} \left[ \frac{\text{vol}(B_{\ell^s_p}(X))}{\text{vol}(B_{\ell^s_p}(X))} \right]^{1/(ds)} \precsim_{p,r} s^{-(1/p-1/r)}, \tag{10}
\]

so that we obtain a covering of \( B_I \) with cardinality \( 2^{c_p,q,sd} \).

Combining the coverings for all possible index sets \( I \) yields an \( \epsilon \)-covering \( U \) of \( B_{\ell^p_p}(X) \) in \( B_{\ell^d_p}(X) \), where \( \epsilon \simeq s^{-(1/p+1/r)} \), with cardinality

\[
\#U = \binom{b}{s} 2^{c_p,q,sd}.
\]

Now, given \( k \in [bd] \), we choose

\[
s \simeq \frac{k}{\log(eb/k) + d}
\]

such that

\[
\log(\#U) \lesssim s(\log(eb/s) + d) \leq k - 1
\]

is assured. Consequently, we obtain the upper bound

\[
e_k(\text{id}: \ell^b_p(X) \rightarrow \ell^b_r(X)) \lesssim s^{-1/p+1/r} \lesssim \left( \frac{\log(eb/k) + d}{k} \right)^{1/p-1/r}.
\]
Remark 8 One way to obtain the matching lower bound in the case \( X = Y \) is to generalize the proof idea underlying Schütt’s theorem (Theorem 1) in the case that \( \log(b) \leq k \leq b \). However, the standard combinatorial lemma is not sufficient here. A suitable packing to do this generalization has already been considered in [6, Prop. 5.3]. See also Remark 12 below.

3.2 The Covering Construction by Edmunds and Netrusov

The generalized Edmunds–Triebel covering is optimal for finite dimensional \( X = Y \), see Proposition 7 in the previous section. In the general situation, where \( X \) is compactly embedded into \( Y \), it seems that the volumetric arguments underlying (10) are too coarse to obtain sharp estimates (at least in the finite dimensional situation). The main contribution of [10] is a covering construction which resolves this shortcoming by not using volumetric arguments at all. In particular, \( X \) and \( Y \) do not have to be finite dimensional. We give a detailed recapitulation of their idea in this section. For some comments concerning the lower bound in Theorem 5, see Remark 12 at the end of this section.

The covering in [10] works in the very general situation where we are given quasi-Banach spaces \( X_1, \ldots, X_b \) and \( Y_1, \ldots, Y_b \), see Proposition 10 below. The basic idea is to cover the unit ball \( B_{\ell^p([X_i]_{i=1}^b)} \) by \( N \) cuboids

\[
U(v^j) = v_1^j B_{X_1} \times \cdots \times v_b^j B_{X_b},
\]

where \( v_1, \ldots, v^N \in \mathbb{R}_+^b \) and \( N \) is exponential in \( b \) (think of each cuboid as an anisotropically rescaled version of \( B_{\ell^\infty([X_i]_{i=1}^b)} \)). The crux is to find suitable vectors \( v^j \) such that an optimal covering can be reached by covering the cuboid \( U(v^j) \) using a product of optimal coverings of \( B_{X_1}, \ldots, B_{X_b} \). Edmunds and Netrusov [10] had the idea to consider vectors that form a *dyadic grid* derived from the simplex

\[
S(b) = \left\{ x \in [0, 1]^b : \sum_{i=1}^b x_i \leq 1 \right\}, \quad b \in \mathbb{N}.
\]

The dyadic grid is constructed with the help of the following mapping. Let

\[
v_0 : \mathbb{R}_+ \to \{ 2^k : k \in \mathbb{N}_0 \}, \quad x \mapsto 2^{\max\{0, \lceil \log_2(x) \rceil \}},
\]

and for \( x \in [0, 1]^b \), put

\[
v(x) := b^{-1}(v_0(bx_1), \ldots, v_0(bx_b)).
\]

This mapping \( v \) leads to a finite grid with the following properties.

Lemma 9 (Simplification of Lemma 2.2 in [10]) For \( b \in \mathbb{N} \), let \( \Gamma(b) = v(S(b)) \). The set \( \Gamma(b) \) has the following properties.
(i) For all \( u \in S(b) \), there is \( v \in \Gamma(b) \) such that \( u_i \leq v_i \) for all \( i \in [b] \).
(ii) For all \( v \in \Gamma(b) \), we have \( \|v\|_1 \leq 2 \).
(iii) For all \( v \in \Gamma(b) \), we have \( bv_i \in \mathbb{N} \) for each \( i \in [b] \).
(iv) We have \( \# \Gamma(b) \leq 23b \).

**Proof** Given \( x \in S(b) \), let \( v = \nu(x) \). We clearly have \( \sum_{i=1}^b v_i \leq 2 \) and \( bv_i \in \mathbb{N} \) for all indices \( i = 1, \ldots, b \). Further

\[
\# \{ i \in [b] : v_i \geq t \} \leq 2/\max_i \left\lfloor \frac{3k}{8b} \right\rfloor \leq m \leq k \left( \frac{m}{k} \right)^{1/p} - \left( \frac{m}{k} \right)^{1/r} e_m(\text{id} : X_i \rightarrow Y_i).
\]

The dyadic grid according to Lemma 9 allow us to establish the following upper bound on entropy numbers.

**Proposition 10** (Reformulation of Lemma 2.3 in [10]) Let \( X_1, \ldots, X_b \) and \( Y_1, \ldots, Y_b \) be quasi-Banach spaces, let \( 0 < p \leq r \leq \infty \), and let \( k \in \mathbb{N} \) such that \( k \geq 8b \). Then, we have

\[
e_{k+1}(\text{id} : \ell_p([X_i]_{i=1}^b) \rightarrow \ell_r([Y_i]_{i=1}^b)) \leq 2^{1/r} 8^{1/p-1/r} \max_{i \in [b]} \max_{\left\lfloor \frac{3k}{8b} \right\rfloor \leq m \leq k} (m/k)^{1/p-1/r} \left( \frac{m}{k} \right)^{1/p} - \left( \frac{m}{k} \right)^{1/r} e_m(\text{id} : X_i \rightarrow Y_i).
\]

**Proof** Consider the transformed grid

\[
\Gamma(b, p) = \{ (v_1^{1/p}, \ldots, v_b^{1/p}) : v \in \Gamma(b) \}.
\]

By Lemma 9 (i), we have

\[
B_{\ell_p([X_i]_{i=1}^b)} \subset \bigcup_{v \in \Gamma(b, p)} U(v),
\]

where \( U(v) \) is the cuboid defined in (11).
Let $v \in \Gamma(b, p)$ be given by $v = (v_1^{1/p}, \ldots, v_b^{1/p})$. For each $m_i = \lfloor 1/2(k/b - 2) \rfloor b v_i$, $i \in [b]$, let $C_i$ be a $e_{m_i}(v_i^{1/p} B_{X_i}, Y_i)$-covering. Then, for every $x \in U(v)$, there is $y \in \ell^b_r(Y)$ such that $y_i \in C_i$ and

$$\|x - y\|_{\ell^b_r(Y)} \leq \left( \sum_{i=1}^b v_i^{r/ p} e_{m_i}(B_{X_i}, Y_i)^r \right)^{1/r} \leq \left( \sum_{i=1}^b v_i \right)^{1/r} \max_{i=1, \ldots, b} \max_{j=1, \ldots, b} v_j^{1/p - 1/r} e_{m_j}(B_{X_i}, Y_i).$$

By construction of the set $\Gamma(b)$, we have $\left( \sum_{i=1}^b v_i \right)^{1/r} \leq 2^{1/r}$ and

$$\max_{j=1, \ldots, b} v_j^{1/p - 1/r} e_{m_j}(B_{X_i}, Y_i) \leq 8^{1/p - 1/r} \max_{m=1, \ldots, k} (m/k)^{1/p - 1/r} e_m(B_{X_i}, Y_i).$$

Finally, note that the product $C_1 \times \cdots \times C_s$ has cardinality

$$\prod_{i=1}^s 2^{m_i - 1} \leq 2^{k - 3b},$$

which, in combination with $\sharp \Gamma(b, p) \leq 2^{3b}$, implies the desired result. \qed

Proposition 10 is not the complete final answer. For $k \leq b$, we have to modify the proof of Proposition 7. We sketch the proof and refer to the proof of [10, Thm 3.1] for technical details.

**Proposition 11** Let $\log_2(b) \leq k \leq b$. Then, we have

$$e_k(\text{id} : \ell^b_p(X) \to \ell^b_r(Y)) \preceq A(k, b),$$

where $A(k, b)$ is defined in Theorem 5.

**Proof** (Proof sketch) Let $s \in [k]$. It is clear that, analogously to (9), we have

$$B_{\ell^b_p(X)} \subseteq \bigcup_{I \subseteq [b]; \sharp I = s} B_I.$$

Similar to Proposition 7, we can use a covering for $B_{\ell^b_p(X)}$ to construct a covering for $B_I$. Consider now $\varepsilon = e_k(B_{\ell^b_p(X)}, \ell^b_r(Y))$ and let $\Gamma_0$ be a minimal $\varepsilon$-covering of
Let $\Gamma_I = \Gamma_0 \times (0, b]^{[I]}$. Then, for every $x \in B_I$, there is $y \in \Gamma_I$ such that

$$
\|x - y\|_{\ell_r(Y)} \lesssim r, \quad p 
$$

where the second term on the right-hand side follows from the best $s$-term approximation result already used in Proposition 7. Consequently, we have

$$
\|x - y\|_{\ell_r(Y)} \lesssim r, \quad p \max \{e_k(B_{\ell_p}(X), \ell_r(Y)), s^{1/r-1/p}\|\text{id} : X \to Y\|, \max_{m \in [k]} (m/k)^{1/p-1/r} e_m(B_X, Y)\}.
$$

(12)

In contrast to Proposition 7, volumetric arguments would now give a suboptimal estimate for the entropy numbers $e_k(B_{\ell_p}(X), \ell_r(Y))$. In this general situation, it requires Proposition 10 with $X_1 = \cdots = X_b = X$ and $Y_1 = \cdots = Y_b = Y$ to get the proper estimate. Concretely, since $s \leq k$, we have

$$
e_k(B_{\ell_p}(X), \ell_r(Y)) \leq e_k(B_{\ell_p}(X), \ell_r(Y)),
$$

which leads in combination with Proposition 10 and (12) to an upper bound of the form

$$
e_k(B_{\ell_p}(X), \ell_r(Y)) \lesssim \max_{m \in [k]} (m/k)^{1/p-1/r} e_m(B_X, Y).
$$

The usual arguments show that it is optimal to choose $s$ of the order $k/\log(ek/b)$. "

Remark 12 We close this section with some remarks concerning the lower bound in Theorem 5. Its proof relies on two surprisingly simple observations, see [10] for details.

(i) Let $M$ be a maximal $\varepsilon$-packing of $B_X$ in $Y$. Using the Gilbert–Varshamov bound, which is well-known in coding theory [14,41], we know that $(2s)^{-1/p}M^2 \subset B_{\ell_p}(X)$ contains $N$ elements of mutual distance $s^{1/r-1/p}\varepsilon$, where $N \simeq \text{card}(M)^s$. This leads to the lower bound

$$
e_m(B_{\ell_p}(X), \ell_r(Y)) \gtrsim s^{1/r-1/p} e_{4m+6}(B_X, Y),
$$

see [27, p. 68] and [10, Lem.2.6] for a more general formulation. Given $k \in \mathbb{N}$, we have to make a good choice for the dimension $s$ to maximize the lower bound. Choose $s = k/m$ for some $m \in [k]$ to obtain

$$
e_k(B_{\ell_p}(X), \ell_r(Y)) \gtrsim (m/k)^{1/p-1/r} e_{4m+6}(B_X, Y).
$$

If $k \leq b$, we conclude

$$
e_k(B_{\ell_p}(X), \ell_r(Y)) \gtrsim \max_{m \in [k]} (m/k)^{1/p-1/r} e_{4m+6}(B_X, Y).
$$
If \( k \geq b \), then \( m \geq k/b \) guarantees \( s = k/m \leq b \) and thus

\[
e_k(B_{\ell_p^b(X)}, \ell_r^b(Y)) \gtrsim \max_{k/b \leq m \leq k} (m/k)^{1/p-1/r} e_{4m+6}(B_X, Y).
\]

(ii) Choose a vector \( x \in B_X \) such that

\[
\|x\|_Y \geq \frac{1}{2}\|\text{id} : X \rightarrow Y\|.
\]

We construct a packing by building row-sparse matrices, where the nonzero rows contain copies of \( x \) and the row support sets are chosen according to the following combinatorial fact that is well-known in various disciplines of mathematics, see e.g., [13, Lemma 10.12], [22], [12], or [30, Prop. 2.21, p. 219]. Given \( s, n \in \mathbb{N} \) such that \( 0 < s < n/2 \), there exist subsets \( I_1, \ldots, I_N \) of \([n]\), where

\[
N \geq \left( \frac{n}{8s} \right)^s,
\]

such that each subset \( I_i \) has cardinality \( 2s \) and

\[
\sharp(I_i \cap I_j) < s \quad \text{whenever } i \neq j.
\]

This leads to the lower bound

\[
e_k(B_{\ell_p^b(X)}, \ell_r^b(Y)) \gtrsim \|\text{id} : X \rightarrow Y\| \left( \frac{\log(eb/k)}{k} \right)^{1/p-1/r}.
\]

In view of the packing construction that we have mentioned in Remark 8 it is somewhat surprising that it is not necessary to combine the combinatorics of the two observations in order to obtain the optimal abstract bound in Theorem 5. An explanation is given in [27, Rem. 4.13, p. 69].

### 4 Consequences of the Edmunds–Netrusov Result

We discuss some consequences of Theorem 5. Let us begin with considering the entropy numbers

\[
e_k := e_k(\text{id} : \ell_p^b(\ell_q^d) \rightarrow \ell_r^b(\ell_u^d)), \quad 0 < p \leq r \leq \infty, \, 0 < q \leq u \leq \infty.
\]

We have the following matching bounds.

**Theorem 13** Let \( 0 < p \leq r \leq \infty \) and \( 0 < q \leq u \leq \infty \). Then, we have

\[
e_k \simeq \begin{cases} 
1 & : 1 \leq k \leq \log(bd), \\
b^{-(1/p-1/r)}d^{-(1/q-1/u)}2^{-\frac{k-1}{bd}} & : k \geq bd.
\end{cases}
\]
For \( \log(bd) \leq k \leq bd \), we have the following case distinctions.

(i) Let \( 1/p - 1/r > 1/q - 1/u \geq 0 \).

(i.a) In the special case \( q = u \), we have

\[
e_k \simeq \begin{cases} 
1 & : \log(bd) \leq k \leq d, \\
\left( \log(eb/k) \right)^{1/p-1/r} & : d \leq k \leq bd.
\end{cases}
\]

(i.b) If \( q < u \) and \( b \leq d \), then

\[
e_k \simeq \begin{cases} 
\left( \log(ek/k) \right)^{1/q-1/u} & : \log(bd) \leq k \leq d, \\
\left( \frac{d}{k} \right)^{1/p-1/r} d^{1/u-1/q} & : d \leq k \leq bd.
\end{cases}
\]

(i.c) If \( q < u \) and \( d \leq b \), then

\[
e_k \simeq \max \left\{ \left( \log(ek/k) \right)^{1/p-1/r}, \left( \frac{d}{k} \right)^{1/p-1/r} d^{1/u-1/q} \right\} : b \leq k \leq bd.
\]

(ii) Let \( 1/q - 1/u \geq 1/p - 1/r \geq 0 \). Then, we have

\[
e_k \simeq \frac{1}{(l/k)^{1/p-1/r}} \epsilon_l \left( \epsilon_q^d \rightarrow \epsilon_u^d \right).
\]

Proof For \( 1 \leq k \leq \log(bd) \) and \( k \geq bd \), it requires only standard volumetric arguments, see [27, Appendix A] for details. Let us also refer to [7, Lemma 3], where this case has been already considered. Let \( D(m, k) \) and \( A(k, b) \) be as defined in Theorem 5. Moreover, throughout the proof, we write for \( k, l \in \mathbb{N} \),

\[
s_{k,l} := (l/k)^{1/p-1/r} \epsilon_l \left( \epsilon_q^d \rightarrow \epsilon_u^d \right).
\]

Ad (i.a). Since \( q = u \), it follows from Theorem 1 that \( \epsilon_l \left( \epsilon_q^d \rightarrow \epsilon_u^d \right) \simeq 1 \) for \( 1 \leq l \leq d \) and consequently that \( D(1, k) = D(k/b, k) \simeq 1 \) and \( A(k, b) \simeq 1 \) for all \( k \leq d \). Now, for \( k \geq d \), we have that \( s_{k,l} \simeq (l/k)^{1/p-1/q} \) for \( 1 \leq l \leq d \), so the sequence is bounded from above by a monotonically increasing sequence. For \( d \leq l \leq k \), we have

\[
s_{k,l} \simeq (l/k)^{1/p-1/r} 2^{-l/d} := t_{k,l}.
\]

Since \( 2^{-l/d} \) decays faster in \( l \) than \((l/k)^{1/p-1/r} \) increases, we conclude that for \( d \leq l \leq k \), the sequence \( s_{k,l} \) is “essentially monotonically decreasing”. To be more precise \( t_{k,l} \)
attains at \( l = \beta_{p,r}d \) its maximum, where the factor \( \beta \) depends only on \( p \) and \( r \). Hence, the maximum of \( s_{k,l} \) can be bounded from above by a constant times the maximum

of \( t_{k,l} \) and therefore by \( c_{p,r}(d/k)^{1/p-1/r} \). Using analogous arguments for \( D(k/b, k) \),

we conclude that \( \tilde{D}(1, k) = D(k/b, k) \simeq (d/k)^{1/p-1/r} \) and

\[
A(k, b) \simeq \max \left\{ \left( \frac{\log(eb/k)}{k} \right)^{1/p-1/r}, \left( \frac{d}{k} \right)^{1/p-1/r} \right\} \simeq \max \left\{ \frac{\log(eb/k)}{k}, \frac{d}{k} \right\}^{1/p-1/r}
\]

for \( d \leq k \leq b \).

Ad \((i.b)\). Consider now \( 0 < q < u \) and \( b \leq d \). For \( \log(bd) \leq k \leq b \), we have, in consequence of Theorem 1, that \( s_{k,l} \simeq (l/k)^{1/p-1/r} \) for \( 1 \leq l \leq \log(d) \) and

\[
s_{k,l} \simeq (l/k)^{1/p-1/r} \left( \frac{\log(ed/l)}{l} \right)^{1/q-1/u}.
\]

Since \( 1/p-1/r > 1/q-1/u \), the sequence \( s_{k,l} \) is bounded from above and below up to a constant by a monotonically increasing sequence and consequently, the maximum is attained at \( l = k \) such that \( D(1, k) \simeq (\log(ed/k)/k)^{1/q-1/u} \). Since \( b \leq d \), we further have

\[
D(1, k) \leq A(k, b) \lesssim D(1, k).
\]

For \( b \leq k \leq d \) we find as before that \( D(k/b, k) \simeq (\log(ed/k)/k)^{1/q-1/u} \) and for \( d \leq k \leq bd \), we have the estimate

\[
D(k/b, k) \simeq \left( \frac{d}{k} \right)^{1/p-1/r} d^{1/u-1/q}.
\]

Ad \((i.c)\). Consider now \( d \leq b \). For \( \log(bd) \leq k \leq d \), we find \( D(1, k) \simeq (\log(ed/k)/k)^{1/q-1/u} \) since the sequence \( s_{k,l} \) is bounded from below and above by a sequence that increases monotonically in \( l \). If \( d \leq k \leq b \), then

\[
D(1, k) \simeq \left( \frac{d}{k} \right)^{1/p-1/r} d^{1/u-1/q}
\]

and

\[
A(k, b) \simeq \max \left\{ \left( \frac{\log(eb/k)}{k} \right)^{1/p-1/r}, \left( \frac{d}{k} \right)^{1/p-1/r} d^{1/u-1/q} \right\}.
\]

Finally, if \( b \leq k \leq bd \), then \( D(k/b, k) \simeq (d/k)^{1/p-1/r} d^{1/u-1/q} \).

Ad \((ii)\). For \( \log(bd) \leq k \leq b \), we observe that

\[
D(1, k) \simeq \left( \frac{\log(d)}{k} \right)^{1/p-1/r}
\]

\( \diamond \) Springer
since the term \( e^\ell_B \), \( \ell^d \) is decaying in \( \ell \) at least as fast as \((\ell/k)^{1/p-1/r}\) is growing. Hence,

\[
A(k, b) \approx \max \left\{ \left( \log\left( \frac{eb}{k} \right) \right)^{1/p-1/r}, \left( \log\left( \frac{d}{k} \right) \right)^{1/p-1/r} \right\}
\]

\[
\approx \left( \frac{\log(ebd/k)}{k} \right)^{1/p-1/r}.
\]

Next, we consider \( b \leq k \leq b \log(d) \). Since \( k/b \leq \log(d) \), we find

\[
D(k/b, k) \approx \left( \frac{\log(ebd/k)}{k} \right)^{1/p-1/r},
\]

where we have used \( b/k \leq 1 \) in the last estimate. At the same time, since \( k/b \leq \log(d) \), we also have \( \log(bd/k) \gtrsim \log(d) \) and thus

\[
D(k/b, k) \approx \left( \frac{\log(ebd/k)}{k} \right)^{1/p-1/r}.
\]

Finally, for \( b \log(d) \leq k \leq bd \) it is easy to see that

\[
D(k/b, k) \approx b^{1/p-1/r} \left( \frac{b \log(ebd/k)}{k} \right)^{1/q-1/u}.
\]

\[\square\]

**Remark 14** The upper bound for \( k \geq bd \) in Theorem 13 also follows from [7, Lem. 3]. The upper bound in Theorem 13 (ii) has also been proved in [42, Lem 3.16] for the range \( b \max\{\log(d), \log(b)\} \leq k \leq bd \). The proof there uses the following covering construction, which as far as we know, first appeared in [23, Proof of Prop. 4]. Let \( X_1, \ldots, X_b \) and \( Y_1, \ldots, Y_b \) be (quasi-)Banach spaces and \( 0 < p, r \leq \infty \). The covering rests on the idea to split the ball \( B_{\ell^p}(X_1, \ldots, X_b) \) into subsets of matrices with non-increasing rows,

\[
B_{\ell_p}^{b}(X_1, \ldots, X_b) \subseteq \bigcup_{\pi} \{ x \in B_{\ell_p}^{b}(X_1, \ldots, X_b) : \|x_\pi(1)\| \geq \cdots \geq \|x_\pi(b)\| \},
\]

where the union is taken over all permutations of \([b]\). This leads to the upper bound

\[
e^{-\sum_{j=1}^{b} n_j + b \log_2(b)} \left( \sum_{j=1}^{b} j^{-r/p} e_{n_j}^{\ell_p} (B_{X_j}, Y_j)^r \right)^{1/r} \leq \left( \sum_{j=1}^{b} j^{-r/p} e_{n_j}^{\ell_p} (B_{X_j}, Y_j)^r \right)^{1/r}
\]

for \( n_1, \ldots, n_b \in \mathbb{N} \). If

\[
X = X_1 = \cdots = X_b = \ell^d_q \quad \text{and} \quad Y = Y_1 = \cdots = Y_b = \ell^d_u
\]
with $0 < q \leq u$, and we choose $n_j \simeq j^{-\alpha}$ for some $0 < \alpha < 1$ such that

$$\alpha(1/q - 1/u) > 1/p - 1/r,$$

then (13) is strong enough to obtain the upper bound in Theorem 13 (ii), provided

$$b \max \{\log(d), \log(b)\} \leq k \leq bd.$$

Now we increase the level of abstraction and consider mixed norms of higher order. Let, for $\mu = 1, \ldots, b$, the weighted spaces $X_\mu$ and $Y_\mu$ be given by

\begin{align*}
X_\mu &= \ell^b_p \left( d_{\mu}^\alpha \ell_q^d \right), \\
Y_\mu &= \ell^b_r \left( d_{\mu}^\beta \ell_u^d \right),
\end{align*}

(14)

with $0 < p \leq r \leq \infty$, $0 < q \leq u \leq \infty$, and $\alpha, \beta \in \mathbb{R}$. The dimensions $(d_{\mu})_{\mu}$ and $(b_{\mu})_{\mu}$ are non-decreasing natural numbers satisfying $d_{\mu} \gtrsim b_{\mu}$. These spaces are used as “inner spaces” in the way that

$$X = \ell^b_p \left( (X_\mu)_{\mu=1}^{b} \right) \text{ and } Y = \ell^b_r \left( (Y_\mu)_{\mu=1}^{b} \right).$$

Note that for $x = (x_{\mu,i,j})_{\mu,i,j} \in X$ with $\mu = 1, \ldots, b, i = 1, \ldots, b_{\mu}, j = 1, \ldots, d_{\mu}$, the norm is given by

$$\|x\|_X = \left\{ \sum_{\mu=1}^{b} \sum_{i=1}^{b_{\mu}} d_{\mu}^\alpha \left( \sum_{j=1}^{d_{\mu}} |x_{\mu,i,j}|^q \right)^{p/q} \right\}^{1/p}.$$

We are interested in the behavior of the entropy numbers

$$e_k(\text{id} : X \to Y) = e_k(\text{id} : \ell^b_p(X_\mu) \to \ell^b_r(Y_\mu))$$

in the special situation $1/q - 1/u < 1/p - 1/r$.

**Proposition 15** Let $0 \leq 1/q - 1/u < 1/p - 1/r$ and $\alpha - \beta \leq 1/p - 1/r - (1/q - 1/u)$. Let further $X, Y$ and $X_\mu, Y_\mu$ be as above. Then we have for all $k \geq 8b$ and $k \geq \max_{\mu=1,\ldots,b} d_{\mu}$

$$e_k(\text{id} : \ell^b_p(X_\mu) \to \ell_r(Y_\mu)) \lesssim \left( \frac{1}{k} \right)^{\alpha - \beta + 1/q - 1/u}.$$

**Proof** We use Theorem 5, in particular the upper bound in Proposition 10. Since $k \geq 8b$ we obtain

$$e_k(\text{id} : \ell^b_p(X_\mu) \to \ell_r(Y_\mu)) \lesssim \max_{\mu=1,\ldots,b} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p - 1/r} e_\ell(\text{id} : X_\mu \to Y_\mu)$$
\[
= \max_{\mu=1,\ldots,b} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{1-(\alpha-\beta)} \mu e_\ell (\text{id} : \ell_p^{b_\mu} (\ell_q^{d_\mu})) \\
= \ell_r^{b_\mu} (\ell_u^{d_\mu})) \\
\lesssim \max_{1 \leq \ell \leq d_\mu} \left\{ \max_{1 \leq \ell \leq d_\mu} \left[ \cdots \right], \max_{d_\mu \leq \ell \leq k} \left[ \cdots \right] \right\}.
\]

Let us evaluate the first \( \max[\cdots] \). With Theorem 13, (i.b), (i.c) we have

\[
\max_{1 \leq \ell \leq d_\mu} \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{1-(\alpha-\beta)} \mu e_\ell (\text{id} : \ell_p^{b_\mu} (\ell_q^{d_\mu})) \rightarrow \ell_r^{b_\mu} (\ell_u^{d_\mu})) \\
\lesssim \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{1-(\alpha-\beta)} \mu \left( \log(e_{d_\mu}/\ell) \right)^{1/q-1/u}.
\]

Because \( 1/p - 1/r > 1/q - 1/u \) the maximum is attained for \( \ell = d_\mu \), which leads to

\[
\max_{1 \leq \ell \leq d_\mu} \left[ \cdots \right] \lesssim d^{1-((1-q-1)/u)} \mu d^{1-((1-q-1)/u)} \mu.
\]

Let us discuss the second \( \max[\cdots] \). Using again Proposition 10 we obtain

\[
\max_{d_\mu \leq \ell \leq k} \left[ \cdots \right] \lesssim k^{-(1/p-1/r)} d^{1-(\alpha-\beta)} \mu d^{1-((1-q-1)/u)} \mu.
\]

Due to our assumption the exponent for \( d_\mu \) is positive in both cases. Since \( k \geq d_\mu \) we may replace \( d_\mu \) by \( k \) to increase the right-hand side. This leads to

\[
e_k (\text{id} : \ell_p^{b_\mu} (X_\mu) \rightarrow \ell_r (Y_\mu)) \lesssim \left( \frac{1}{k} \right)^{\alpha-\beta+1/q-1/u}.
\]

\( \square \)

We are now aiming for a similar relation for small \( k \).

**Proposition 16** Let \( \alpha - \beta > 0 \) and \( 1/p - 1/r > 1/q - 1/u \geq 0 \). Then we have for \( 8b \leq k \leq \min_{\mu=1,\ldots,b} d_\mu \) the estimate

\[
e_k (\text{id} : \ell_p^{b_\mu} (X_\mu) \rightarrow \ell_r (Y_\mu)) \lesssim \left( \frac{1}{k} \right)^{\alpha-\beta+1/q-1/u}.
\]
Proof Again we use Theorem 5, in particular the upper bound in Proposition 10. This gives

\[ e_k \lesssim \max_{\mu} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p-1/r} e_\ell (\text{id} : X_\mu \to Y_\mu) \]

\[ = \max_{\mu} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{-(\alpha-\beta)} e_\ell (\text{id} : X^\mu_\ell \to Y^\mu_\ell) \]

\[ \lesssim \max_{\mu} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{-(\alpha-\beta)} \left( \frac{\log (ed_\mu/\ell)}{\ell} \right)^{1/q-1/u}, \]

where we used once again Theorem 13, (i.b). Clearly, we get

\[ e_k \lesssim \max_{\mu} \max_{1 \leq \ell \leq k} \left( \frac{\ell}{k} \right)^{1/p-1/r} d^{-(\alpha-\beta)} \left( \frac{\log (ed_\mu/\ell)}{\ell} \right)^{1/q-1/u} \]

\[ \lesssim k^{-(\alpha-\beta)} k^{-(1/q-1/u)} \left( \frac{k}{d_\mu} \right)^{\alpha-\beta} \left( \log (ed_\mu/k) \right)^{1/q-1/u}. \]

Since the function \( x \mapsto x^{-(\alpha-\beta)} \left[ \log (ex) \right]^{(1/q-1/u)} \) is bounded on \([1, \infty)\) we conclude with

\[ e_k \lesssim k^{-(\alpha-\beta+1/q-1/u)}. \]

\[ \square \]

5 Polynomial Decay of Entropy Numbers for Multivariate Function Space Embeddings

We come to the main subject of this paper, improved upper bounds for entropy numbers of function space embeddings (1) in regimes of small mixed smoothness.

5.1 Function Spaces of Dominating Mixed Smoothness

Besov and Triebel–Lizorkin spaces of mixed smoothness are typically defined via a dyadic decomposition on the Fourier side. Let \( \{ \varphi_j \}_{j \in \mathbb{N}_0^n} \) be the standard tensorized dyadic decomposition of unity, see [37] and [42]. We further denote by \( S^r(\mathbb{R}^n) \) the space of tempered distributions and by \( D^r(\Omega) \) the space of distributions (dual space of \( D(\Omega) \), which represents the space of test functions on the bounded domain \( \Omega \subset \mathbb{R}^n \)). The Besov space of dominating mixed smoothness \( S^r_{p,q} B(\mathbb{R}^n) \) with smoothness parameter \( r > 0 \) and integrability parameters \( 0 < p, q \leq \infty \) is given by

\[ S^r_{p,q} B(\mathbb{R}^n) := \left\{ f \in S^r(\mathbb{R}^n) : \| f \|_{S^r_{p,q} B} := \left( \sum_{j \in \mathbb{N}_0^n} 2^{rq j \| \cdot \|_1} \| \mathcal{F}^{-1} \{ \varphi_j \mathcal{F} f \} \|_p^q \right)^{1/q} < \infty \right\}. \]
with the usual modification in the case \( q = \infty \). The Triebel–Lizorkin space of dominating mixed smoothness \( S_{p,q}^r F(\mathbb{R}^n) \) is given by \((p < \infty)\)

\[
S_{p,q}^r F(\mathbb{R}^n) := \left\{ f \in S'(\mathbb{R}^n) : \|f\|_{S_{p,q}^r F} := \left\| \left( \sum_{j \in \mathbb{N}^n_0} 2^{rq|j|1}|F^{-1}[\varphi_j F f](\cdot)|^q \right)^{1/q} \right\|_p < \infty \right\}.
\]

The latter scale of spaces contains the classical \( L_p \) spaces and Sobolev spaces with dominating mixed smoothness if \( 1 < p < \infty \) and \( q = 2 \), namely we have \( S_{0,2}^0 F(\mathbb{R}^n) = L_p(\mathbb{R}^n) \) and \( S_{p,2}^k F(\mathbb{R}^n) = S_p^k W(\mathbb{R}^n) \) for \( k \in \mathbb{N} \). Note that we also have \( S_{p,p}^r B(\mathbb{R}^n) = S_{p,p}^r F(\mathbb{R}^n) \) for all \( 0 < p < \infty \) and \( r \in \mathbb{R} \). Though we have the embedding

\[
S_{p_0,q_0}^r A(\Omega) \hookrightarrow S_{p_1,q_1}^0 \hat{A}(\mathbb{R}^n), \quad A, \hat{A} \in \{ B, F \},
\]

for \( p_0 \leq p_1 \) and \( r_0 - r_1 > 1/p_0 - 1/p_1 \), see [37, Chapt. 2], the embedding (16) is never compact. Hence, the entropy numbers of embeddings between function spaces defined on the whole \( \mathbb{R}^n \) do not converge to zero. We restrict our considerations to spaces on bounded domains \( \Omega \). Let \( \Omega \) be an arbitrary bounded domain in \( \mathbb{R}^n \). Then, we define \( S_{p,q}^r A(\Omega) \) for \( A \in \{ B, F \} \) as

\[
S_{p,q}^r A(\Omega) := \{ f \in D'(\Omega) : \exists g \in S_{p,q}^r A(\mathbb{R}^n) \text{ such that } g|_{\Omega} = f \}
\]

and its (quasi-)norm is given by \( \|f\|_{S_{p,q}^r A(\Omega)} := \inf_{g|_{\Omega} = f} \|g\|_{S_{p,q}^r A} \). The embedding (16) transfers to the bounded domain \( \Omega \) and is compact such that the entropy numbers decay and converge to zero.

### 5.2 Sequence Spaces

The key to establishing the decay rate of entropy numbers for the embedding (1) is a discretization technique which has been developed over the years by several authors beginning with Maiorov [26]. Later, after wavelet isomorphisms had been established, this technique was refined by Lemarie, Meyer, Triebel, and many others. In [42, Thm. 2.10] Vybřal gave the necessary modifications to deal with the above defined \( S_{p,q}^r A(\Omega) \) spaces in detail. The main advantage of this approach is to transfer questions for function space embeddings to certain sequence spaces.

Using sufficiently smooth wavelets with sufficiently many vanishing moments (and the notation from [42]) the mapping

\[
f \mapsto \lambda_{j,m}(f) := \langle f, \psi_{j,m} \rangle, \quad j \in \mathbb{N}^n_0, m \in \mathbb{Z}^n,
\]
represents a sequence spaces isomorphism between $S_{p,q}^r B(\mathbb{R}^n)$, $S_{p,q}^r F(\mathbb{R}^n)$ and

$$s_{p,q}^r b := \left\{ \lambda = (\lambda_j, m) \in \mathbb{N}_0^n : \|\lambda\|_{s_{p,q}^r b} < \infty \right\},$$

$$s_{p,q}^r f := \left\{ \lambda = (\lambda_j, m) \in \mathbb{N}_0^n : \|\lambda\|_{s_{p,q}^r f} < \infty \right\},$$

respectively, with the usual modification in the case $\max\{p, q\} = \infty$. Here we denote for $j \in \mathbb{N}_0^n$ and $m \in \mathbb{Z}^n$

$$Q_{j,m} := \prod_{i=1}^n 2^{-j_i} [m_i - 1, m_i + 1]$$

and

$$A_j^\Omega = \{ m \in \mathbb{Z}^n : Q_{j,m} \cap \Omega \neq \emptyset \}.$$

Further, $\chi_{j,m}$ denotes the characteristic function of $Q_{j,m}$. Consider the sequence spaces

$$s_{p,q}^r b(\Omega) := \left\{ \lambda = (\lambda_j, m) \in \mathbb{N}_0^n, m \in A_j^\Omega : \|\lambda\|_{s_{p,q}^r b(\Omega)} < \infty \right\},$$

$$s_{p,q}^r f(\Omega) := \left\{ \lambda = (\lambda_j, m) \in \mathbb{N}_0^n, m \in A_j^\Omega : \|\lambda\|_{s_{p,q}^r f(\Omega)} < \infty \right\}$$

with (quasi-)norms given by

$$\|\lambda\|_{s_{p,q}^r b(\Omega)} := \left( \sum_{j \in \mathbb{N}_0^n} 2^{\|j\|_{1(r-1/p)q}} \left( \sum_{m \in A_j^\Omega} |\lambda_{j,m}|^p \right)^{q/p} \right)^{1/q},$$

$$\|\lambda\|_{s_{p,q}^r f(\Omega)} := \left( \sum_{j \in \mathbb{N}_0^n} 2^{\|j\|_{1rq}} \left( \sum_{m \in A_j^\Omega} \lambda_{j,m} \chi_{j,m}(\cdot)^q \right)^{p/q} \right)^{1/p}.$$
Let us also define the following building blocks for $\mu \in \mathbb{N}_0$ fixed

$$s_{p,q}^r b(\Omega)_\mu := \left\{ \lambda : \|\lambda\|_{s_{p,q}^r b(\Omega)_\mu} \right\},$$

$$:= \left\{ \sum_{\|j\|_1=\mu} 2^{\|j\|_r (r-1/p)q} \left( \sum_{m \in A_j^\Omega} |\lambda_{j,m}|^p \right)^{q/p} \right\}^{1/q} < \infty,$$

$$s_{p,q}^r f(\Omega)_\mu := \left\{ \lambda : \|\lambda\|_{s_{p,q}^r f(\Omega)_\mu} \right\},$$

$$:= \left\{ \sum_{\|j\|_1=\mu} 2^{\|j\|_r q} \left( \sum_{m \in A_j^\Omega} \lambda_{j,m} \chi_{j,m}(\cdot)^q \right)^{1/q} \right\} \|_{p} < \infty \right\}. \quad (17)$$

Clearly, for $\mu \in \mathbb{N}_0$ we have

$$\# \{ j \in \mathbb{N}_0^n : \|j\|_1 = \mu \} = \left( \begin{array}{c} \mu + n - 1 \\ \mu \end{array} \right) \simeq d (\mu + 1)^{n-1},$$

and $\# A_j^\Omega \simeq 2^{\|j\|_1} = 2^\mu.$ Consider

$$\text{id} : s_{p_0,q_0}^{r_0} a(\Omega) \to s_{p_1,q_1}^{r_1} a^\dagger (\Omega)$$

for $r_0 - r_1 > 1/p_0 - 1/p_1$ such that the embedding is compact. Defining the building blocks

$$(\text{id}_\mu \lambda)_{j,m} := \left\{ \begin{array}{ll} \lambda_{j,m} : \|j\|_1 = \mu, m \in A_j^\Omega \\ 0 & : \text{otherwise} \end{array} \right\},$$

we have $\text{id} = \sum_{\mu=0}^{\infty} \text{id}_\mu$. Of course, the identity

$$e_k (\text{id}_\mu : s_{p_0,q_0}^{r_0} a(\Omega) \to s_{p_1,q_1}^{r_1} a^\dagger (\Omega)_\mu) = e_k (\text{id}_\mu' : s_{p_0,q_0}^{r_0} a(\Omega)_\mu \to s_{p_1,q_1}^{r_1} a^\dagger (\Omega)_\mu)$$

holds true, where $\text{id}_\mu'$ denotes the corresponding embedding operator on the respective block (17). Although these operators have the same mapping properties we use different notations to formally distinguish between them. If $a = a^\dagger = b$ we also have, for a finite index set $I$, that

$$e_k \left( \sum_{\mu \in I} \text{id}_\mu : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{p_1,q_1}^{r_1} b(\Omega) \right) = e_k \left( \text{id}' : \ell_{q_0} \left( (s_{p_0,q_0}^{r_0} b(\Omega)_\mu)_{\mu \in I} \right) \to \ell_{q_1} \left( (s_{p_1,q_1}^{r_1} b(\Omega)_\mu)_{\mu \in I} \right) \right) \simeq e_k \left( \text{id}'' : \ell_{q_0} \left( (X_\mu)_{\mu \in I} \right) \to \ell_{q_1} \left( (Y_\mu)_{\mu \in I} \right) \right), \quad (18)$$
where $X_\mu = 2^{\mu (r_0 - 1/p_0)} \ell_{q_0}^{(\mu+1)\sigma -1} (\ell_{p_0}^{2\mu})$ and $Y_\mu = 2^{\mu (r_1 - 1/p_1)} \ell_{q_1}^{(\mu+1)\sigma -1} (\ell_{p_1}^{2\mu})$, which means $d_\mu = 2^{\mu}$ and $b_\mu = (\mu + 1)^{n - 1}$ in the notation of (14). In particular, we have $b_\mu \lesssim d_\mu$.

### 5.3 Entropy Numbers

As a consequence of the boundedness of certain restriction and extension operators, see [42, 4.5], the investigation of entropy numbers of Besov space embeddings can be shifted to the sequences spaces side. We formulate our first result in the framework of sequence spaces, which improves the upper bound. More specifically, we prove that the lower bound in (23) is sharp in the case that $0 \leq 1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1$, which also includes the limiting case $r_0 - r_1 = 1/q_0 - 1/q_1$. What is known in this direction is summarized in Remark 20 below.

**Proposition 17** Let $\Omega$ be a bounded domain and $0 < q_0 < q_1 \leq \infty$, $0 < p_0 \leq p_1 \leq \infty$ such that

$$1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1.$$  

Then we have

$$e_m(\text{id} : s_{p_0,q_0}^{r_0} b(\Omega) \rightarrow s_{p_1,q_1}^{r_1} b(\Omega)) \simeq m^{-(r_0 - r_1)} , \quad m \in \mathbb{N}.$$  

**Proof** The lower bound follows by [42, Thm. 3.18]. The upper bound is the actual contribution. We argue as follows.

**Step 1.** Put $\varrho := \min\{1, p_1, q_1\}$ and fix $m \geq m_0$, where $m_0$ is large enough (depending on $p_0, p_1, q_0, q_1, r_0, r_1$). We decompose the identity operator $\text{id}$ as follows

$$\text{id} = \left( \sum_{\mu=0}^{L_m} \text{id}_\mu \right) + \left( \sum_{\mu=L_m+1}^{M_m+L_m} \text{id}_\mu \right) + \left( \sum_{\mu=M_m+L_m+1}^{\infty} \text{id}_\mu \right), \quad (19)$$

where $L_m := \lfloor \log_2(m) \rfloor$ and $M_m := \lfloor m/8 \rfloor$. With an eye on Proposition 10, this means, in particular, that $m \geq 8L_m$ and $m \geq 8M_m$ (for $m$ large enough). Using (4) we obtain

$$e_{2m}(\text{id})^\varrho \lesssim e_m \left( \sum_{\mu=0}^{L_m} \text{id}_\mu \right)^\varrho + e_m \left( \sum_{\mu=L_m+1}^{M_m+L_m} \text{id}_\mu \right)^\varrho + \sum_{\mu=M_m+L_m+1}^{\infty} ||\text{id}_\mu||^\varrho. \quad (20)$$

**Step 2.** We estimate the first summand. By (18) this breaks down to the entropy numbers

$$e_m \left( \text{id} : \ell_{q_0} (X_\mu)_{\mu \in I} \rightarrow \ell_{q_1} (Y_\mu)_{\mu \in I} \right) \quad (21)$$

with $X_\mu, Y_\mu$ chosen as after (18) and $I$ denotes the range for $\mu$. Putting

$$p := q_0, r := q_1, q := p_0, u := p_1, d_\mu := 2^{\mu}, b_\mu := (\mu + 1)^{n - 1}, \alpha := r_0 - 1/p_0$$

we obtain

$$p := q_0, r := q_1, q := p_0, u := p_1, d_\mu := 2^{\mu}, b_\mu := (\mu + 1)^{n - 1}, \alpha := r_0 - 1/p_0$$

$$e_m(\text{id} : \ell_{q_0} (X_\mu)_{\mu \in I} \rightarrow \ell_{q_1} (Y_\mu)_{\mu \in I}) \quad (21)$$

with $X_\mu, Y_\mu$ chosen as after (18) and $I$ denotes the range for $\mu$. Putting

$$p := q_0, r := q_1, q := p_0, u := p_1, d_\mu := 2^{\mu}, b_\mu := (\mu + 1)^{n - 1}, \alpha := r_0 - 1/p_0$$

we obtain
and $\beta := r_1 - 1/p_1$ in (14) we may apply Proposition 15. Since $m \geq \max\{8L_m, \max_{\mu} d_{\mu}\}$ we may apply Proposition 15 to obtain
\begin{equation}
    e_m \lesssim \left(\frac{1}{m}\right)^{\alpha - \beta + 1/q - 1/u} = \left(\frac{1}{m}\right)^{r_0 - r_1}. \tag{22}
\end{equation}

Note that, due to Proposition 15, we only used that $r_0 - r_1 \leq 1/q_0 - 1/q_1$. To estimate the first summand in (20) it is not needed that $r_0 - r_1 > 1/p_0 - 1/p_1$.

Step 3. Let us address the second summand in (20). Clearly, it can be reduced to (21) with spaces $X_\mu, Y_\mu$ defined analogously, but with $\mu$ running this time in the range
\[ I = \{L_m + 1, \ldots, L_m + M_m\}. \]

Hence, we have $b := \#I = M_m \leq \min_{\mu} d_{\mu}$. We apply Proposition 16 to end up with (22). Note, that we have used here only that $\alpha - \beta > 0$, or, equivalently, $r_0 - r_1 > 1/p_0 - 1/p_1$.

Step 4. Finally, we deal with the third summand in (20). Clearly, we have
\[ \|\text{id}_{\mu}\| \lesssim 2^{-\mu(r_0 - r_1 - 1/p_0 + 1/p_1)}. \]

This gives
\begin{align*}
    \sum_{\mu=M_m+L_m+1}^{\infty} \|\text{id}_{\mu}\|^\varrho & \lesssim \sum_{\mu=M_m+L_m+1}^{\infty} 2^{-\varrho \mu(r_0 - r_1 - 1/p_0 + 1/p_1)} \\
    & \simeq 2^{-\varrho(m/8 + L_m)(r_0 - r_1 - 1/p_0 + 1/p_1)} \\
    & \lesssim m^{-\varrho(r_0 - r_1)}. 
\end{align*}

This concludes the proof. $\Box$

In the next theorem we consider the situation where a Besov type sequence space compactly embeds into a Triebel–Lizorkin type sequence space. This setting is particularly important, since it leads to results with target space $L_p$.

**Proposition 18** Let $\Omega$ be a bounded domain and $0 < q_0 < q_1 \leq \infty$, $0 < p_0 \leq p_1 < \infty$, $q_0 < p_0$, $r_0 > r_1$ such that
\[ 1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1. \]

Then we have
\[ e_m(\text{id} : s_{p_0,q_0}^{r_0} b(\Omega) \rightarrow s_{p_1,q_1}^{r_1} f(\Omega)) \simeq m^{-(r_0 - r_1)}. \]

**Proof** Again, the lower bounds follow from [42, Thm. 3.18].

**Step 1.** In the case $p_1 > q_1$ we use the commutative diagram in Fig. 1.
Then we have by (5) and (6)

\[ e_m(\text{id} : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{p_1,q_1}^{r_1} f(\Omega)) \leq e_m(\text{id} : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{p_1,q_1}^{r_1} b(\Omega)) \cdot \|\text{id} : s_{p_1,q_1}^{r_1} b(\Omega) \to s_{p_1,q_1}^{r_1} f(\Omega)\|. \]

Hence, we may use Proposition 17 and obtain

\[ e_m(\text{id} : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{p_1,q_1}^{r_1} f(\Omega)) \leq m^{-(r_0-r_1)}. \]

**Step 2.** Now we consider \( p_1 < q_1 \). After decomposing the identity operator in an analogous way as in (19) and (20) we use the commutative diagrams in Fig. 2 for the first and second summand, respectively. In fact, for the first summand in (20) we obtain by (6)

\[ e_m\left( \sum_{\mu=0}^{L_m} \text{id}_\mu : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{p_1,q_1}^{r_1} f(\Omega) \right) \leq e_m\left( \sum_{\mu=0}^{L_m} \text{id}_\mu : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{q_1,q_1}^{r_1} b(\Omega) \right) \cdot \|\text{id} : s_{q_1,q_1}^{r_1} b(\Omega) \to s_{p_1,q_1}^{r_1} f(\Omega)\|. \]

Note, that the identity operator is bounded since \( \Omega \) is a bounded domain.

Furthermore, the entropy numbers

\[ e_m\left( \sum_{\mu=0}^{L_m} \text{id}_\mu : s_{p_0,q_0}^{r_0} b(\Omega) \to s_{q_1,q_1}^{r_1} b(\Omega) \right) \leq m^{-(r_0-r_1)}. \]
can be estimated by the same reasoning as in Step 2 of the proof of Proposition 17. Note that $r_0 - r_1$ may be smaller than $1/p_0 - 1/q_1$. However, this is not important for the argument (based on Proposition 15). It remains to consider the second summand in (20). Here we use the right diagram in Fig. 2 and obtain

\[ e_m \left( \sum_{\mu = M_m + 1}^{M_m + L_m} \text{id}_\mu : s_{p_0, q_0}^{r_0} f(\Omega) \to s_{p_1, q_1}^{r_1} f(\Omega) \right) \leq e_m \left( \sum_{\mu = M_m + 1}^{M_m + L_m} \text{id}_\mu : s_{p_0, q_0}^{r_0} b(\Omega) \to s_{p_1, p_1}^{r_1} b(\Omega) \right) \cdot \| \text{id} : s_{p_1, p_1}^{r_1} b(\Omega) \to s_{p_1, q_1}^{r_1} f(\Omega) \|. \]

We continue to estimate the appearing entropy numbers as in Step 3 of the proof of Proposition 17. Note that $r_0 - r_1$ might be larger than $1/q_0 - 1/p_1$. However, for the argument, we only need $r_0 - r_1 > 1/p_0 - 1/p_1$. This concludes the proof.

Let us finally consider the situation, where a Triebel–Lizorkin type sequence space compactly embeds into a Besov type sequence space.

**Proposition 19** Let $0 < q_0 < q_1 \leq \infty$, $0 < p_0 \leq p_1 < \infty$, $q_1 > p_1$, $r_0 > r_1$ such that

\[ 1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1 , \]

and let $\Omega$ be a bounded domain. Then we have

\[ e_m(\text{id} : s_{p_0, q_0}^{r_0} f(\Omega) \to s_{p_1, q_1}^{r_1} f(\Omega)) \simeq m^{-(r_0 - r_1)} . \]

**Proof** The lower bound follows from [42, Thm. 3.18].

**Step 1.** For the upper bound in the case $p_0 < q_0$ we may use the commutative diagram in Fig. 3 below to decompose the identity operator. Afterwards, we use (5) to reduce everything to the situation in Proposition 17.

**Step 2.** In the case $p_0 > q_0$ we argue analogously to Step 2 of Proposition 18. This time we use the decompositions in Fig. 4 for the first and second summand in (20), respectively.

Unfortunately, we were not able to find a corresponding result for the $f - f$ situation. So, this remains an open problem.
Remark 20 To clarify the contribution of this paper, let us briefly recapitulate the known results and open questions which motivated this work. For several results and historical remarks on the subject we refer to [8] and the references therein. In particular, Vybíral [42, Thm. 4.9] proved for $0 < p_0 \leq p_1 \leq \infty$ and $0 < q_0 \leq q_1 \leq \infty$ in the case of small smoothness

$$1/p_0 - 1/p_1 < r \leq 1/q_0 - 1/q_1,$$

that there is for any $\varepsilon > 0$ a number $C_\varepsilon > 0$ such that

$$cm^{-r} \leq e_m(\text{id} : s_{p_0,q_0}^r b(\Omega) \to s_{p_1,q_1}^r b(\Omega)) \leq C_\varepsilon m^{-r} (\log m)^{\varepsilon}, \quad m \geq 2. \tag{23}$$

The result is a direct consequence of the bound for $r > \max\{1/p_0 - 1/p_1, 1/q_0 - 1/q_1\}$ (the case of “large smoothness”), saying that

$$e_m(\text{id} : s_{p_0,q_0}^r b(\Omega) \to s_{p_1,q_1}^r b(\Omega)) \simeq m^{-r} (\log m)^{(n-1)(r_0-r_1-1/q_0+1/q_1)+}. \tag{24}$$

In fact, the entropy numbers in (23) can be bounded from above by

$$e_m(\text{id} : s_{p_0,q_*}^r b(\Omega) \to s_{p_1,q_1}^r b(\Omega))$$

if $q_* \geq q_0$. Now choose $q_1 > q_* > q_0$ such that $1/q^* - 1/q_1 + \varepsilon/(n-1) = r_0 - r_1 > 1/q^* - 1/q_1$, which, together with (24) and $q_0$ replaced by $q^*$, implies (23).

The propositions proved above allow us to improve a number of existing results for the entropy numbers of the embedding

$$\text{Id} : S_{p_0,q_0}^r A(\Omega) \to S_{p_1,q_1}^r A^\dag(\Omega).$$

Theorem 21 Let $\Omega$ be a bounded domain and $A, A^\dag \in \{B, F\}$ but $(A, A^\dag) \neq (F, F)$. Let $0 < q_0 < q_1 \leq \infty$, $0 < p_0 \leq p_1 < \infty$ and $r_0 > r_1$ such that

$$1/p_0 - 1/p_1 < r_0 - r_1 \leq 1/q_0 - 1/q_1.$$

In addition, we assume $q_0 < p_0$ if $(A, A^\dag) = (B, F)$ and $q_1 > p_1$ if $(A, A^\dag) = (F, B)$, respectively. Then the following holds

$$e_m(\text{Id} : S_{p_0,q_0}^r A(\Omega) \to S_{p_1,q_1}^r A^\dag(\Omega)) \simeq m^{-(r_0-r_1)}, \quad m \in \mathbb{N}.$$

Proof The result is a direct consequence of Propositions 17, 18, 19 and the machinery described in the proof of [42, Thm. 4.11]. \qed

As a corollary of Theorem 21, we obtain the following result, which settles Open Problem 6.4 in [8].

Springer
Corollary 22 Let $\Omega$ be as above. Let further $0 < q < p_0 \leq p_1$, $1 < p_1 < \infty$ and $1/p_0 - 1/p_1 < r \leq 1/q - 1/2$. Then we have
\[ e_m(\text{Id} : S_{p_0,q}^r B(\Omega) \to L_{p_1}(\Omega)) \simeq m^{-r}. \] (25)

Proof Identifying $S_{p_1,2}^0 F(\Omega) = L_{p_1}(\Omega)$ in the case $1 < p_1 < \infty$, the result is a direct consequence of Theorem 21. \(\Box\)

With the final corollary below (from Theorem 21) we close some more gaps in [42, Thm. 4.18 (ii), (iii)].

Corollary 23 Let $\Omega$ be as above. We have the following sharp bounds for entropy numbers.
(i) Let $1 < p \leq \infty$ and $1/p < r \leq 1$. Then, we have
\[ e_m(\text{Id} : S_{p,1}^r B(\Omega) \to S_{0,\infty}^0 B(\Omega)) \simeq m^{-r}. \]
(ii) Let $1 < p < q < \infty$ and $1/p - 1/q < r \leq 1/2$. Then, we have
\[ e_m(\text{Id} : S_{p}^r W(\Omega) \to S_{q,\infty}^0 B(\Omega)) \simeq m^{-r}. \]
(iii) Let $0 < q < p \leq \infty$, $q < 1$ and $1/p < r \leq 1/q - 1$. Then, we have
\[ e_m(\text{Id} : S_{p,q}^r B(\Omega) \to L_{\infty}(\Omega)) \simeq m^{-r}. \]

Remark 24 Entropy numbers of mixed smoothness Sobolev-Besov embeddings into $L_p$, where $1 \leq p \leq \infty$, recently gained significant interest, see [40] and [33–36]. There are some fundamental open problems connected with $p = \infty$, see [8, 2.6, 6.4, 6.5]. Interestingly, when choosing the third index $q$ small enough in Corollaries 22, 23 we get rid of the logarithm.

Acknowledgements The authors would like to thank Dinh Dung, Thomas Kühn, Van Kien Nguyen, Winfried Sickel and Vladimir N. Temlyakov for several discussions on the topic. They would also like to thank two anonymous referees for their valuable comments. T.U. and S.M. would like to acknowledge support by the DFG UI-403/2-1 and by the Fraunhofer Cluster of Excellence “Cognitive Internet Technologies”.

Funding Open Access funding enabled and organized by Projekt DEAL.
References

1. Akashi, S.: An operator theoretical characterization of $\varepsilon$-entropy in Gaussian processes. Kodai Math. J. 9(1), 58–67 (1986)
2. Aoki, T.: Locally bounded linear topological spaces. Proc. Imp. Acad. 18(10), 588–594 (1942)
3. Billingsley, P.: Ergodic Theory and Information. Wiley, New York (1965)
4. Carl, B.: Entropy numbers, $s$-numbers, and eigenvalue problems. J. Funct. Anal. 41(3), 290–306 (1981)
5. Carl, B., Stephani, I.: Entropy, Compactness and the Approximation of Operators. Cambridge University Press, Cambridge (1990)
6. Dirksen, S., Ullrich, T.: Gelfand numbers related to structured sparsity and Besov space embeddings with small mixed smoothness. J. Complex. 48, 69–102 (2018)
7. Düng, D.: Non-linear approximations using sets of finite cardinality or finite pseudo-dimension. In: 3rd Conference of the Foundations of Computational Mathematics, vol. 17, pp. 467–492. 2001, Oxford (1999)
8. Düng, D., Temlyakov, V.N., Ullrich, T.: Hyperbolic Cross Approximation. Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, Basel/Berlin (2018)
9. Edmunds, D.E., Netrusov, Y.: Entropy numbers of embeddings of Sobolev spaces in Zygmund spaces. Studia Math. 128(1), 71–102 (1998)
10. Edmunds, D.E., Netrusov, Y.: Schütt’s theorem for vector-valued sequence spaces. J. Approx. Theory 178, 13–21 (2014)
11. Edmunds, D.E., Triebel, H.: Function Spaces, Entropy Numbers, Differential Operators. Cambridge Tracts in Mathematics, vol. 120. Cambridge University Press, Cambridge (1996)
12. Foucart, S., Pajor, A., Rauhut, H., Ullrich, T.: The Gelfand widths of $\ell_p$-balls for $0 < p \leq 1$. J. Complex. 26(6), 629–640 (2010)
13. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York (2013)
14. Gilbert, E.: A comparison of signalling alphabets. Bell Syst. Techol. J. 31, 504–522 (1952)
15. Guédon, O., Litvak, A.E.: Euclidean projections of a $p$-convex body. In: Geometric Aspects of Functional Analysis, volume 1745 of Lecture Notes in Math., pp. 95–108. Springer, Berlin (2000)
16. Hinrichs, A., Kolleck, A., Vybíral, J.: Carl’s inequality for quasi-Banach spaces. J. Funct. Anal. 271(8), 2293–2307 (2016)
17. Kempka, H., Vybíral, J.: Volumes of unit balls of mixed sequence spaces. Math. Nachr. 290(8–9), 1317–1327 (2017)
18. Kolmogorov, A.N.: On certain asymptotic characteristics of completely bounded metric spaces. Dokl. Akad. Nauk SSSR (N.S.) 108, 385–388 (1956)
19. Kolmogorov, A.N.: A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. Dokl. Akad. Nauk SSSR (N.S.) 119, 861–864 (1958)
20. König, H.: Eigenvalue Distribution of Compact Operators. Birkhäuser Verlag, Basel (1986)
21. Kuelbs, J., Li, W.V.: Metric entropy and the small ball problem for Gaussian measures. J. Funct. Anal. 116(1), 133–157 (1993)
22. Kühn, T.: A lower estimate for entropy numbers. J. Approx. Theory 110(1), 120–124 (2001)
23. Kühn, T., Leopold, H.-G., Sickel, W., Skrzypczak, L.: Entropy numbers of embeddings of weighted Besov spaces. Constr. Approx. 23(1), 61–77 (2006)
24. Li, W.V., Linde, W., et al.: Approximation, metric entropy and small ball estimates for Gaussian measures. Ann. Probab. 27(3), 1556–1578 (1999)
25. Lorentz, G.G., Golitschek, M.V., Makovoz, Y.: Constructive Approximation, volume 304 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, Advanced problems (1996)
26. Malorov, V.E.: Discretization of the problem of diameters. Uspehi Mat. Nauk 30(6(186)), 179–180 (1975)
27. Mayer, S.: Preasymptotics via metric entropy. Dr. Hut Verlag, Munich, PhD thesis (2018)
28. Pietsch, A.: Operator ideals, volume 16. Deutscher Verlag der Wissenschaften, (1978)
29. Pinkus, A.: Eigenvalues and s-Numbers. Cambridge Studies in Advanced Mathematics, vol. 13. Cambridge University Press, Cambridge (1987)
30. Pietsch, A.: n-widths in approximation theory, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 7. Springer-Verlag, Berlin (1987)
31. Pontrjagin, L., Schnirelmann, L.: Sur une propriété métrique de la dimension. Ann. Math. (2) 33(1), 156–162 (1932)
32. Rolewicz, S.: On a certain class of linear metric spaces. Bull. Acad. Polon. Sci 5, 471–473 (1957)
33. Romanyuk, A.S.: Entropy numbers and widths of the classes $B^r_{p,q}$ of periodic functions of many variables. Ukrain. Mat. Zh. 68(10), 1403–1417 (2016)
34. Romanyuk, A.S.: Estimation of the entropy numbers and Kolmogorov widths for the Nikol’skii-Besov classes of periodic functions of many variables. Ukrainian Math. J. 67(11), 1739–1757 (2016)
35. Romanyuk, A. S.: Translation of Ukrain. Mat. Zh. 67, no. 11, 1540–1556 (2015)
36. Romanyuk, A.S.: Entropy numbers and widths for the Nikol’skii–Besov classes of functions of many variables in the space $L_\infty$. Anal. Math. 45(1), 133–151 (2019)
37. Schmeisser, H.-J., Triebel, H.: Topics in Fourier Analysis and Function Spaces. Wiley, Chichester (1987)
38. Schütt, C.: Entropy numbers of diagonal operators between symmetric Banach spaces. J. Approx. Theory 40(2), 121–128 (1984)
39. Temlyakov, V.: Greedy Approximation. Cambridge Monographs on Applied and Computational Mathematics, vol. 20. Cambridge University Press, Cambridge (2011)
40. Temlyakov, V.: On the entropy numbers of the mixed smoothness function classes. J. Approx. Theory 217, 26–56 (2017)
41. Varshamov, R.: Estimate of the number of signals in error correcting codes. Dokl. Acad. Nauk SSSR 117, 739–741 (1952)
42. Vybiral, J.: Function spaces with dominating mixed smoothness. Dissertationes Math. (Rozpravy Mat.) 436, 73 (2006)
43. Williamson, R.C., Smola, A.J., Schölkopf, B.: Entropy numbers, operators and support vector kernels. In: Fischer, P., Simon, H.U. (eds.) Computational Learning Theory, pp. 285–299. Springer, Berlin (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.