Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings II

Stephan Dahlke*, Erich Novak, Winfried Sickel

dahlke@mathematik.uni-marburg.de, novak@math.uni-jena.de, sickel@math.uni-jena.de

August 22, 2005

Abstract

We study the optimal approximation of the solution of an operator equation $\mathcal{A}(u) = f$ by four types of mappings: a) linear mappings of rank n; b) n-term approximation with respect to a Riesz basis; c) approximation based on linear information about the right hand side f; d) continuous mappings. We consider worst case errors, where f is an element of the unit ball of a Sobolev or Besov space $B^r_q(L_p(\Omega))$ and $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain; the error is always measured in the H^s-norm. The respective widths are the linear widths (or approximation numbers), the nonlinear widths, the Gelfand widths, and the manifold widths. As a technical tool we also study the Bernstein numbers. Our main results are the following. If $p \geq 2$ then the order of convergence is the same for all four classes of approximations. In particular, the best linear approximations are of the same order as the best nonlinear ones. The best linear approximation can be quite difficult to realize as a numerical algorithm since the optimal Galerkin space usually depends on the operator and of the shape of the domain Ω. For $p < 2$ there is an essential difference, nonlinear approximations are better than linear ones. Also in this case it turns out, however, that linear information about the right hand side f is optimal. As a main theoretical tool we study best n-term approximation with respect to an optimal Riesz basis and related nonlinear widths. Then we study the Poisson equation in a polygonal domain. In this case it turns out that best n-term wavelet approximation is (almost) optimal. The main

*The work of this author has been supported through the European Union's Human Potential Programme, under contract HPRN–CT–2002–00285 (HASSIP), and through DFG, Grant Da 360/4–2, Da 360/4–3.

results are about approximation, not about computation. However, we also discuss consequences of the results for the numerical complexity of operator equations.

AMS subject classification: 41A25, 41A46, 41A65, 42C40, 65C99

Key Words: Elliptic operator equations, worst case error, linear and nonlinear approximation methods, best n-term approximation, Besov spaces, Gelfand widths, Bernstein widths, manifold widths.

1 Introduction

We study the optimal approximation of the solution of an operator equation

$$\mathcal{A}(u) = f, \tag{1}$$

where $\mathcal{A}$ is a linear operator

$$\mathcal{A} : H \to G \tag{2}$$

from a Hilbert space H to another Hilbert space G. We always assume that $\mathcal{A}$ is boundedly invertible, hence (1) has a unique solution for any $f \in G$. We have in mind the more specific situation of an elliptic operator equation which is given as follows. Assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and assume that

$$\mathcal{A} : H_0^s(\Omega) \to H^{-s}(\Omega) \tag{3}$$

is an isomorphism, where $s > 0$. A standard case (for second order elliptic boundary value problems for PDEs) is $s = 1$, but also other values of s are of interest. Now we put $H = H_0^s(\Omega)$ and $G = H^{-s}(\Omega)$. Since $\mathcal{A}$ is boundedly invertible, the inverse mapping $S : G \to H$ is well defined. This mapping is sometimes called the solution operator – in particular if we want to compute the solution $u = S(f)$ from the given right-hand side $\mathcal{A}(u) = f$.

We use linear and (different kinds of) nonlinear mappings S_n for the approximation of the solution $u = \mathcal{A}^{-1}(f)$ for f contained in $F \subset G$. We consider the worst case error

$$e(S_n, F, H) = \sup_{\|f\|_F \leq 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H, \tag{4}$$

where F is a normed (or quasi-normed) space, $F \subset G$. In our main results, F will be a Sobolev or Besov space.[1] Hence we use the following commutative diagram

$$\begin{array}{ccc} G & \xrightarrow{S} & H \\ I \;\nwarrow & & \nearrow\; S_F \\ & F. & \end{array}$$

Here $I: F \to G$ denotes the identity and S_F the restriction of S to F. In the specific case (3) this diagram is given by

$$\begin{array}{ccc} H^{-s}(\Omega) & \xrightarrow{S} & H_0^s(\Omega) \\ I \;\nwarrow & & \nearrow\; S_t \\ & B_q^{-s+t}(L_p(\Omega)), & \end{array}$$

where $B_q^{-s+t}(L_p(\Omega))$ denotes a Besov space compactly embedded into $H^{-s}(\Omega)$, cf. the Appendix for a definition, and S_t the restriction of S to $B_q^{-s+t}(L_p(\Omega))$. We are interested in approximations that have the optimal order of convergence depending on n, where n denotes the *degree of freedom*. In general our results are *constructive in a mathematical sense*, because we can describe optimal approximations S_n in mathematical terms. This does not mean, however, that these descriptions are constructive in a practical sense since it might be very difficult to convert those descriptions into a practical algorithm. We will discuss this more thoroughly in Section 3.4. As a consequence, most of our results give optimal benchmarks and can serve for the evaluation of old and new algorithms. We study and compare *four kinds of approximation methods*, see Section 2.1 for details.

- We consider the class $\mathcal{L}_n$ of all continuous linear mappings $S_n : F \to H$,

  $$S_n(f) = \sum_{i=1}^{n} L_i(f) \cdot \tilde{h}_i$$

  with arbitrary $\tilde{h}_i \in H$. The worst case error of optimal linear mappings is given by the *approximation numbers* or *linear widths*

  $$e_n^{\mathrm{lin}}(S, F, H) = \inf_{S_n \in \mathcal{L}_n} e(S_n, F, H).$$

---

[1] Formally we only deal with Besov spaces. Because of the embeddings $B_1^{-s+t}(L_p(\Omega)) \subset W_p^{-s+t}(\Omega) \subset B_\infty^{-s+t}(L_p(\Omega))$, which hold for $1 \le p \le \infty$, $t \ge s$, see [84], our results are valid also for Sobolev spaces.

- For a given basis $\mathcal{B}$ of H we consider the class $\mathcal{N}_n(\mathcal{B})$ of all (linear or nonlinear) mappings of the form

$$S_n(f) = \sum_{k=1}^{n} c_k \, h_{i_k},$$

where the c_k and the i_k depend in an arbitrary way on f. We also allow that the basis $\mathcal{B}$ is chosen in a nearly arbitrary way. Then the *nonlinear widths* $e_{n,C}^{\rm non}(S, F, H)$ are given by

$$e_{n,C}^{\rm non}(S, F, H) = \inf_{\mathcal{B} \in \mathcal{B}_C} \inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, F, H).$$

Here $\mathcal{B}_C$ denotes a set of Riesz bases for H where C indicates the stability of the basis. These numbers are the main topic of our analysis.

- We also study methods S_n with $S_n = \varphi_n \circ N_n$, where $N_n : F \to \mathbb{R}^n$ is linear and continuous and $\varphi_n : \mathbb{R}^n \to H$ is arbitrary. This is the class of all (linear or nonlinear) approximations S_n that use *linear information of cardinality n* about the right hand side f. The respective widths are

$$r_n(S, F, H) := \inf_{S_n} e(S_n, F, H),$$

they are closely related to the *Gelfand numbers.*

- Let $\mathcal{C}_n$ be the class of continuous mappings, given by arbitrary continuous mappings $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$. Again we define the worst case error of optimal continuous mappings by

$$e_n^{\rm cont}(S, F, H) = \inf_{S_n \in \mathcal{C}_n} e(S_n, F, H),$$

where $S_n = \varphi_n \circ N_n$. These numbers are called *manifold widths* of S.

For problems (3) with $F = B_q^r(L_p(\Omega))$ our main results are the following. If $p \geq 2$ then the order of convergence is the same for all four classes of approximations. In particular, the best linear approximations are of the same order as the best nonlinear ones. The best linear approximation can be quite difficult to realize as a numerical algorithm since the optimal Galerkin space usually depends on the operator and of the shape of the domain Ω. For $p < 2$ there is an essential difference, nonlinear approximations are better than linear ones. Also in this case it turns out, however, that linear information about the right hand side f is optimal. As a main theoretical tool we study best n-term approximation with respect to an optimal Riesz basis and related nonlinear widths. The main results are about approximation, not about computation. However, we also discuss consequences of the results for the numerical complexity of operator equations.

The paper is organized as follows:

We add a few comments. The main results of our paper are contained in Section 3.2. They are further illustrated for the case of the Poisson equation in Section 3.3. A discussion in connection with *uniform approximation*, *adaptive/nonadaptive information*, *adaptive numerical schemes*, and *complexity* is contained in Section 3.4. All proofs are contained in Section 4. Of independent interest are the estimates of the widths of embedding operators for Besov spaces, see Section 4.3.

Notation. We write $a \asymp b$ if there exists a constant $c > 0$ (independent of the context dependent relevant parameters) such that

$$c^{-1}\, a \leq b \leq c\, a\,.$$

All unimportant constants will be denoted by c, sometimes with additional indices.

2 Linear and Nonlinear Widths

Widths represent concepts of optimality. In this section we shall discuss several variants. Most important for us will be the nonlinear widths $e_n^{\rm non}$ and the linear widths $e_n^{\rm lin}$. We also study Gelfand and manifold widths and, as a vehicle of the proofs, Bernstein widths.

2.1 Classes of Admissible Mappings

Linear Mappings S_n

Here we consider the class $\mathcal{L}_n$ of all continuous linear mappings $S_n : F \to H$,

$$S_n(f) = \sum_{i=1}^{n} L_i(f)\, h_i \tag{5}$$

where the $L_i : F \to \mathbb{R}$ are linear functionals and h_i are elements of H. We consider the worst case error

$$e(S_n, F, H) := \sup_{\|f\|_F \le 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H, \tag{6}$$

where F is a normed (or quasi-normed) space, $F \subset G$. According to this we ask for the optimal linear approximation and the numbers

$$e_n^{\text{lin}}(S, F, H) = \inf_{S_n \in \mathcal{L}_n} e(S_n, F, H), \tag{7}$$

usually called *approximation numbers* or *linear widths* of $S : F \to H$, cf. [57, 69, 70, 78].

Nonlinear Mappings S_n

Let $\mathcal{B} = \{h_1, h_2, \dots\}$ be a subset of H. Then the *best n-term approximation* of an element $u \in H$ with respect to this set $\mathcal{B}$ is defined as

$$\sigma_n(u, \mathcal{B})_H := \inf_{i_1,\dots,i_n} \inf_{c_1,\dots c_n} \|u - \sum_{k=1}^{n} c_k\, h_{i_k}\|_H\,. \tag{8}$$

This subject is widely studied, see the surveys [28] and [77]. Now we continue by looking for an optimal set $\mathcal{B}$ as has been done in Kashin [51] and Donoho [37], cf. also [77]. Temlyakov [77] has introduced the quantities

$$\inf_{\mathcal{B} \text{ basis of } H} \sup_{\|u\|_Y \le 1} \sigma_n(u, \mathcal{B})_H\,,$$

where Y denotes a subspace of H. For our purposes this would be too general. We will study certain approximations of S based on Riesz bases, cf. e.g. Meyer [59, page 21].

Definition 1. *Let H be a Hilbert space. Then the sequence $h_1, h_2, \dots$ of elements of H is called a* Riesz basis *for H if there exist positive constants A and B such that,* *for every sequence of scalars $\alpha_1, \alpha_2, \ldots$ with $\alpha_k \neq 0$ for only finitely many k, we have*

$$A\Big(\sum_k |\alpha_k|^2\Big)^{1/2} \le \Big\| \sum_k \alpha_k\, h_k \Big\|_H \le B\Big(\sum_k |\alpha_k|^2\Big)^{1/2} \tag{9}$$

and the vector space of finite sums $\sum \alpha_k\, h_k$ is dense in H.

Remark 1. *The constants A, B reflect the stability of the basis. Orthonormal bases are those with $A = B = 1$. Typical examples of Riesz bases are the biorthogonal wavelet bases on $\mathbb{R}^d$ or on certain Lipschitz domains, cf.* Cohen [11, Sect. 2.6, 2.12].

In what follows

$$\mathcal{B} = \{h_i \mid i \in \mathbb{N}\} \tag{10}$$

will always denote a Riesz basis of H and A and B are the corresponding optimal constants in (9).
For a given basis $\mathcal{B}$ we consider the class $\mathcal{N}_n(\mathcal{B})$ of all (linear or nonlinear) mappings of the form

$$S_n(f) = \sum_{k=1}^{n} c_k\, h_{i_k}, \tag{11}$$

where the c_k and the i_k depend in an arbitrary way on f. By the arbitrariness of S_n one obtains immediately

$$\inf_{S_n \in \mathcal{N}_n(\mathcal{B})} \sup_{\|f\|_F \le 1} \|\mathcal{A}^{-1} f - S_n(f)\|_H = \sup_{\|f\|_F \le 1} \sigma_n(\mathcal{A}^{-1} f, \mathcal{B})_H\,. \tag{12}$$

It is natural to assume some common stability of the bases under consideration. For a real number $C \ge 1$ we put

$$\mathcal{B}_C := \Big\{\mathcal{B} :\; B/A \le C\Big\}. \tag{13}$$

We are ready to define the nonlinear widths $e^{\mathrm{non}}_{n,C}(S, F, H)$ by

$$e^{\mathrm{non}}_{n,C}(S, F, H) = \inf_{\mathcal{B} \in \mathcal{B}_C} \inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, F, H). \tag{14}$$

These numbers are the main topic of our analysis. They could be called the *widths of best n-term approximation* (with respect to the collection $\mathcal{B}_C$ of Riesz basis of H).

Remark 2. *i) It should be clear that the class $\mathcal{N}_n(\mathcal{B})$ contains many mappings that are difficult to compute. In particular, the number n just reflects the dimension of a nonlinear manifold and has nothing to do with a computational cost. In this paper we are interested also in lower bounds and hence it is useful to define such a large class of approximations.*

ii) The inequality

$$e_{n,C}^{\mathrm{non}}(S,F,H) \le e_n^{\mathrm{lin}}(S,F,H) \tag{15}$$

is trivial.

(iii) Because of the homogeneity of σ_n, i.e., $\sigma_n(\lambda u, \mathcal{B})_H = |\lambda|\, \sigma_n(u, \mathcal{B})_H$, $\lambda \in \mathbb{R}$, it does not change the asymptotic behaviour of e_n^{non} if we replace $\sup_{\|f\|_F \le 1}$ by $\sup_{\|f\|_F \le c}$, $c > 0$.

Continuous Mappings S_n

Linear mappings S_n are of the form $S_n = \varphi_n \circ N_n$ where both $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$ are linear and continuous. If we drop the linearity condition then we obtain the class of all continuous mappings $\mathcal{C}_n$, given by arbitrary continuous mappings $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$. Again we define the worst case error of optimal continuous mappings by

$$e_n^{\mathrm{cont}}(S,F,H) = \inf_{S_n \in \mathcal{C}_n} e(S_n, F, H). \tag{16}$$

These numbers, or slightly different numbers, were studied by different authors, cf. [29, 30, 39, 57]. Sometimes these numbers are called manifold widths of S, see [30], and we will use this notation here. The inequality

$$e_n^{\mathrm{cont}}(S,F,H) \le e_n^{\mathrm{lin}}(S,F,H) \tag{17}$$

is obvious.

Gelfand Widths and Minimal Radii of Information

We can also study methods S_n with $S_n = \varphi_n \circ N_n$, where $N_n : F \to \mathbb{R}^n$ is linear and continuous and $\varphi_n : \mathbb{R}^n \to H$ is arbitrary. The respective widths are

$$r_n(S,F,H) := \inf_{S_n} e(S_n, F, H). \tag{18}$$

These numbers are called the n-th *minimal radii of information*, they are closely related to Gelfand widths, see Lemma 1 below. The *n-th Gelfand width* of the linear operator $S : F \to H$ is given by

$$d^n(S,F,H) := \inf_{L_1,\ldots,L_n} \sup \Big\{ \|Sf\|_H \,:\, \|f\|_F \le 1,\ L_i(f) = 0\,,\ i = 1, \ldots n \Big\}, \tag{19}$$

where the $L_i : F \to \mathbb{R}$ are continuous linear functionals.

Bernstein Widths

A well-known tool for deriving lower bounds of widths consists in the investigation of Bernstein widths, see [69, 70, 78].

Definition 2. *The number* $b_n(S,F,H)$*, called the* n*-th* Bernstein width *of the operator* $S: F \to H$*, is the radius of the largest* $(n+1)$*-dimensional ball that is contained in* $S(\{\|f\|_F \le 1\})$*.*

Remark 3. *In the literature there are used different definitions of Bernstein widths. E.g. in Pietsch* [68] *the following version is given. Let* X_n *denote subspaces of* F *of dimension* n*. Then*

$$\widetilde{b}_n(S,F,H) := \sup_{X_n \subset F} \inf_{x \in X_n, x \neq 0} \frac{\|Sx\|_H}{\|x\|_F} .$$

As long as S *is an injective mapping we obviously have* $b_n(S,F,H) = \widetilde{b}_{n+1}(S,F,H)$*.*

2.2 Properties of Widths and Relations Between Them

Lemma 1. *Let* $n \in \mathbb{N}$ *and assume that* $F \subset G$ *is quasi-normed.*
(i) *We have* $d^n \le r_n \le 2d^n$ *if* F *is normed and* $d^n \asymp r_n$ *in general.*
(ii) *The inequality*

$$b_n(S,F,H) \le \min\Big(e_n^{\mathrm{cont}}(S,F,H), d^n(S,F,H)\Big) \tag{20}$$

holds for all n*.*

Remark 4. *The inequality* $b_n \le e_n^{\mathrm{cont}}$ *is known, compare e.g. with* [29]*, and the proof technique (via Borsuk's theorem) is often used for the proof of similar results.*

The Bernstein widths b_n can also be used to prove lower bounds for the $e_{n,C}^{\mathrm{non}}$. The following inequality has been proved in [23].

Lemma 2. *Assume that* $F \subset G$ *is quasi-normed. Then*

$$e_{n,C}^{\mathrm{non}}(S,F,H) \ \ge \ \frac{1}{2C}\, b_m(S,F,H) \tag{21}$$

holds for all $m \ge 4\,C^2\,n$*.*

More important for us will be a direct comparison of e_n^{non} and e_n^{cont}. Best n-term approximation yields a mapping

$$S_n(u) = \sum_{k=1}^{n} c_k\, h_{i_k}$$

which is in general not continuous. It is known, however, that certain discontinuous mappings can be suitably modified in order to obtain a continuous n-term approximation with an error which is only slightly worse. See, for example, [30] and [40]. We prove that, under general assumptions, the numbers $e_{n,C}^{\rm non}$ can be bounded from below by the manifold widths $e_n^{\rm cont}$.

Theorem 1. *Let $S : G \to H$ be an isomorphism. Suppose that the embedding $F \hookrightarrow G$ is compact. Then we have for all $C \ge 1$ and all $n \in \mathbb{N}$*

$$e_{4n+1}^{\rm cont}(S,F,H) \le 2\,C\,\|\,S\,\|^2\,\|\,S^{-1}\,\|^2\;e_{n,C}^{\rm non}(S,F,H)\,. \tag{22}$$

Finally we collect some further properties of the quantities $e_n^{\rm cont}$ and $e_n^{\rm non}$.

Lemma 3. (i) *Let $m,n \in \mathbb{N}$, and let F be a subset of the quasi-normed linear space X, where X itself is a subset of the quasi-normed linear space Y. Denote by I_j identity operators. Then*

$$e_{m+n}^{\rm cont}(I_1,F,Y) \le e_m^{\rm cont}(I_2,F,X)\,e_n^{\rm cont}(I_3,X,Y) \tag{23}$$

holds.
(ii) *Let F be a quasi-normed subset of G and let $I: F \to G$ be the identity. Then*

$$e_n^{\rm cont}(I,F,G) \le \|S^{-1}\|\,e_n^{\rm cont}(S,F,H) \le \|S^{-1}\|\,\|S\|\,e_n^{\rm cont}(I,F,G) \tag{24}$$

and for any $C \ge \|S^{-1}\|\,\|S\|$

$$e_{n,C\,\|S^{-1}\|\,\|S\|}^{\rm non}(I,F,G) \le \|S^{-1}\|\,e_{n,C}^{\rm non}(S,F,H) \le \|S^{-1}\|\,\|S\|\,e_{n,C/(\|S^{-1}\|\,\|S\|)}^{\rm non}(I,F,G) \tag{25}$$

holds.

Remark 5. *Let us point out the following which is part of the proof of Lemma 3. Let $\mathcal{B} = \{h_1, h_2, \dots\}$ be a Riesz basis of G. Let S_n be an approximation of the identity $I : F \to G$. Then $S(\mathcal{B})$ is a Riesz basis of H and $S \circ S_n$ is an approximation of $S : F \to H$ satisfying*

$$\|f - S_n(f)\|_G \le \|S^{-1}\| \cdot \|\,Sf - S\circ S_n(f)\,\|_H \le \|S^{-1}\| \cdot \|\,S\,\| \cdot \|\,f - S_n(f)\,\|_G\,. \tag{26}$$

This makes clear that if $\mathcal{B}$ and S_n are order optimal for the triple I,F,G, then $S(\mathcal{B})$ and $S\circ S_n$ are order optimal for the triple S,F,H. Consequently, instead of looking for good approximations of $S : F \to H$ it will be enough to study approximations of the identity $I : F \to G$.

Remark 6. *The assertion in part* (i) *of the Lemma is essentially proved in* [39] *but traced there to Khodulev. The inequality* (23) *can be made more transparent by means of the diagram*

$$\begin{array}{ccc} X & \stackrel{I_3}{\longrightarrow} & Y \\ I_2 \ \nwarrow & & \nearrow \ I_1 \\ & F. & \end{array}$$

Remark 7. *The approximation numbers e_n^{lin}, the Gelfand widths d^n, the manifold widths e_n^{cont} and Bernstein widths b_n are particular examples of s-numbers in the sense of* Pietsch [68], *cf.* [57] *for the manifold widths. They have some properties in common, e.g. the multiplicativity: with s_n instead of e_n^{lin}, d^n, e_n^{cont} and b_n we have*

$$s_n(T_2 \circ T_1 \circ T_0) \le \| T_0 \| \, \| T_2 \| \, s_n(T_1)\,, \tag{27}$$

where $T_0 \in \mathcal{L}(E_0, E)$, $T_1 \in \mathcal{L}(E, F)$, $T_2 \in \mathcal{L}(F, F_0)$ and E_0, E, F, F_0 are arbitrary Banach spaces. For these four types of s-numbers the assertion remains true also for quasi-Banach spaces.

Another property concerns additivity. For s_n instead of e_n^{lin} and d^n we have

$$s_{2n}(T_0 + T_1) \le c\,\Big(s_n(T_0) + s_n(T_1)\Big)\,, \tag{28}$$

where $T_0, T_1 \in \mathcal{L}(E, F)$, E, F are arbitrary quasi-Banach spaces, and c does not depend on n, T_0, T_1, cf. [10]. *In case that F is a Banach space one can take $c = 1$.*

3 Optimal Approximation of Elliptic Problems

Let $s, t > 0$. We consider the diagram

$$\begin{array}{ccc} H^{-s}(\Omega) & \stackrel{S}{\longrightarrow} & H^s_0(\Omega) \\ I \ \nwarrow & & \nearrow \ S_t \\ & B^{-s+t}_q(L_p(\Omega)), & \end{array}$$

where S_t denotes the restriction of S to $B^{-s+t}_q(L_p(\Omega))$ and I denotes the identity. We assume (3) and $S = \mathcal{A}^{-1}$.

3.1 Optimal Linear Approximation of Elliptic Problems

Theorem 2. *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $0 < p, q \le \infty$, $s > 0$, and*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+ . \tag{29}$$

Then

$$e_n^{\text{lin}}(S, B_q^{-s+t}(L_p(\Omega))), H_0^s(\Omega)) \asymp \begin{cases} n^{-t/d} & \text{if} \quad 2 \le p \le \infty\,, \\ n^{-\frac{t}{d}+\frac{1}{p}-\frac{1}{2}} & \text{if} \quad 0 < p < 2\,. \end{cases}$$

Remark 8. *i) The restriction* (29) *is necessary and sufficient for the compactness of the embedding* $I : B_q^{-s+t}(L_p(\Omega)) \hookrightarrow H^{-s}(\Omega)$, *cf. the Appendix, Proposition* 7.

ii) The proof is constructive. First of all one has to determine a linear mapping S_n *which approximates the identity* $I : B_q^{-s+t}(L_p(\Omega)) \to H^{-s}(\Omega)$ *with the optimal order. How this can be done is described in Remark* 25, *Subsection* 4.3.3. *Finally, the linear mapping* $S \circ S_n$ *realizes an in order optimal approximation of* S_t.

iii) There are hundreds of references dealing with approximation numbers of linear operators. Most useful for us have been the monographs [42, 69, 70, 78, 76, 87] *to which we refer also for further references.*

3.2 Optimal Nonlinear Approximation of Elliptic Problems

To begin with, we consider the manifold and the Gelfand widths. There we have a rather final answer.

Theorem 3. *Let* $\Omega \subset \mathbb{R}^d$ *be a bounded Lipschitz domain. Let* $0 < p, q \le \infty$, $s > 0$, *and*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+ \,.$$

Then

$$e_n^{\text{cont}}(S, B_q^{-s+t}(L_p(\Omega))), H_0^s(\Omega)) \asymp n^{-t/d}\,.$$

If, in addition, $p \ge 1$ *(and* $t > d/2$ *if* $1 \le p < 2$*), then*

$$d^n(S, B_q^{-s+t}(L_p(\Omega))), H_0^s(\Omega)) \asymp n^{-t/d}\,.$$

From Theorem 1 and Theorem 3 we conclude that the order of $e_{n,C}^{\text{non}}$ is also at least $n^{-t/d}$. For the respective upper bound of the nonlinear widths $e_{n,C}^{\text{non}}$ we need a few more restrictions with respect to the domain Ω. Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$ and let $s > 0$. We assume that for any fixed triple (t, p, q) of parameters the spaces $B_q^{-s+t}(L_p(\Omega))$ and $H^{-s}(\Omega)$ allow a discretization by one common wavelet system $\mathcal{B}^*$, i.e. (106) - (111) should be satisfied with $B_q^{-s+t}(L_p(\Omega))$ and $B_2^{-s}(L_2(\Omega))$, respectively, cf. Appendix 5.10. By assumption such a wavelet system belongs to $\mathcal{B}_{C^*}$ for some $1 \le C^* < \infty$.

Theorem 4. *Under the above conditions on Ω and if $0 < p, q \le \infty$, $s > 0$, $t > d(\frac{1}{p} - \frac{1}{2})_+$, we have for any $C \ge C^*$*

$$e^{\text{non}}_{n,C}(S, B^{-s+t}_q(L_p(\Omega))), H^s_0(\Omega)) \asymp n^{-t/d}\,.$$

Remark 9. *Comparing Theorems 3, 4 and Theorem 2 there is a clear message. For $p < 2$ there are nonlinear approximations which are better in order than any linear approximation.*

Remark 10. *The proof of the upper bound in Theorem 4 is* constructive in a theoretical sense *that we now describe. Given a right-hand side $f \in B^{-s+t}_q(L_p(\Omega))$ we have to calculate all wavelet coefficients $\langle f, \widetilde{\psi}_{j,\lambda}\rangle$. The sequence of these coefficients belongs to the space $b^{-s+t}_{p,q}(\nabla)$, cf. Subsection 4.2. With*

$$a = (a_{j,\lambda})_{j,\lambda}\,, \qquad a_{j,\lambda} := \langle f, \widetilde{\psi}_{j,\lambda}\rangle\,, \quad \textit{for all} \quad j, \lambda\,,$$

we find a good approximation $S_n(a)$ of a with n components with respect to the norm $\| \cdot |b^s_{2,2}(\nabla)\|$ in Proposition 2. To get an optimal approximation of the solution $u = Sf$ in $\| \cdot |H^s(\Omega)\|$ we have to apply the solution operator to $S_n(a)$. Hence

$$u_n = (S \circ S_n)(a) = \sum_{j=0}^{K} \sum_{\lambda \in \Lambda^*_j} a^*_{j,\lambda}\, S\psi_{j,\lambda}\,, \tag{30}$$

where $K = K(a, n)$, $a^*_{j,\lambda}$ and Λ^*_j are as in Proposition 2, cf. in particular (61) and (64), represents such a good approximation of u. To calculate u_n a lot of computations have to be done. The coefficients $a^*_{j,\lambda}$ are the largest in a weighted sense (the weight depends on n and j, cf. the proof of Proposition 2 for explicit formulas). Having these coefficients at hand one has finally to solve all the equations

$$\mathcal{A}u_{j,\lambda} = \psi_{j,\lambda}\,, \qquad 0 \le j \le K, \quad \lambda \in \Lambda^*_j \tag{31}$$

to obtain $u_{j,\lambda} = S\psi_{j,\lambda}$. The number of equations is $O(n)$.

In this way we obtain a nonlinear approximation with respect to the Riesz basis given by the $S\psi_{j,\lambda}$. Observe that this Riesz basis depends on the operator equation. It would be much better to use a known Riesz basis, such as a wavelet basis, that does not depend on $\mathcal{A}$. See Theorem 5 for a step into that direction.

Remark 11. *At least if Ω is a cube, all required properties are known to be satisfied if in addition $1 < p, q < \infty$. The latter restriction allows to use duality arguments, cf. Proposition 10 in Appendix 5.8. There also exist results for domains with piecewise analytic boundary such as polygonal or polyhedral domains. One natural way as,* *e.g., outlined in* [8] *and* [25]*, is to decompose the domain into a disjoint union of parametric images of reference cubes. Then, one constructs wavelet bases on the reference cubes and glues everything together in a judicious fashion. However, due to the glueing procedure, only Sobolev spaces* H^s *with smoothness* $s < 3/2$ *can be characterized. This bottleneck can be circumvented by the approach in* [26]*. There a much more tricky domain decomposition method involving certain projection and extension operators is used. By proceeding this way, norm equivalences for all spaces* $B^t_q(L_p(\Omega))$ *can be derived, at least for the case* $p > 1$*, see* [26]*, Theorem* 3.4.3*. However, the authors also mention that their results can be generalized to the case* $p < 1$*, see* [26]*, Remark* 3.1.2.

Remark 12. *Comparing Theorems* 3 *and* 4 *we see that the numbers* $e^{\mathrm{non}}_{n,C}$*,* e^{cont}_n*, and* d^n *have the same asymptotic behaviour, at least for* $p > 1$*. Using the relation* $d^n \asymp r_n$*, see Lemma* 1*, we actually can get the optimal order* $n^{-t/d}$ *with an approximation of the form*

$$f \mapsto S \circ \varphi_n \circ N_n(f)\,, \tag{32}$$

where

$$N_n : B_q^{-s+t}(L_p(\Omega)) \to \mathbb{R}^n$$

is linear and does not depend on S *(this mapping gives the* information *that is used about the right hand side),*

$$\varphi_n : \mathbb{R}^n \to H^{-s}(\Omega)$$

is nonlinear but also does not depend on S*. The mapping* $\varphi_n \circ N_n$ *gives a good approximation of the embedding from* $B_q^{-s+t}(L_p(\Omega))$ *to* H^{-s}*.*

Remark 13. *There is a further little difference between linear and nonlinear approximation. Let us consider the limiting case* $t = d(\frac{1}{p} - \frac{1}{2})$*,* $0 < p < 2$*. Then the embedding* $B_p^{-s+t}(L_p(\Omega)) \hookrightarrow H^{-s}(\Omega)$ *is continuous, not compact. As a consequence*

$$e_n^{\mathrm{lin}}(S, B_p^{-s+t}(L_p(\Omega))), H_0^s(\Omega)) \not\to 0 \qquad \text{if} \quad n \to \infty\,,$$

but

$$e_n^{\mathrm{non}}(S, B_p^{-s+t}(L_p(\Omega))), H_0^s(\Omega)) \to 0 \qquad \text{if} \quad n \to \infty\,,$$

cf. Remark 23.

3.3 The Poisson Equation

The next step is to discuss the specific case of the Poisson equation on a Lipschitz domain Ω contained in $\mathbb{R}^2$

$$\begin{aligned} -\triangle u &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega. \end{aligned} \tag{33}$$

As usual, we study (33) in the weak formulation. Then, it can be shown that the operator $\mathcal{A} = \triangle : \; H_0^1 \longrightarrow H^{-1}$ is boundedly invertible, see, e.g., [49] for details. Hence Theorems 2 and 3 apply with $s = 1$, for the upper bound of Theorem 4 we need some restrictions with respect to Ω. For the proof of Theorem 4 we used the Riesz basis $S\psi_{j,\lambda}$ which depends on $\mathcal{A}$. Now we want to approximate the solution u by wavelets.
We shall restrict ourselves to the case that Ω is a simply connected polygonal domain. The segments of $\partial\Omega$ are denoted by $\overline{\Gamma}_l, \Gamma_l$ open, $l = 1, \ldots, N$ numbered in positive orientation. Furthermore, Υ_l denotes the endpoint of Γ_l and ω_l denotes the measure of the interior angle at Υ_l. For such a domain appropriate wavelet systems can be constructed, see Remark 11. Then we obtain the following.

Theorem 5. *Let Ω be a polygonal domain in $\mathbb{R}^2$. Let $1 < p \leq 2$ and let $k \geq 1$ be a nonnegative integer such that*

$$\frac{m\pi}{\omega_l} \neq k + 1 - \frac{2}{p} \qquad \textit{for all} \quad m \in \mathbb{N},\; l = 1, \ldots, N.$$

Then, for an appropriate wavelet system $\mathcal{B}^*$ the best n-term approximation of problem (33) *yields*

$$\sup_{\|f|B_p^{k-1}(L_p(\Omega))\| \leq 1} \sigma_n(u, \mathcal{B}^*) \leq c_\varepsilon \, n^{-k/2+\varepsilon} \tag{34}$$

where $\varepsilon > 0$ and c_ε do not depend on n.

Remark 14. *There is an important difference to the approximation described in Remark* 10*. Here we can work with one given wavelet system to approximate the solution u. We are not forced to work with the solutions of the system* (31)*. A more detailed discussion of these relationships including possible numerical realizations of wavelet methods will follow in Section* 3.4*.*

3.4 Algorithms and Complexity

So far we studied the error $e(S_n, F, H)$ of approximations S_n. We compared nonlinear S_n and linear S_n with respect to their errors and proved results on the optimal rate of convergence. We assume that (1) is a given and fixed operator equation and hence, in the case of (3), also Ω is fixed.

In this section we briefly discuss algorithms and their complexity, and for simplicity we still assume that the operator equation (3) is given and fixed. Observe that in practice it is important to construct also algorithms for more general problems: We want to input information about Ω and $\mathcal{A}$ and the right hand side f, and we want to obtain an ε-approximation of the solution u. In our more restricted case we *only* have to input information concerning the right hand side f because Ω and $\mathcal{A}$ are fixed.

We use, as usual in numerical analysis, the real number model of computation (see [61] for the details and [63] and [64] for further comments). Any algorithm computes and/or uses some information (consisting in finitely many numbers) describing the right hand side f of (3). There are different ways how an algorithm may use information concerning f, we describe two of them in turn.

1. The used information about f is very explicit if S_n is linear (5): Then the algorithm uses $L_1(f), \dots, L_n(f)$ and we assume that we have an oracle (or subroutine) for the $L_i(f)$. In practical applications the computation of a functional $L_i(f)$ can be very easy or very difficult or anything between. One often assumes that the cost of obtaining a value $L_i(f)$ is c where $c > 0$ is small or large, depending on the circumstances.

   Now we imagine S_n as in (11) as the input-output mapping of a numerical algorithm: on input $f \in F$ we obtain the output $S_n(f) = u_n = \sum_{k=1}^n c_k \, h_{i_k}$. More formally we should say that the output is

   $$\text{out}(f) = (i_1, c_1, i_2, c_2, \dots, i_n, c_n) \tag{35}$$

   but we identify $\text{out}(f)$ with u_n. Of course we cannot consider arbitrary mappings S_n of the form (11) as the input-output mapping of an algorithm, since not all such S_n are computable.

   We still assume that we only have an oracle for the computation of linear functionals $L_i(f)$. Then it is not so clear what the information cost of (11) is, since (11) only describes the (desired) output of an algorithm, it is not an algorithm by itself. We need an algorithm that uses information $L_1(f), \dots, L_N(f)$, where N might be bigger than n, to produce the i_k and the c_k of $\text{out}(f)$. The information cost of such a procedure would be cN.

2. One also can assume that a good approximation f_n can easily be precomputed

with negligible cost. Hence the algorithm starts with an approximation

$$f_n = \sum_{k=1}^{n} c_k \, g_{i_k}, \tag{36}$$

such as a best n-term approximation (or a greedy approximation) of f with respect to a basis $\{g_i,\ i \in \mathbb{N}\}$.

This is a good place for a short remark about adaption. The use of *adaptive methods* is quite widespread but we want to stress that the notion of adaptive methods is not uniformly used in the literature. Some confusion is almost unavoidable if different such notions are mixed. To avoid such a confusion we do not use the notion of an "adaptive method". Instead we speak first about *adaptive (or nonadaptive) information* and then about *adaptive numerical schemes*.

- *Nonadaptive information*: The algorithm uses certain functionals $L_1, L_2, \ldots, L_n$ and for each input $f \in F$ the algorithm needs $L_1(f), L_2(f), \ldots, L_n(f)$, hence the functionals L_i do not depend on f. In this case we say that the algorithm uses *nonadaptive information*.
- *Adaptive information*: The algorithm uses $L_1(f)$ and, depending on this number, the next functional L_2 is chosen. In general, the chosen functional L_k may depend on the values $L_1(f), \ldots, L_{k-1}(f)$ that are already known to the algorithm. Observe that L_k cannot depend in an arbitrary way on f since the algorithm can only use the known information about f. In this case we say that the algorithm uses *adaptive information*.

We give an example. Assume that a certain S_n of the form (11) can be realized in such a way that we first compute $L_1(f), \ldots, L_N(f)$, where the L_i do not depend on $f \in F$. In the further parts of the algorithm we only use the $L_i(f)$ for the n largest values of $|L_i(f)|$, together with the respective values of i, to compute the output $\mathrm{out}(f)$. Such an algorithm uses nonadaptive information (of cardinality N), the information cost is cN.

There is a large stream of results under which conditions adaptive information is superior (or not superior) compared to nonadaptive information, we mention the pioneering paper by Bakhvalov [2], the results on operator equations by Gal and Micchelli [43] and by Traub and Woźniakowski [79] and the survey [62]. It is known, for example, that adaptive information does not help (up to a factor of 2) for linear operator equations and the worst case error with respect to the unit ball of a normed space F. If F is only quasi-normed then the proofs must be modified, with a possible change of the constant 2. Nevertheless nonadaptive information is almost as good as adaptive information.

How much information is needed about the right hand side $f \in F$ in order that we can solve the equation (1) with an error ε? This question is answered by the minimal radii of information $r_n(S, F, H)$ (or the closely related Gelfand numbers). These numbers are a good measure for the *information complexity* of the operator equation. In contrast, the *output complexity* of the problem is measured by the nonlinear widths $e_{n,C}^{\mathrm{non}}(S, F, H)$. These numbers measure the cost of just outputting the approximation (with respect to an optimal basis $\mathcal{B} \in \mathcal{B}_C$). It is quite remarkable that, under general conditions, we obtain the same order

$$r_n(S, F, H) \asymp d^n(S, F, H) \asymp e_{n,C}^{\mathrm{non}}(S, F, H) \asymp n^{-t/d},$$

see Theorem 3 and Theorem 4.

Now we discuss *adaptive numerical schemes* for the numerical treatment of elliptic partial differential equations. Usually, these operator equations are solved by a Galerkin scheme, i.e., one defines an increasing sequence of finite dimensional approximation spaces $G_{\Lambda_l} := \mathrm{span}\{\eta_\mu : \mu \in \Lambda_l\}$, where $G_{\Lambda_l} \subset G_{\Lambda_{l+1}}$, and projects the problem onto these spaces, i.e.,

$$\langle \mathcal{A} u_{\Lambda_l}, v \rangle = \langle f, v \rangle \quad \text{for all} \quad v \in G_{\Lambda_l}.$$

To compute the actual Galerkin approximation, one has to solve a linear system

$$\mathbf{A}_{\Lambda_l} \mathbf{c}_{\Lambda_l} = \mathbf{f}_{\Lambda_l}, \qquad \mathbf{A}_{\Lambda_l} = (\langle A\eta_{\mu'}, \eta_\mu \rangle)_{\mu,\mu' \in \Lambda_l}, \qquad (\mathbf{f}_\Lambda)_\mu = \langle f, \eta_\mu \rangle, \ \mu \in \Lambda_l.$$

Then the question arises how to choose the approximation spaces in a suitable way, for doing that in a somewhat clumsy fashion would yield huge linear systems and a very unefficient scheme. One natural way would be to use an updating strategy, i.e., one starts with a small set Λ_0, tries to estimate the (local) error, and only in regions where the error is large the index set is *refined*, i.e., further basis functions are added. Such an updating strategy is usually called an *adaptive numerical scheme* and it is characterized by the following facts: the sequence of approximation spaces is not a priori fixed but depends on the *unknown* solution u of the operator equation, and the whole scheme should be self-regulating, i.e., it should work without a priori information on the solution. In principle, such an adaptive scheme consists of the following three steps:

solve	–	estimate	–	refine
$\mathbf{A}_{\Lambda_l} \mathbf{c}_{\Lambda_l} = \mathbf{f}_{\Lambda_l}$		$\lVert u - u_{\Lambda_l} \rVert = ?$ a posteriori error estimator		add functions if necessary.

Already the second step is highly nontrivial since the exact solution u is unknown, so that clever a posteriori error estimators are needed. These error estimators should be local since only in regions where the local error is large we want to refine, i.e., to add functions. Then another challenging task is to show that the refinement strategy leads to a convergent scheme and to estimate its order of convergence, if possible.

Recent developments, for instance in the finite element context, indeed indicate the promising potential of adaptive numerical schemes, see, e.g., [1, 3, 4, 5, 38, 86]. However, to further explain the ideas and to make comparisons as simple as possible, we shall restrict ourselves to adaptive schemes based on wavelets. For simplicity, we shall primary discuss the approach in [20], for more sophisticated versions the reader is referred to [12, 13, 14, 21]. The first step clearly must be the development of an a posteriori error estimator. Using the fact that $\mathcal{A}$ is boundedly invertible and the usual norm equivalences, compare with (111), we obtain

$$(37)\qquad \begin{aligned} \|u-u_\Lambda\|_{H^s} &\asymp \|\mathcal{A}(u-u_\Lambda)\|_{H^{-s}} \\ &\asymp \|f-\mathcal{A}(u_\Lambda)\|_{H^{-s}} \\ &\asymp \|r_\Lambda\|_{H^{-s}} \\ &\asymp \left(\sum_{(j,\lambda)\in J\setminus\Lambda} 2^{-2sj}|\langle r_\Lambda,\psi_{j,\lambda}\rangle|^2\right)^{1/2} \\ &= \left(\sum_{(j,\lambda)\in J\setminus\Lambda} \delta_{j,\lambda}^2\right)^{1/2}, \end{aligned}$$

where the *residual weights* $\delta_{j,\lambda}$ can be computed as

$$\delta_{j,\lambda}=2^{-sj}|f_{j,\lambda}-\sum_{(j',\lambda')\in\Lambda}\langle\mathcal{A}\psi_{j',\lambda'},\psi_{j,\lambda}\rangle u_{j',\lambda'}|,\qquad f_{j,\lambda}=\langle f,\psi_{j,\lambda}\rangle.$$

From (37) we observe that the sum of the residual weights gives rise to an efficient and reliable a posteriori error estimator. Each residual weight $\delta_{j,\lambda}$ can be interpreted as a local error indicator, so that the following natural refinement strategy suggests itself: add wavelets in regions where the residual weights are large, that is, try to catch the bulk of the residual expansion in (37). Indeed, it can be shown that this stratey produces a convergent adaptive scheme, in principle. However, we are faced with a serious problem: the index set J will not have finite cardinality, so that neither the error estimator nor the adaptive refinement strategy can be implemented. Nevertheless, there exists implementable variants, see again [12, 20] for details. We start with the set

$$J_{j,\lambda,\varepsilon}:\ \{(j',\lambda')|\ |\langle\mathcal{A}\psi_{j',\lambda'},\psi_{j,\lambda}\rangle|\ \varepsilon\text{-significant}\}$$

and define

$$a_{j,\lambda}(\Lambda,\varepsilon) := 2^{-sj} | \sum_{(j',\lambda')\in\Lambda\cap J_{j,\lambda,\varepsilon}} \langle \mathcal{A}\psi_{j',\lambda'}, \psi_{j,\lambda}\rangle u_{j',\lambda'}|.$$

(The expresssion 'ε-significant' can be made precise by using the locality and the cancellation properties of a wavelet basis). By employing the $a_{j,\lambda}(\Lambda,\varepsilon)$ we obtain another error erstimator:

$$\|u-u_\Lambda\| \lesssim \left(\sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}^2\right)^{1/2} + \varepsilon\|f\|_{H^{-s}} + \inf_{v\in\tilde{V}_\Lambda} \|F-v\|_{H^{-s}}.$$

Here $\tilde{V}_\Lambda$ denotes the approximation space spanned by the dual wavelets corresponding to Λ, see Section 5.3 for details. Now, playing the same game for the $a_{j,\lambda}(\Lambda,\varepsilon)$ instead of the $\delta_{j,\lambda}$, we end up with a convergent and implementable adaptive strategy. For the proof and further details, the reader is again refered to [20].

Theorem 6. *Choose a final accuracy eps and a parameter $\theta \in (0,1)$. Determine Λ such that $\inf_{v\in\tilde{V}_\Lambda} \|f-v\|_{H^{-s}} \lesssim$ eps and compute $\varepsilon(f, eps, \theta)$. Then, whenever $\tilde{\Lambda}\subset J,\ \Lambda\subset\tilde{\Lambda}$ is chosen so that*

$$\left(\sum_{(j,\lambda)\in\tilde{\Lambda}\setminus\Lambda} a_{j,\lambda}(\Lambda,\varepsilon)^2\right)^{1/2} \geq (1-\theta)\left(\sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}(\Lambda,\varepsilon)^2\right)^{1/2} \tag{38}$$

there exists a constant $\kappa\in(0,1)$ such that either

$$\|u-u_{\tilde{\Lambda}}\| \leq \kappa\|u-u_\Lambda\|, \qquad \kappa\in(0,1) \tag{39}$$

or

$$\left(\sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}(\Lambda,\varepsilon)^2\right)^{1/2} \leq eps \tag{40}$$

which implies that

$$\|u-u_\Lambda\| \lesssim eps. \tag{41}$$

Remark 15. *i) The norm $\|\cdot\|$ in* (39) *and* (41) *clearly denotes the energy norm $\|v\| := \langle\mathcal{A}v, v\rangle$ which is equivalent to the Sobolev norm H^1, see again* [49] *for details.*

ii) Theorem 6 *obviously implies that the adaptive strategy in* (38) *converges. Indeed, the error is reduced by κ step by step until the sum of the significant coefficients in* (40) *is smaller than the final accuracy, which by* (41) *means that the same property holds for the current Galerkin approximation.*

iii) Although the sum in the right-hand side of (38) *formally still contains unfinitely many coefficients, it can be checked that this sum in fact runs over a finite set, so that the adaptive strategy is implementable.*

Let us now compare this concept of adaptivity with the notion of adaptive information explained above:

- From Theorem 6 we observe that adaptive wavelet schemes are not performed by gaining more and more information from the right-hand side f in an adaptive fashion. Instead they use the *residual* which depends on the right-hand side, the operator, and the domain. Moreover, we see that the starting index set Λ in Theorem 6 is determined by the wavelet expansion of the right-hand side. That is, Λ is given by some kind of best n-term approximation of f which is assumed to be available or to be easily computable. In this sense, the adaptive wavelet schemes require *nonlinear information* of the problem.

- In the wavelet setting, the benchmark for the performance is the approximation order of the best n-term approximation of the solution, i.e., the numbers

$$\sup_{\|f\|_F \leq 1} \sigma_n(\mathcal{A}^{-1} f, \mathcal{B})_H.$$

  Quite recently, it has been shown in [12] that a judicious variant of the algorithm in Theorem 6 indeed gives rise to the same order of approximation as best n-term approximation while the arithmetic operations that are needed stay proportional to the number of unknowns. Here the authors assume that u is in a certain Besov space $B_p^\alpha(L_p(\Omega))$, hence $F = \mathcal{A}(B_p^\alpha(L_p(\Omega)))$, i.e., the admissible class of right hand sides depends on the operator $\mathcal{A}$.

- The performance of an adaptive scheme is not compared with an *arbitrary linear* scheme. The reason for that is simple and has already been explained earlier. It is indeed true that linear approximation often produces the same order as nonlinear (best n-term) approximations, compare with Theorem 2 and Theorem 4. However, for nonregular problems, it would be necessary to precompute the optimal basis $S(g_i)$ in advance which is mostly too expensive and should be avoided in practice, see [23] for further details. One usually compares adaptive schemes with *uniform* methods for then a precomputation is not necessary. Therefore the use of an adaptive wavelet scheme is justified if it performs better than any uniform scheme. It is known that the order of approximation of uniform schemes is determined by the Sobolev regularity $H^t(\Omega)$ of the object we want to approximate whereas the approximation order of best n-term approximation depends on the regularity in the specific Besov scale $B^t_\tau(L_\tau(\Omega))$, $\frac{1}{\tau} = \frac{t-s}{d} + \frac{1}{2}$, see [19, 28] for details. Therefore adaptive schemes are justified if and only if the Besov regularity of the exact solution is higher than its Sobolev regularity. For elliptic boundary value problems, there exist by now many results in this direction, see, e.g., [15, 16, 17, 18, 22].

- In approximation theory, an approximation scheme that comes from a sequence of linear spaces that are uniformly refined is also called *linear approximation scheme* which sometimes causes misunderstandings because these schemes are only special cases of the linear schemes considered, e.g., in Theorem 4. To avoid this confusion, we used the term uniform methods instead of linear methods.

4 Proofs

4.1 Properties of Widths

Proof of Lemma 1. *Step 1.* Part (i) is proved in [80] for the case where F is normed. The general case is similar.
Step 2. To prove part (ii) we assume that $S(\{\|f\|_F \le 1\})$ contains an $(n+1)$-dimensional ball $B \subset H$ of radius r and that $N_n : F \to \mathbb{R}^n$ is continuous. Since $S^{-1}(B)$ is an $(n+1)$-dimensional bounded and symmetric neighborhood of 0, it follows from the Borsuk Antipodality Theorem, see [27, paragraph 4], that there exists an $f \in \partial S^{-1}(B)$ with $N_n(f) = N_n(-f)$ and hence

$$S_n(f) = \varphi_n(N_n(f)) = \varphi_n(N_n(-f)) = S_n(-f)$$

for any mapping $\varphi_n : \mathbb{R}^n \to G$. Observe that $\|f\|_F = 1$. Because of $\|S(f)-S(-f)\| = 2r$ and $S_n(f) = S_n(-f)$ we obtain that the maximal error of S_n on $\{\pm f\}$ is at least r. This proves

$$b_n(S,F,H) \le e_n^{\text{cont}}(S,F,H)\,.$$

Since we did not use the continuity of φ_n also $b_n(S,F,H) \le d^n(S,F,H)$ follows. □

Proof of Lemma 3. *Step 1.* Proof of (i). A corresponding assertion with X and Y normed linear spaces has been proved in [39]. This proof carries over without changes.
Step 2. Proof of (25). Let $\mathcal{B} = \{h_1, h_2, \dots\}$ be a Riesz basis of G with Riesz constants $A, B > 0$. Let this basis $\mathcal{B}$ and a corresponding mapping S_n be optimal with respect

to I, F, G (up to some $\varepsilon > 0$ if necessary). Then the image of $\mathcal{B}$ under the mapping S is also a Riesz basis, now of H and with Riesz constants $A' = A/\|S^{-1}\|$ and $B' = B\,\|S\|$. From

$$\| Sf - (S \circ S_n)f \|_H \le \|S\| \, \| f - S_n(f) \|_G$$

it follows

$$e^{\text{non}}_{n,C\,\|S^{-1}\|\,\|S\|}(S, F, H) \le \|S\| \, e^{\text{non}}_{n,C}(I, F, G) \, .$$

Replacing C by $C/(\|S^{-1}\|\,\|S\|)$ the right-hand side in (25) follows.
Now, let $\mathcal{B} \subset H$ be a Riesz basis with Riesz constants $A, B > 0$. Let $\mathcal{B}$ and a corresponding S_n be optimal with respect to S, F, H (again up to some $\varepsilon > 0$ if necessary). From

$$\| If - (S^{-1} \circ S_n)f \|_G \le \|S^{-1}\| \, \| Sf - S_n(f) \|_H$$

it follows

$$e^{\text{non}}_{n,C\,\|S^{-1}\|\,\|S\|}(I, F, G) \le \|S^{-1}\| \, e^{\text{non}}_{n,C}(S, F, H) \, .$$

The proof of (24) follows from (27). □

Next we turn to the proof of Theorem 1. It is convenient for us to start with a simplified situation. For this we assume that $K \subset H$ is compact. We define

$$e^{\text{non}}_{n,C}(K, H) = \inf_{\mathcal{B} \in \mathcal{B}_C} \sup_{u \in K} \sigma(u, \mathcal{B}) \tag{42}$$

and

$$e^{\text{cont}}_{n}(K, H) = \inf_{N_n, \varphi_n} \sup_{u \in K} \|\varphi_n(N_n(u)) - u\|, \tag{43}$$

where the infimum runs over all continuous mappings $\varphi_n : \mathbb{R}^n \to H$ and $N_n : K \to \mathbb{R}^n$. We prove the following result.

Proposition 1. *Let $K \subset H$ be compact. Then*

$$e^{\text{cont}}_{4n+1}(K, H) \le 2C \, e^{\text{non}}_{n,C}(K, H). \tag{44}$$

Proof. Let $\mathcal{B} \in \mathcal{B}_C$ be given. Since K is compact, we only need finitely many of the elements of $\mathcal{B}$ in the sense that

$$\sup_{u \in K} \|u - L_N(u)\| \le \varepsilon \tag{45}$$

for

$$L_N(u) = \sum_{j=1}^{N} a_j h_j. \tag{46}$$

Here L_N is the orthogonal projection onto the space that is generated by $h_1, \dots, h_N$. The functionals a_j are linear and continuous. Moreover, we know that

$$A \left(\sum_{j=1}^{N} |\alpha_j|^2 \right)^{1/2} \le \| \sum_{j=1}^{N} \alpha_j h_j \| \le B \left(\sum_{j=1}^{N} |\alpha_j|^2 \right)^{1/2} \tag{47}$$

with $B/A \le C$. We may assume that $A = 1$. For a suitable $\mathcal{B} \in \mathcal{B}_C$ we obtain

$$\sup_{u \in K} \| \sum_{k=1}^{n} c_k \, h_{i_k} - L_N(u) \| \le e_{n,C}^{\text{non}}(K, H) + \varepsilon. \tag{48}$$

Let $\beta > 0$. We define a modification of L_N by

$$L_N^*(u) = \sum_{j=1}^{N} a_j^* h_j \tag{49}$$

where $a_j^* = a_j$ if $|a_j| \ge 2\beta$ and $a_j^* = 0$ if $|a_j| \le \beta$. To make the a_j^* continuous we define elsewhere

$$a_j^* = 2 \, \text{sgn}(a_j) \cdot (|a_j| - \beta).$$

We prove certain statements about L_N^* and denote the best n-term approximation of u by u_n.

Assume that, for $u \in K$, there are $m > n$ of the a_j, see (46), such that $|a_j| \ge \beta$. Then we obtain

$$\|u_n - L_N(u)\| \ge (m - n)^{1/2} \beta$$

and with (48) we obtain

$$m - n \le \frac{1}{\beta^2} (e_n^{\text{non}}(K, H) + \varepsilon)^2. \tag{50}$$

Now we consider the sum $\sum_{|a_j| < \beta} a_j^2$ for $u \in K$. We distinguish between those j that are used for u_n (there are only n of those j) and the other indices and obtain

$$\sum_{|a_j|^2 < \beta} a_j^2 \le n\beta^2 + (e_n^{\text{non}}(K, H) + \varepsilon)^2.$$

Now we are ready to estimate $\|L_N^*(u) - L_N(u)\|$ for $u \in K$. Observe that $|a_j^* - a_j| \le \beta$ for any j. We obtain

$$\|L_N^*(u) - L_N(u)\| \le B(m\beta^2 + n\beta^2 + (e_n^{\text{non}}(K, H) + \varepsilon)^2)^{1/2}.$$

Using the estimate (50) for m we obtain

$$\|L_N^*(u) - L_N(u)\| \le B(2n\beta^2 + 2(e_n^{\rm non}(K,H)+\varepsilon)^2)^{1/2}.$$

Now we define β by

$$n\beta^2 = (e_n^{\rm non}(K,H)+\varepsilon)^2$$

and obtain the final error estimate (where we replace, for general A, the number B by B/A)

$$\|L_N^*(u) - L_N(u)\| \le 2\frac{B}{A}(e_n^{\rm non}(K,H)+\varepsilon).$$

In addition we obtain

$$m \le 2n$$

and therefore L_N^* yields a continuous $2n$-term approximation of $u \in K$ with error at most

$$\sup_{u\in K} \|L_N^*(u) - u\| \le 2\frac{B}{A}(e_n^{\rm non}(K,H)+\varepsilon) + \varepsilon.$$

The mapping L_N^* is continuous and the image is a complex of dimension $2n$, cf. e.g. [30]. Hence we have an upper bound for the so-called Aleksandrov widths, see [30] and [75] for a discussion of these widths. By the famous theorem of Nöbeling, any such mapping can be factorized as $L_N^* = \varphi_{4n+1} \circ N_{4n+1}$ where $N_{4n+1} : K \to \mathbb{R}^{4n+1}$ and $\varphi_{4n+1} : \mathbb{R}^{4n+1} \to H$ are continuous. Hence the result is proved. □

Proof of Theorem 1. The unit ball of F is a compact subset of G by assumption. From Proposition 1 we derive

$$e_{4n+1}^{\rm cont}(I,F,G) \le 2C\, e_{n,C}^{\rm non}(I,F,G)\,.$$

Next we apply Lemma 3(ii) and obtain

$$e_n^{\rm cont}(S,F,H) \le \|S\|\, e_n^{\rm cont}(I,F,G)$$

as well as

$$e_{n,C}^{\rm non}(I,F,G) \le \|S^{-1}\|\, e_{n,C/(\|S^{-1}\|\,\|S\|)}^{\rm non}(S,F,H)\,.$$

Combining these inequalities we are done. □

4.2 Widths of Embeddings of Weighted Sequence Spaces

Having the wavelet characterization of Besov spaces in mind, cf. Subsections 5.3, 5.4, we introduce the following scale of sequence spaces.

Definition 3. *Let $0 < p, q \le \infty$ and let $s \in \mathbb{R}$. Let $\nabla := (\nabla_j)_j$ be a sequence of subsets of finite cardinality of the set $\{1, 2, \dots, 2^d - 1\} \times \mathbb{Z}^d$. We suppose that there exist $0 < C_1 \le C_2$ and $J \in \mathbb{N}$ such that the cardinality $|\nabla_j|$ of ∇_j satisfies*

$$C_1 \le 2^{-jd}\, |\nabla_j| \le C_2 \qquad \text{for all} \quad j \ge J\,. \tag{51}$$

Then $b^s_{p,q}(\nabla), 0 < q < \infty$, denotes the collection of all sequences $a = (a_{j,\lambda})_{j,\lambda}$ of complex numbers such that

$$\|\, a\,\|_{b^s_{p,q}} := \left(\sum_{j=0}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \Big(\sum_{\lambda \in \nabla_j} |\, a_{j,\lambda}|^p \Big)^{q/p} \right)^{1/q} < \infty\,. \tag{52}$$

For $q = \infty$, we use the usual modification

$$\|\, a\,\|_{b^s_{p,\infty}} := \sup_{j=1,2,\dots} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))} \left(\sum_{\lambda \in \nabla_j} |a_{j,\lambda}|^p \right)^{1/p} < \infty. \tag{53}$$

If there is no danger of confusion we shall write $b^s_{p,q}$ instead of $b^s_{p,q}(\nabla)$.

Remark 16. *In what follows we shall denote by $e_{j,\lambda}$ the elements of the canonical orthonormal basis of $b^0_{2,2}$. Let $\sigma \in \mathbb{R}$. It is obvious that the linear mapping L_σ defined by*

$$L_\sigma\, e_{j,\lambda} := 2^{-\sigma j}\, e_{j,\lambda} \qquad \text{for all} \quad j,\ \lambda\,,$$

extends to an isomorphism from $b^s_{p,q}$ onto $b^{s+\sigma}_{p,q}$ (simultaneously for all s, p, q) and $\|\, L_\sigma\,\| = 1$.

In the framework of these sequence spaces it is very easy to prove embedding theorems, cf. [54].

Lemma 4. *Let $0 < p_0, p_1, q_0, q_1 \le \infty$, $s \in \mathbb{R}$, and $t \ge 0$.*
(i) *The embedding*

$$b^{s+t}_{p_0,q_0}(\nabla) \hookrightarrow b^s_{p_1,q_1}(\nabla)$$

exists if and only if it is continuous if and only if either

$$t > d(\frac{1}{p_0} - \frac{1}{p_1})_+ \tag{54}$$

or

$$t = d(\frac{1}{p_0} - \frac{1}{p_1})_+ \qquad \text{and} \qquad q_0 \le q_1\,.$$

(ii) *The embedding*

$$b^{s+t}_{p_0,q_0}(\nabla) \hookrightarrow b^s_{p_1,q_1}(\nabla)$$

is compact if and only if (54) holds.

The main result of this subsection consists in the following:

Theorem 7. *Let* $0 < p, p_0, p_1 \le \infty$, $0 < q, q_0, q_1 \le \infty$, *and* $s \in \mathbb{R}$.
(i) *Suppose that*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+ \tag{55}$$

holds. Then, for any $C \ge 1$, *we have*

$$e^{\mathrm{non}}_{n,C}(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp n^{-\frac{t}{d}} .$$

(ii) *Suppose that* (55) *holds. Then we have*

$$e^{\mathrm{lin}}_n(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp \begin{cases} n^{-\frac{t}{d}} & \text{if} \quad 0 < p \le 2\,, \\ n^{-\frac{t}{d}+\frac{1}{p}-\frac{1}{2}} & \text{if} \quad 2 < p \le \infty\,. \end{cases}$$

(iii) *Suppose that* (54) *holds. Then we have*

$$e^{\mathrm{cont}}_n(I, b^{s+t}_{p_0,q_0}, b^s_{p_1,q_1}) \asymp n^{-\frac{t}{d}} .$$

Remark 17. *In part (i) there is an interesting limiting case. Suppose* $0 < p < 2$ *and* $t = d(\frac{1}{p} - \frac{1}{2})$. *Then the embedding* $b^{s+t}_{p,p} \hookrightarrow b^s_{2,2}$ *exists, cf. Lemma 4, and*

$$\left(\sum_{n=1}^{\infty} \left[n^{t/d}\, \sigma_n(a, \mathcal{B})_{b^s_{2,2}}\right]^p \frac{1}{n}\right)^{1/p} < \infty \quad \textit{if and only if} \quad a \in b^{s+t}_{p,p}\,.$$

In view of Lemma 4(ii) this shows that $\lim_{n\to\infty} e^{\mathrm{non}}_{n,C}(S, F, H) = 0$ *does not imply compactness of* S.

The proof of Theorem 7 requires some preparations. It will be given in Subsections 4.2.2-4.2.4.

4.2.1 The Bernstein Widths of the Identity Operator

We concentrate on the estimate from below. For later use we treat a more general situation.

Lemma 5. *Let* $0 < p_0, p_1, q_0, q_1 \le \infty$, $s \in \mathbb{R}$ *and* $t > 0$ *such that* (54) *holds. Then there exists a positive constant* c *such that*

$$b_n(I, b^{s+t}_{p_0,q_0}, b^s_{p_1,q_1}) \ge c \begin{cases} n^{-\frac{t}{d}} & \text{if} \quad 0 < p_0 \le p_1 \le \infty\,, \\ n^{-\frac{t}{d}+\frac{1}{p_0}-\frac{1}{p_1}} & \text{if} \quad 0 < p_1 < p_0 \le \infty\,. \end{cases} \tag{56}$$

holds for all n.

Proof. The Bernstein numbers are monotone in n. So it will be enough to prove the assertion for sufficiently large n. Consequently, we may assume that there is a natural number $N \geq J$ and positive constants c_1 and c_2 such that

$$c_1\, 2^{Nd} \leq n \leq c_2\, 2^{Nd}\, .$$

Step 1. Let $0 < p_0 \leq p_1$. Using Hölder's inequality we find

$$\begin{aligned}
\| \sum_{\lambda \in \nabla_N} b_\lambda\, e_{N,\lambda} | b^{s+t}_{p_0,q_0} \| &= 2^{N(s+t+\frac{d}{2}-\frac{d}{p_0})} \Big(\sum_{\lambda \in \nabla_N} |b_\lambda|^{p_0} \Big)^{1/p_0} \\
&\leq 2^{N(s+t+\frac{d}{2}-\frac{d}{p_0})}\, |\nabla_N|^{\frac{1}{p_0}-\frac{1}{p_1}} \Big(\sum_{\lambda \in \nabla_N} |b_\lambda|^{p_1} \Big)^{1/p_1} \\
&\leq C_2\, 2^{Nt}\, \| \sum_{\lambda \in \nabla_N} b_\lambda\, e_{N,\lambda}\, | b^{s}_{p_1,q_1} \| \\
&\leq c_3\, n^{t/d}\, \| \sum_{\lambda \in M_N} b_\lambda\, e_{N,\lambda}\, | b^{s}_{p_1,q_1} \|\, ,
\end{aligned}$$

where C_2 corresponds to (51). Consequently, the unit ball in $b^{s+t}_{p_0,q_0}$ contains the n-dimensional ball (spanned by the vectors $e_{N,\lambda}$, $\lambda \in \nabla_N$) with radius $c_3^{-1}\, n^{-t/d}$. This proves

$$b_n(I, b^{s+t}_{p_0,q_0}, b^{s}_{p_1,q_1}) \geq c\, n^{-t/d}$$

for some positive constant c independent of n.
Step 2. If $p_0 > p_1$, then Hölder's inequality (used in the second line of the estimate in Step 1) will be replaced by the monotonicity of the ℓ_r-norms and we obtain

$$\begin{aligned}
\| \sum_{\lambda \in \nabla_N} b_\lambda\, e_{N,\lambda}\, | b^{s+t}_{p_0,q_0} \| &= 2^{N(s+t+\frac{d}{2}-\frac{d}{p_0})} \Big(\sum_{\lambda \in \nabla_N} |b_\lambda|^{p_0} \Big)^{1/p_0} \\
&\leq 2^{N(s+t+\frac{d}{2}-\frac{d}{p_0})} \Big(\sum_{\lambda \in \nabla_N} |b_\lambda|^{p_1} \Big)^{1/p_1} \\
&\leq c_5\, 2^{N(t+\frac{d}{p_1}-\frac{d}{p_0})}\, \| \sum_{\lambda \in \nabla_N} b_\lambda\, e_{N,\lambda}\, | b^{s}_{p_1,q_1} \|\, .
\end{aligned}$$

This time the unit ball in $b^{s+t}_{p_0,q_0}$ contains the n-dimensional ball with radius

$$c_5^{-1}\, 2^{-N(t+\frac{d}{p_1}-\frac{d}{p_0})}.$$

This proves our claims. □

Remark 18. *In the one-dimensional periodic situation also estimates of the Bernstein numbers from above are known, due to Tsarkov and Maiorov, cf.* [78, Thm. 12,

p. 194]. *Let* $1 \le p \le \infty$ *and* $s > 0$. *By* $\mathring{W}^s_p$ *we denote the collection of all* 2π-*periodic functions* f *with Weyl derivative of order* s *belonging to* $L_p(\mathbb{T})$ *and satisfying* $\int_{-\pi}^{\pi} f(x)\, dx = 0$. *It holds*

$$b_n(I, \mathring{W}^t_{p_0}, L_{p_1}) \asymp \begin{cases} n^{-t} & \text{if} \quad 1 \le p_0 \le p_1 \le \infty \quad \text{or} \\ & \qquad 1 \le p_1 \le p_0 \le 2 \quad \text{and} \quad t > 0\,, \\ n^{-t+\frac{1}{p_0}-\frac{1}{p_1}} & \text{if} \quad 2 \le p_1 < p_0 \le \infty \quad \text{and} \quad t > \frac{1}{p_0}\,, \\ n^{-t+\frac{1}{p_0}-\frac{1}{2}} & \text{if} \quad 1 \le p_1 \le 2 \le p_0 \le \infty \quad \text{and} \quad t > \frac{1}{p_0}\,. \end{cases}$$

This should be compared with Lemma 5 for $s = 0$ *and* $d = 1$.

4.2.2 Best m-Term Approximation in the Framework of Sequence Spaces

We prepare the proof of part (i) of Theorem 7. Also here we treat a more general situation. Let $\mathcal{B}$ denote the canonical basis $(e_{j,\lambda})_{j,\lambda}$ in $b^0_{2,2}(\nabla)$. Then our aim in this subsection consists in a characterization of the behaviour of the best m-term approximation of a given element $a \in b^{s+t}_{p_0,q_0}$ with respect to $\mathcal{B}$.
The main result of this subsection reads as follows.

Theorem 8. *Let* $0 < p_0, p_1, q_0, q_1 \le \infty$, $s \in \mathbb{R}$ *and* $t > 0$ *such that* (54) *holds. Then we have*

$$\sup\Big\{\sigma_n(a, \mathcal{B})_{b^s_{p_1,q_1}} \; : \; \|\, a\|_{b^{s+t}_{p_0,q_0}} \le 1\Big\} \asymp n^{-\frac{t}{d}}\,. \tag{57}$$

We start with some preparations. Let U denote the unit ball in $b^{s+t}_{p_0,\infty}$. Then

$$a = \sum_{j=0}^{\infty} \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda} \qquad \text{and} \qquad \sup_{j=0,1,\ldots} \; 2^{j(s+t+d(\frac{1}{2}-\frac{1}{p_0}))} \Big(\sum_{\lambda \in \nabla_j} |a_{j,\lambda}|^{p_0}\Big)^{1/p_0} \le 1\,.$$

The following lemma will be of some use.

Lemma 6. *Let* $0 < p_0 \le p_1$ *and suppose*

$$t > d\Big(\frac{1}{p_0} - \frac{1}{p_1}\Big)\,. \tag{58}$$

For all $a \in U$ *and all* $n \ge 1$ *there exists a natural number* $K := K(a,n)$ *such that*

$$\Big\| a - \sum_{j=0}^{K} \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda} \Big| b^s_{p_1,q_1} \Big\| \; \le \; n^{-\frac{t}{d}}$$

holds.

Proof. We define

$$T_j \;:=\; \sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda}\,, \qquad j = 0,1\ldots\,.$$

Then one has

$$a - \sum_{j=0}^{K}\sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda} \;=\; \sum_{j>K} T_j.$$

Because of $0 < p_0 \le p_1 \le \infty$ the monotonicity of the ℓ_q-norms and $a \in U$ lead to

$$\begin{aligned}\|\, T_j\, |b^s_{p_1,q_1}\| \;&\le\; 2^{j(s+\frac{d}{2}-\frac{d}{p_1})}\Big(\sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^{p_0}\Big)^{1/p_0}\\ &\le\; 2^{-j(t+d(\frac{1}{p_0}-\frac{1}{p_1}))}.\end{aligned}$$

Let $u = \min(1,p_1,q_1)$. Consequently, using (58) and chosing K large enough, we find

$$\begin{aligned}\Big\|\sum_{j\ge K} T_j \,\Big| b^s_{p_1,q_1}\Big\|^u \;&\le\; \sum_{j\ge K} \|\, T_j\, |b^s_{p_1,q_1}\|^u \;\le\; \sum_{j\ge K} 2^{-ju\left[t+d\left(\frac{1}{p_0}-\frac{1}{p_1}\right)\right]}\\ &\le\; C_1\, 2^{-Ku(t+d(\frac{1}{p_0}-\frac{1}{p_1}))} \le n^{-\frac{tu}{d}}.\end{aligned}$$

This proves the claim. □

The basic step in deriving an upper estimate of $\sigma_n(a,\mathcal{B})$ is the following proposition. Again U denotes the unit ball in $b^{s+t}_{p_0,\infty}$.

Proposition 2. *Let $0 < p_0 \le p_1 \le \infty$. Let $a \in U$, $n \in \mathbb{N}$ and $K = K(a,n)$ as in Lemma 6. Then there exists an approximation*

$$S_n a \;:=\; \sum_{j=0}^{K}\sum_{\lambda\in\nabla_j} a^*_{j,\lambda}\, e_{j,\lambda} \tag{59}$$

of a, which satisfies

i) the coefficients $a^*_{j,\lambda}$ depend continuously on a;

ii) the number of nonvanishing entries is bounded by $c\cdot n$;

iii) $\|\, a - S_n a\, |b^s_{p_1,q_1}\| \;\le\; c\, n^{-t/d}\,, \quad n = 1,2,\ldots\,,$

and c can be chosen independent of a and n.

Proof. Observe that it will be enough to prove the claim for natural numbers $n = 2^{Nd}$, $N \in \mathbb{N}$. We define

$$\delta := \frac{t - d\left(\frac{1}{p_0} - \frac{1}{p_1}\right)}{2\left(\frac{1}{p_0} - \frac{1}{p_1}\right)},$$

$$\varepsilon_j := \begin{cases} 0 & \text{if} \quad 1 \le j \le N \\ n^{-\frac{1}{p_0}} 2^{-jd(\frac{1}{2} - \frac{1}{p_0})} 2^{-jt} 2^{(j-N)\frac{\delta}{p_0}} & \text{if} \quad j > N, \end{cases} \tag{60}$$

$$\Lambda_j^* := \left\{\lambda \in \nabla_j \ : \ |a_{j,\lambda}|\, 2^{sj} \ge \varepsilon_j\right\}, \qquad j = 0, 1, \ldots\,. \tag{61}$$

Then, if $j > N$,

$$\begin{aligned} |\Lambda_j^*| &= \sum_{\lambda \in \Lambda_j^*} 1 \le \sum_{\lambda \in \Lambda_j^*} 2^{jsp_0} \frac{|a_{j,\lambda}|^{p_0}}{\varepsilon_j^{p_0}} \\ &\le \sum_{\lambda \in \nabla_j} n\, 2^{jd(\frac{1}{2} - \frac{1}{p_0})p_0} 2^{jtp_0} 2^{-(j-N)\delta} 2^{jsp_0} |a_{j,\lambda}|^{p_0} \\ &= n\, 2^{-(j-N)\delta} \sum_{\lambda \in \nabla_j} 2^{j(s+t+d(\frac{1}{2} - \frac{1}{p_0}))p_0} \, |a_{j,\lambda}|^{p_0} \\ &\le n\, 2^{-(j-N)\delta} \, \| \, a \, | b^{s+t}_{p_0,\infty} \|^{p_0} \\ &\le n\, 2^{-(j-N)\delta}\,. \end{aligned} \tag{62}$$

Now a typical method to approximate a would be to choose $a^*_{j,\lambda} = a_{j,\lambda}\,,\ j \in \Lambda_j^*$ and zero otherwise. However, this selection does not depend continuously on a. Therefore we use the following variant. Let g_j denote the following piecewise linear and odd function,

$$g_j(x) := \begin{cases} 0 & \text{if} \quad 0 \le x \le 2^{-js}\varepsilon_j\,, \\ x & \text{if} \quad x \ge 2 \cdot 2^{-js}\varepsilon_j\,, \\ \text{linear} & \text{if} \quad x \in (2^{-js}\varepsilon_j, 2 \cdot 2^{-js}\varepsilon_j)\,. \end{cases} \tag{63}$$

Then we set

$$a^*_{j,\lambda} := g_j(a_{j,\lambda}) \tag{64}$$

and consider the associated approximation (59). Let us prove that S_n will do the job.

Step 1. We shall prove (i). Observe

$$\Big| \bigcup_{j=0}^{K} \Lambda_j^* \Big| \le c_1 \sum_{j=0}^{N} 2^{jd} + \sum_{j=N+1}^{K} n\, 2^{-(j-N)\delta} \le c_2\, n\,,$$

cf. (62). The constant c_2 is independent of a, K, and n. This proves (i) and (ii).
Step 2. Proof of (iii). We have

$$a - S_n a \; = \; a - \sum_{j=0}^{K} \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda} \; + \; \sum_{j=0}^{K} T_j^* \; =: \; \Sigma_1 + \Sigma_2,$$

where

$$T_j^* \; = \; \sum_{\lambda \in \nabla_j} \big(a_{j,\lambda} - a^*_{j,\lambda}\big)\, e_{j,\lambda}.$$

From Lemma 6, we can conclude that $\| \Sigma_1 \, | b^s_{p_1,q_1} \| \; \le \; n^{-t/d}$ for K large enough. Therefore it remains to estimate $\| T_j^* \, | b^s_{p_1,q_1} \|$. Since $|g_j(x) - x| \; \le \; |x|$ and $a^*_{j,\lambda} \; = \; a_{j,\lambda}$ for $|a_{j,\lambda}| \; \ge \; 2\varepsilon_j 2^{-js}$, we obtain

$$\begin{aligned} |\, a_{j,\lambda} - a^*_{j,\lambda}\, |^{p_1} & \le \; |a_{j,\lambda}|^{p_1} \\ & \le \; |a_{j,\lambda}|^{p_0} |a_{j,\lambda}|^{p_1 - p_0} \\ & \le \; |a_{j,\lambda}|^{p_0} (2\varepsilon_j)^{p_1 - p_0} 2^{-js(p_1 - p_0)} \, . \end{aligned}$$

This will be used to estimate the norm of T_j^* as follows:

$$\begin{aligned} \| \, T_j^* \, | b^s_{p_1,q_1} \| \; & = \; 2^{j(s + d(\frac{1}{2} - \frac{1}{p_1}))} \Big(\sum_{k \in \nabla_j} |a_{j,\lambda} - a^*_{j,\lambda}|^{p_1} \Big)^{1/p_1} \\ & \le \; c_1 \, 2^{jd(\frac{1}{2} - \frac{1}{p_1})} \, 2^{js\frac{p_0}{p_1}} \, \varepsilon_j^{1 - \frac{p_0}{p_1}} \Big(\sum_{k \in \nabla_j} |a_{j,\lambda}|^{p_0} \Big)^{1/p_1} \\ & \le \; c_1 \, \varepsilon_j^{1 - \frac{p_0}{p_1}} 2^{jd/2} 2^{-jt\frac{p_0}{p_1}} 2^{-jd\frac{p_0}{2p_1}} \Big(\sum_{\lambda \in \nabla_j} 2^{j(s+t+d(\frac{1}{2} - \frac{1}{p_0}))p_0} |a_{j,\lambda}|^{p_0} \Big)^{1/p_1} \\ & \le \; c_2 \, \varepsilon_j^{1 - \frac{p_0}{p_1}} 2^{-j(t + \frac{d}{2} - \frac{dp_1}{2p_0})\frac{p_0}{p_1}} \; \| \, a \, | b^{s+t}_{p_0,\infty} \|^{p_0/p_1} \\ & \le \; c_2 \, \varepsilon_j^{1 - \frac{p_0}{p_1}} 2^{-j(t + \frac{d}{2} - \frac{dp_1}{2p_0})\frac{p_0}{p_1}} \, , \end{aligned}$$

where again c_2 does not depend on a and n. For $j > N$ we continue by employing the concrete value of ε_j and obtain

$$\begin{aligned} \| \, T_j^* \, | b^s_{p_1,q_1} \| \; & \le \; c_2 \Big(n^{-\frac{1}{p_0}} 2^{-jd(\frac{1}{2} - \frac{1}{p_0})} 2^{-jt} 2^{(j-N)\frac{\delta}{p_0}} \Big)^{1 - \frac{p_0}{p_1}} 2^{-j(t + \frac{d}{2} - \frac{dp_1}{2p_0})\frac{p_0}{p_1}} \\ & = \; c_2 \, n^{\frac{1}{p_1} - \frac{1}{p_0}} \, 2^{-N\delta(\frac{1}{p_0} - \frac{1}{p_1})} 2^{-j(t - d(\frac{1}{p_0} - \frac{1}{p_1}) - \frac{\delta}{p_0} + \frac{\delta}{p_1})} \, . \end{aligned}$$

By construction $T_j^* = 0$ if $j \le N$ and by definition

$$t - d\Big(\frac{1}{p_0} - \frac{1}{p_1}\Big) > \delta \Big(\frac{1}{p_0} - \frac{1}{p_1}\Big) \, .$$

Hence, with $u = \min(1, p_1, q_1)$

$$\begin{aligned} \| \Sigma_2 \,|b^s_{p_1,q_1}\|^u &\le c_2^u \Big(n^{\frac{1}{p_1}-\frac{1}{p_0}}\, 2^{-N\delta(\frac{1}{p_0}-\frac{1}{p_1})}\Big)^u \sum_{j=N+1}^{K} 2^{-ju(t-d(\frac{1}{p_0}-\frac{1}{p_1})-\frac{\delta}{p_0}+\frac{\delta}{p_1})} \\ &\le c_3 \Big(n^{\frac{1}{p_1}-\frac{1}{p_0}}\, 2^{-N\delta(\frac{1}{p_0}-\frac{1}{p_1})}\Big)^u\, 2^{-Nu(t-d(\frac{1}{p_0}-\frac{1}{p_1})-\frac{\delta}{p_0}+\frac{\delta}{p_1})} \\ &= c_3 \Big(n^{\frac{1}{p_1}-\frac{1}{p_0}}\Big)^u\, 2^{-Nu(t-d(\frac{1}{p_0}-\frac{1}{p_1}))}\,, \end{aligned}$$

with c_3 independent of K, n and a. Recalling that $2^{Nd} = n$, we end up with

$$\| \Sigma_2 \,|b^s_{p_1,q_1}\| \le c_3\, n^{-t/d}\,.$$

This finishes the proof of Proposition 2. □

For completeness and better reference we formulate the counterpart of Proposition 2 in case $p_0 \ge p_1$.

Proposition 3. *Let $0 < p_1 \le p_0 \le \infty$. Let $a \in U$ (the unit ball in $b^{s+t}_{p_0,\infty}$) and $2^{Nd} \le n \le 2^{(N+1)}d$. Then the approximation*

$$S_n a := \sum_{j=0}^{N} \sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda} \tag{65}$$

of a satisfies

i) the coefficients $a_{j,\lambda}$ depend continuously on a;

ii) the number of nonvanishing entries is bounded by $c \cdot n$;

iii) $\| a - S_n a \,|b^s_{p_1,q_1}\| \le c\, n^{-t/d}, \quad n = 1, 2, \ldots,$

and c can be chosen independent of a and n.

Proof. The proof is elementary. □

Proof of Theorem 8. The estimate from above follows from Propositions 2, 3, and the continuous embedding $b^{s+t}_{p_0,q_0} \hookrightarrow b^{s+t}_{p_0,\infty}$. For the estimate from below it will be enough to consider $n = 2^{Nd}$, $N \ge J$ and $N \in \mathbb{N}$. Let K be the smallest natural number such that $C_1\, 2^{Kd} \ge 2$ (here C_1 is the same constant as in (51)). Then

$$n \le \frac{C_1\, 2^{(N+K)d}}{2} \le \frac{1}{2}\, |\nabla_{N+K}|\,.$$

Let $\Gamma \subset \nabla_{N+K}$ with $|\Gamma| = n$. We define

$$a = |\nabla_{N+K}|^{-1/p_0}\, 2^{-(N+K)(s+t+d(\frac{1}{2}-\frac{1}{p_0}))} \sum_{\lambda\in\nabla_{N+K}} e_{N+K,\lambda}\,.$$

Consequently $\| \, a \, \|_{b^{s+t}_{p_0,q_0}} = 1$ for any q_0. Furthermore, we find

$$\begin{aligned}
\|a - S_n a\|_{b^s_{p_1,q_1}} &\geq \Big\| \sum_{\lambda \in \nabla_{N+K} \setminus \Gamma} |\nabla_{N+K}|^{-1/p_0} \, 2^{-(N+K)(s+t+d(\frac{1}{2}-\frac{1}{p_0}))} \, e_{N+K,\lambda} \Big\|_{b^s_{p_1,q_1}} \\
&= |\nabla_{N+K}|^{-1/p_0} \, 2^{-(N+K)(t+d(\frac{1}{p_1}-\frac{1}{p_0}))} |\nabla_{N+K} \setminus \Gamma|^{1/p_1} \\
&\geq \frac{C_1^{1/p_1}}{2^{1/p_1} \, C_2^{1/p_0}} \, 2^{-(N+K)t} \\
&= \frac{C_1^{1/p_1}}{2^{1/p_1} \, C_2^{1/p_0}} \, 2^{-Kt} \, n^{-t/d} \, ,
\end{aligned}$$

(also C_2 has the same meaning as in (51)). It is clear that an optimal Γ with $|\Gamma| = n$ has to be a subset of ∇_{N+K}. This completes the proof of the estimate from below. □

Proof of Theorem 7(i). The estimate from above is covered by Theorem 8, the estimate from below follows from Theorem 1 and Theorem 7(iii). □

4.2.3 The Manifold Widths of the Identity

Proof of Theorem 7(iii). Without loss of generality we may choose $s = 0$, cf. Lemma 3(ii) and Remark 16.
Step 1. The estimate from above. In case $p_1 = q_1 = 2$ we may use Propositions 1, 2 and 3 to get the desired inequality. However, for the general case we have to modify the argument. We follow the arguments used in [30]. Let U denote the unit ball in $b^t_{p_0,q_0}$. As explained there Propositions 2, 3 guarantee

$$a^n(U, b^0_{p_1,q_1}) \leq c n^{-t/d} \, ,$$

where a^n denotes the Alexandroff-co-width, cf. [30] for details. But

$$e^{\text{cont}}_{2n+1}(U, b^0_{p_1,q_1}) \leq a^n(U, b^0_{p_1,q_1}) \, ,$$

cf. [30] and [39]. Let us mention that in the quoted literature the target space was always a normed linear space. But the arguments carry over to quasi-normed linear spaces.
Step 2. The estimate from below. Lemmata 5 and 1 yield the lower estimate in case $0 < p_0 \leq p_1 \leq \infty$.

Now, let $p_1 < p_0 \le \infty$. Let $\varepsilon > 0$. We consider the diagram

$$\begin{array}{ccc} b^0_{p_1,q_1} & \stackrel{I_3}{\longrightarrow} & b^{-d(\frac{1}{p_1}-\frac{1}{p_0})-\varepsilon}_{p_0,\infty} \\ I_2 \nwarrow & & \nearrow I_1 \\ & b^t_{p_0,q_0}, & \end{array}$$

where I_1, I_2 and I_3 are identity operators. Then (23) yields

$$e^{\mathrm{cont}}_{2n}(I_1, b^t_{p_0,q_0}, b^{-d(\frac{1}{p_1}-\frac{1}{p_0})-\varepsilon}_{p_0,\infty}) \le e^{\mathrm{cont}}_n(I_2, b^t_{p_0,q_0}, b^0_{p_1,q_1})\; e^{\mathrm{cont}}_n(I_3, b^0_{p_1,q_1}, b^{-d(\frac{1}{p_1}-\frac{1}{p_0})-\varepsilon}_{p_0,\infty})$$

which implies

$$c_1\, n^{-\frac{t}{d}-\frac{1}{p_1}+\frac{1}{p_0}-\frac{\varepsilon}{d}} \le c_2\, e^{\mathrm{cont}}_n(I_2, b^t_{p_0,q_0}, b^0_{p_1,q_1})\; n^{-\frac{1}{p_1}+\frac{1}{p_0}-\frac{\varepsilon}{d}}$$

for some positive c_1 and c_2 (independent of n), see Lemmata 5, 1, and Step 1. □

Remark 19. *As it is clear from the proof given above that to establish the estimate from below of* e^{cont}_n *the knowledge of the behavior of the Bernstein widths is not enough. Here the multiplicativity of the numbers* e^{cont}_n*, cf.* (23)*, is crucial. This seems to be overlooked in* [30].

4.2.4 The Approximation Numbers of the Identity

Proof of Theorem 7(ii). *Step 1.* Let $2 \le p \le \infty$. From Proposition 3 the estimate from above with S_n as in (65) follows. The estimate from belows is covered by (57).
Step 2. Let $0 < p < 2$. Without loss of generality we assume $s = 0$. Let S_n be defined as in (65). The estimate from above is easily derived by using the monotonicity of the ℓ_r-norms and $t + d(\frac{1}{2} - \frac{1}{p}) > 0$:

$$\begin{aligned} \|\, a - S_n a\, |b^0_{2,2}\|^2 &\le \sum_{j=N+1}^{\infty} \Big(\sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \Big)^{2/p} \\ &\le \Big(\sum_{j=N+1}^{\infty} 2^{-2j(t+d(\frac{1}{2}-\frac{1}{p}))} \Big) \Big(\sup_{j\ge N+1} 2^{j(t+d(\frac{1}{2}-\frac{1}{p}))} \Big(\sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \Big)^{1/p} \Big)^2 \\ &\le c\, 2^{-2N(t+d(\frac{1}{2}-\frac{1}{p}))} \|\, a\, |b^t_{p,\infty}\|^2 \\ &\le c \Big(n^{-\frac{t}{d}-\frac{1}{2}+\frac{1}{p}} \|\, a\, |b^t_{p,q}\| \Big)^2 , \end{aligned}$$

where c does not depend on n and a. For the estimate from below we use the obvious fact that the optimal approximation of an element in a Hilbert space is given by the partial sum with respect to an orthonormal basis. Hence, if $\widetilde{S}_n$ is a linear operator of rank $\le n$ then

$$\| \, a - \widetilde{S}_n a | b_{0,0} \| \ge \| \, a - S_n a | b_{0,0} \| \, ,$$

where S_n is defined as in (65). We put

$$a := \sum_{j=0}^{N+1} e_{j,\lambda_j} \, ,$$

where $\lambda_j \in \nabla_j$ can be chosen arbitrarily. Then

$$\| \, a \, | b^t_{p,q} \| = \Big(\sum_{j=0}^{N+1} 2^{j(t+d(\frac{1}{2}-\frac{1}{p}))q} \Big)^{1/q} \ge 2^{N(t+d(\frac{1}{2}-\frac{1}{p}))}$$

for some positive c independent of n and

$$\| \, a - S_n a | b^0_{2,2} \| = 1 \, .$$

This implies

$$\| \, I - S_n \, | b^t_{p,q} \| \ge \frac{1}{2^{N(t+d(\frac{1}{2}-\frac{1}{p}))}} \, ,$$

which finishes the proof of the lower bound. □

Remark 20. *Notice that in any case an in order optimal approximation is given by an appropriate partial sum, see* (65)*.*

4.2.5 The Gelfand Widths of the Identity

All what we will do here relies on a result of Gluskin [44, 45] about the Gelfand widths of the embedding $\ell_p^m \to \ell_2^m$ which we now recall: let $1/p + 1/p' = 1$. For all natural numbers m and n, $n \le m$, it holds

(66)

$$d^n(I, \ell_p^m, \ell_2^m) \asymp \begin{cases} (m-n+1)^{\frac{1}{2}-\frac{1}{p}} & \text{if} \quad 2 \le p \le \infty \, , \\ 1 & \text{if} \quad 1 \le p < 2 \quad \text{and} \quad 1 \le n \le m^{2/p'} \, , \\ m^{1/p'} \, n^{-1/2} & \text{if} \quad 1 \le p < 2 \quad \text{and} \quad m^{2/p'} \le n \le m \, . \end{cases}$$

The simple monotonicity argument leads to the following supplement to $p = 1$. There exists a constant c, independent of m and n, such that

$$d^n(I, \ell_p^m, \ell_2^m) \le c \, n^{-1/2} \qquad (67)$$

if $0 < p < 1$ and $1 \le n \le m$.

The Gelfand widths are examples of so-called s-numbers, cf. [70, 69] and [10]. Following Pietsch [69, 2.2.4, p. 80] we associate to the sequence of Gelfand widths the following operator ideals. Let F and E be quasi-Banach spaces and denote by $\mathcal{L}(F,E)$ the class of all linear continuous operators $T : F \to E$. Then we put for $0 < r < \infty$

$$\mathcal{L}^{(c)}_{r,\infty} := \Big\{ T \in \mathcal{L}(F,E) : \quad \sup_{n\in\mathbb{N}} n^{1/r}\, d^n(T) < \infty \Big\} .$$

Equipped with the quasi-norm

$$\lambda_r(T) := \sup_{n\in\mathbb{N}} n^{1/r}\, d^n(T)$$

the set $\mathcal{L}^{(c)}_{r,\infty}$ becomes a quasi-Banach space. For such quasi-Banach spaces there always exist a real number $\varrho \in (0,1]$ and an equivalent quasi-norm, here denoted by $\| \cdot \,|\mathcal{L}^{(c)}_{r,\infty}\|$, such that

$$\| T_1 + T_2 \,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \le \| T_1 \,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho + \| T_2 \,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \tag{68}$$

holds for all $T_1, T_2 \in \mathcal{L}^{(c)}_{r,\infty}$.
To shorten notation we shall use the abbreviation $I^m_{p,q}$ for the identity $I : \ell^m_p \to \ell^m_q$. It is not complicated to check that (66), (67) imply the following estimates for $\| I^m_{p,2} \,|\mathcal{L}^{(c)}_{r,\infty}\|$, cf. [55].

Lemma 7. *Let* $0 < r < \infty$.
(i) *Let* $2 \le p \le \infty$. *Then*

$$\| I^m_{p,2} \,|\mathcal{L}^{(c)}_{r,\infty}\| \asymp m^{\frac{1}{r}-\frac{1}{p}+\frac{1}{2}} \tag{69}$$

holds.
(ii) *Let* $1 < p < 2$. *Then*

$$\| I^m_{p,2} \,|\mathcal{L}^{(c)}_{r,\infty}\| \asymp \begin{cases} m^{\frac{1}{r}-\frac{1}{p}+\frac{1}{2}} & \text{if} \quad 0 < r \le 2 , \\ m^{\frac{2}{rp'}} & \text{if} \quad 2 < r < \infty , \end{cases} \tag{70}$$

holds.
(iii) *Let* $0 < p \le 1$. *Then there exists a constant* c *such that*

$$\| I^m_{p,2} \,|\mathcal{L}^{(c)}_{r,\infty}\| \le c \begin{cases} m^{\frac{1}{r}-\frac{1}{2}} & \text{if} \quad 0 < r \le 2 , \\ 1 & \text{if} \quad 2 < r < \infty , \end{cases} \tag{71}$$

holds for all $m \in \mathbb{N}$.

In order to prove the estimates of the Gelfand numbers from above it turns out to be useful to split the identity I into two parts $\mathrm{id}^1, \mathrm{id}^2$ and to treat them independently. In fact, we shall investigate $\| \mathrm{id}^i \,|\mathcal{L}^{(c)}_{r_i,\infty}\|$, $i = 1,2$, and r_1 and r_2 have to be chosen in different ways. For basic properties of the Gelfand numbers we refer to Remark 7 and [10, 2.3].

Theorem 9. *Let* $0 < q \le \infty$.
(i) *Let* $1 \le p < 2$ *and suppose* $t > d/2$. *Then*

$$d^n(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp n^{-\frac{t}{d}} .$$

(ii) *Let* $2 < p \le \infty$ *and suppose* $t > 0$. *Then*

$$d^n(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp n^{-\frac{t}{d}} .$$

(iii) *Let* $0 < p < 1$ *and suppose*

$$t > d\Big(\frac{1}{p} - \frac{1}{2}\Big) . \tag{72}$$

Then there exist two constants c_1 *and* c_2 *such that*

$$c_1\, n^{-\frac{t}{d}} \le d^n(I, b^{s+t}_{p,q}, b^s_{2,2}) \le c_2 n^{-\frac{t}{d}-1+\frac{1}{p}} .$$

Proof. Without loss of generality we may assume $s = 0$. To see this consider the diagram

$$\begin{array}{ccc} b^{s+t}_{p,q} & \xrightarrow{\ I_1\ } & b^s_{2,2} \\ {\scriptstyle L_{-s}}\big\downarrow & & \big\uparrow{\scriptstyle L_s} \\ b^t_{p,q} & \xrightarrow{\ I_2\ } & b^0_{2,2}\,, \end{array}$$

where L_s denotes the isomorphism introduced in Remark 16. The multiplicativity of the Gelfand numbers implies

$$d^n(I_1, b^{s+t}_{p,q}, b^s_{2,2}) \le \| L_{-s} \| \, \| L_s \| \; d^n(I_2, b^t_{p,q}, b^0_{2,2}) ,$$

compare with Remark 7. Changing L_{-s} into L_s and vice versa in the above diagram we end up with

$$d^n(I_1, b^{s+t}_{p,q}, b^s_{2,2}) = d^n(I_2, b^t_{p,q}, b^0_{2,2}).$$

Step 1. Estimate from above. We concentrate on natural numbers $n = 2^{Nd}$, $N \in \mathbb{N}$ (the remaining can be treated by the monotonicity of the d^n). Let id_j denote the projection given by

$$\Big(\mathrm{id}_j\, a\Big)_{m,\lambda} := \begin{cases} a_{j,\lambda} & \text{if} \quad m = j\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Further, depending on N we split the identity I into a sum $I = \mathrm{id}^1 + \mathrm{id}^2$, where

$$\mathrm{id}^1 := \sum_{j=0}^{N} \mathrm{id}_j \qquad \text{and} \qquad \mathrm{id}^2 := \sum_{j=N+1}^{\infty} \mathrm{id}_j\,.$$

Later on we shall apply the following observation. Consider the diagram

$$\begin{array}{ccc} b^t_{p,q}(\nabla) & \xrightarrow{\ \mathrm{id}_j\ } & b^0_{2,2}(\nabla) \\ P \big\downarrow & & \big\uparrow Q \\ \ell_p^{|\nabla_j|} & \xrightarrow{\ I^{|\nabla_j|}_{p,2}\ } & \ell_2^{|\nabla_j|} \, . \end{array}$$

where P and Q are defined as follows. Let $a = (a_{\ell,\lambda})_{\ell,\lambda}$. Then

$$(P(a))_\lambda := a_{j,\lambda} \, .$$

For $b = (b_\lambda)_\lambda$ we define

$$(Q(b))_{\ell,\lambda} := \begin{cases} a_{j,\lambda} & \text{if} \quad j = \ell \, , \\ 0 & \text{otherwise} \, . \end{cases}$$

Obviously,

$$\| \, P \, \| = 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))} \qquad \text{and} \qquad \| \, Q \, \| = 1 \, .$$

Then the multiplicativity of the Gelfand numbers yields

$$\begin{aligned} d^n(\mathrm{id}_j, b^{s+t}_{p,q}, b^s_{2,2}) & \leq \| \, P \, \| \, \| \, Q \, \| \, d^n(I^{|\nabla_j|}_{p,2}) \\ & \leq 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))} \, d^n(I^{|\nabla_j|}_{p,2}) \, . \end{aligned} \tag{73}$$

Substep 1.1. The estimate of $d^n(\mathrm{id}^1, b^t_{p,q}, b^0_{2,2})$, $n = 2^{Nd}$. First we suppose $2 \leq p \leq \infty$. Thanks to (68), (69), and (73) we find

$$\begin{aligned} \| \, \mathrm{id}^1 \, |\mathcal{L}^{(c)}_{r,\infty}\|^\varrho & \leq \sum_{j=0}^{N} \| \, \mathrm{id}_j \, |\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\ & \leq \sum_{j=0}^{N} 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))\varrho} \, \| \, I^{|\nabla_j|}_{p,2} \, |\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\ & \leq c_1 \sum_{j=0}^{N} 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))\varrho} \, 2^{jd(\frac{1}{r}-\frac{1}{p}+\frac{1}{2})\varrho} \\ & \leq c_2 \, 2^{N(\frac{d}{r}-t)\varrho} \end{aligned} \tag{74}$$

if $d > t\,r$. Choosing r small enough we derive from the definition of $\mathcal{L}^{(c)}_{r,\infty}$

$$d^n(\mathrm{id}^1) = d^{2^{Nd}}(\mathrm{id}^1) \leq c_3 \, 2^{-Nt} = c_3 \, n^{-t/d} \, . \tag{75}$$

Now we consider the case $1 \leq p < 2$. Similar as above, but using (70) instead of (69), we find

$$\| \, \mathrm{id}^1 \, |\mathcal{L}^{(c)}_{r,\infty}\| \leq c_2 \, 2^{N(\frac{d}{r}-t)}$$

if $\frac{1}{r} > t/d$ and $1/r \ge 2$. Choosing r small enough we obtain

$$d^{2^{Nd}}(\mathrm{id}^1) \le c_4\, 2^{-Nt}\,. \tag{76}$$

Finally, we investigate the case $0 < p < 1$. As above we obtain

$$d^{2^{Nd}}(\mathrm{id}^1) \le c_5\, 2^{-N(t+d-\frac{d}{p})} = c_5\, n^{-\frac{t}{d}-1+\frac{1}{p}}\,. \tag{77}$$

Substep 1.2. The estimate of $d^n(\mathrm{id}^2, b^t_{p,q}, b^0_{2,2})$, $n = 2^{Nd}$.
Again we split our considerations into the three cases $p \ge 2$ and $1 \le p < 2$ and $0 < p < 1$. First, let $2 \le p \le \infty$. Using (68), (69), and (73) we find

$$\begin{aligned} \|\,\mathrm{id}^2\,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho &\le \sum_{j=N+1}^{\infty} \|\,\mathrm{id}_j\,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\ &\le \sum_{j=N+1}^{\infty} 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))\varrho}\, \|\, I^{|\nabla_j|}_{p,2}\,|\mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\ &\le c_1 \sum_{j=N+1}^{\infty} 2^{-j(t+d(\frac{1}{2}-\frac{1}{p}))\varrho}\; 2^{jd(\frac{1}{r}-\frac{1}{p}+\frac{1}{2})\varrho} \\ &\le c_2\, 2^{N(\frac{d}{r}-t)\varrho} \end{aligned} \tag{78}$$

if $t\,r > d$. Choosing r large enough ($t > 0$ by assumption) we derive

$$d^{2^{Nd}}(\mathrm{id}^2) \le c_3\, 2^{-Nt}\,. \tag{79}$$

Now we consider $1 \le p < 2$. Similarly

$$\|\,\mathrm{id}^2\,|\mathcal{L}^{(c)}_{r,\infty}\| \le c_3\, 2^{N(\frac{d}{r}-t)} \qquad \text{if} \quad \frac{1}{2} \le \frac{1}{r} < \frac{t}{d}\,.$$

Because of $t > d/2$ such a choice is always possible. Consequently

$$d^{2^{Nd}}(\mathrm{id}^2) \le c_4\, 2^{-Nt}\,. \tag{80}$$

Finally, let $0 < p < 1$. Then

$$d^{2^{Nd}}(\mathrm{id}^1) \le c_5\, 2^{-N(t+d-\frac{d}{p})} \qquad \text{if} \qquad \frac{t}{d} + 1 - \frac{1}{p} > \frac{1}{r} \ge \frac{1}{2}\,. \tag{81}$$

Such a choice is always possible if (72) holds.
Substep 1.3. The additivity of the Gelfand widths yields

$$d^{2n}(\mathrm{id}) \le d^n(\mathrm{id}^1) + d^n(\mathrm{id}^2)\,.$$

In view of this inequality the estimate from above of the Gelfand widths follows from (75)-(81).

Step 2. Estimate from below. Since $b_n \le c\, d^n$, cf. Lemma 1(i), we may use Lemma 5 here to derive the lower bound in case $0 < p \le 2$. For $p > 2$ we shall use a different argument. Again we restrict ourselves to a subsequence of the natural numbers n,

$$\frac{|\nabla_N|}{2} \le n < \frac{|\nabla_N|}{2} + 1\,, \qquad N \in \mathbb{N}\,.$$

Consider the diagram

$$\begin{array}{ccc} \ell_p^{|\nabla_N|} & \xrightarrow{\ I_1\ } & \ell_2^{|\nabla_N|} \\ {\scriptstyle P}\big\downarrow & & \big\uparrow{\scriptstyle Q} \\ b^t_{p,q}(\nabla) & \xrightarrow{\ I_2\ } & b^0_{2,2}(\nabla)\,, \end{array}$$

where I_1 and I_2 denote identities and this time P and Q are defined as follows. Let $b = (b_\lambda)_{\lambda \in \nabla_N}$. Then

$$(P(b))_{j,\lambda} := \begin{cases} b_\lambda & \text{if} \quad j = N\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

For $a = (a_{j,\lambda})_{j,\lambda}$ we define

$$(Q(a))_\lambda := a_{N,\lambda}\,, \qquad \lambda \in \nabla_N\,.$$

Obviously,

$$\|\, P \,\| = 2^{N(t+d(\frac{1}{2}-\frac{1}{p}))} \qquad \text{and} \qquad \|\, Q \,\| = 1\,.$$

Then the multiplicativity of the Gelfand numbers yields

$$d^n(I_1, \ell_p^{|\nabla_N|}, \ell_2^{|\nabla_N|}) \le \|\, P \,\| \; \|\, Q \,\| \; d^n(I_2, b^t_{p,q}(\nabla), b^0_{2,2}(\nabla))$$

which, in view of Gluskin's estimates (66), implies

$$c\, 2^{Nd(\frac{1}{2}-\frac{1}{p})} \le \; 2^{N(t+d(\frac{1}{2}-\frac{1}{p}))}\, d^n(I_2, b^t_{p,q}, b^0_{2,2})$$

for some positive c (independent of N). This completes the estimate from below. □

Remark 21. *The use of operator ideals in such a connection and the associated splitting technique applied in Step 1 has some history, cf.* [9, 55, 53]. *Closest to us is* [53]*, where these methods have been used in connection with entropy numbers.*

4.3 Widths of Embeddings of Besov Spaces

Here we do not formulate a general result since the restrictions on the domains are different for different widths.

4.3.1 The Manifold Widths of the Identity

The main result of this subsection consists in the following non-discrete counterpart of Theorem 7.

Theorem 10. *Let Ω be a bounded Lipschitz domain. Let $0 < p_0, p_1 \le \infty$, $0 < q_0, q_1 \le \infty$, and $s \in \mathbb{R}$. Suppose that* (54) *holds. Then we have*

$$e_n^{\rm cont}(I, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega))) \asymp n^{-\frac{t}{d}} \,. \tag{82}$$

Remark 22. *Theorem* 10 *has several forerunners. We would like to mention DeVore, Howard, and Micchelli* [29], *DeVore, Kyriazis, Leviatan, and Tikhomirov* [30], *and Dung and Thanh* [39]. *In these quoted papers the authors consider the quantities $e_n^{\rm cont}(I, B^t_{q_0}(L_{p_0}(\Omega)), L_{p_1}(\Omega))$. Note, that from the continuous embeddings*

$$B^0_1(L_p(\Omega)) \hookrightarrow L_p(\Omega) \hookrightarrow B^0_\infty(L_p(\Omega)) \,, \qquad 1 \le p \le \infty \,,$$

we obtain as a direct consequence of Theorem 10

$$e_n^{\rm cont}(I, B^t_{q_0}(L_{p_0}(\Omega)), L_{p_1}(\Omega)) \asymp n^{-\frac{t}{d}} \,, \tag{83}$$

as long as $1 \le p_1 \le \infty$ and $t > (\frac{1}{p_0} - \frac{1}{p_1})_+$. So, Theorem 10 covers the results obtained before. However, let us mention that we used the ideas from [30] *for our estimate from above and the ideas from* [39] *to derive the estimate from below (here on the level of sequence spaces).*

Proof of Theorem 10. Let $\mathcal{E}$ denote a universal bounded linear extension operator corresponding to Ω, see Proposition 6 in Subsection 5.5. Let $\operatorname{diam}\Omega$ be the diameter of Ω and let x^0 be a point in $\mathbb{R}^d$ such that

$$\Omega \subset \{y : \, |x^0 - y| \le \operatorname{diam}\Omega\} \,.$$

Without loss of generality we assume

$$\operatorname{supp}\mathcal{E}f \subset \{y : \, |x^0 - y| \le 2\operatorname{diam}\Omega\} \,.$$

Let ∇ be defined as in (98) and (99) (with Ω replaced by the ball with radius $2\operatorname{diam}\Omega$ and center x^0). By R we denote the restriction operator with respect to Ω. By T we denote the linear and continuous operator which associates to f its wavelet series and by T^{-1} the inverse operator. Here we assume that we can characterize the Besov spaces $B^{s+t}_{p_0,q_0}(\mathbb{R}^d)$ as well as $B^s_{p_1,q_1}(\mathbb{R}^d)$ in the sense of Proposition 5 in Subsection 5.3. Then we consider the following diagram

$$
\begin{array}{ccc}
B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{\mathcal{E}} B^{s+t}_{q_0}(L_{p_0}(\mathbb{R}^d)) \xrightarrow{T} & b^{s+t}_{p_0,q_0}(\nabla) \\
I_1 \downarrow & & \downarrow I_2 \\
B^{s}_{q_1}(L_{p_1}(\Omega)) & \xleftarrow{R} B^{s}_{q_1}(L_{p_1}\mathbb{R}^d)) \xleftarrow{T^{-1}} & b^{s}_{p_1,q_1}(\nabla)\,.
\end{array}
\tag{84}
$$

Observe $I_1 = R \circ T^{-1} \circ I_2 \circ T \circ \mathcal{E}$. From (84) and the multiplicativity of $e^{\rm cont}$, cf. (27), we derive

$$
e_n^{\rm cont}(I_1, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega)) \le \|\mathcal{E}\|\, \|T\|\, \|T^{-1}\|\; e_n^{\rm cont}(I_2, b^{s+t}_{p_0,q_0}(\nabla), b^s_{p_1,q_1}(\nabla))\,.
$$

For the converse inequality we choose $\nabla^* = (\nabla^*_j)_j$ such that

$$
\operatorname{supp} \psi_{j,\lambda} \subset \Omega\,, \qquad \lambda \in \nabla^*_j\,, \quad j = -1, 0, 1, \dots\,,
$$

and $\inf_j\, 2^{-jd}\, |\nabla^*_j| > 0$. Then we consider the diagram

$$
\begin{array}{ccc}
b^{s+t}_{p_0,q_0}(\nabla^*) & \xrightarrow{I_2} & b^s_{p_1,q_1}(\nabla^*) \\
{\scriptstyle T^{-1}}\big\downarrow & & \big\uparrow{\scriptstyle T} \\
B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{I_1} & B^s_{q_1}(L_{p_1}(\Omega))\,,
\end{array}
\tag{85}
$$

and conclude

$$
e_n^{\rm cont}(I_2, b^{s+t}_{p_0,q_0}(\nabla^*), b^s_{p_1,q_1}(\nabla^*)) \le \|T\|\, \|T^{-1}\|\; e_n^{\rm cont}(I_1, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega)))\,.
$$

Now Theorem 7 yields the desired result. □

4.3.2 The Widths of Best m-Term Approximation of the Identity

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$. We assume that for any fixed triple (t, p, q) of parameters the spaces $B^{s+t}_q(L_p(\Omega))$ and $B^s_2(L_2(\Omega))$ allow a discretization by one common wavelet system $\mathcal{B}^*$. More exactly, we assume that (106) - (111) are satisfied simultaneously for both spaces, cf. Appendix 5.10. From this it follows that $\mathcal{B}^* \in \mathcal{B}_{C^*}$ for some $1 \le C^* < \infty$.

Theorem 11. *Let Ω be as above. Let $0 < p \le \infty$, $0 < q \le \infty$, $s \in \mathbb{R}$ and*

$$
t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+
$$

holds. Then, for any $C \ge C^*$ we have

$$
e^{\rm non}_{n,C}(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-\frac{t}{d}}\,.
$$

Remark 23. *We also recall the following limiting case. Let* $0 < p < 2$ *and* $t = d(\frac{1}{p} - \frac{1}{2})$*. Then the embedding* $B_p^{s+t}(L_p(\Omega)) \hookrightarrow B_2^s(L_2(\Omega))$ *is continuous but not compact, cf. Proposition* 7*. Here we have*

$$\left(\sum_{n=1}^{\infty}\Big[n^{t/d}\,\sigma_n(u,\mathcal{B}^*)_{B_2^s(L_2(\Omega))}\Big]^p\,\frac{1}{n}\right)^{1/p} < \infty \quad \textit{if and only if} \quad u \in B_p^{s+t}(L_p(\Omega))\,.$$

A proof can be found in [19, Prop. 1], *but the argument there is mainly based on DeVore and Popov* [33], *see also* [31].

Proof of Theorem 11. Let $\mathcal{B}^*$ be a wavelet basis as in Appendix 5.10. Let $\mathcal{B}$ denote the canonical orthonormal basis of $b^0_{2,2}(\nabla)$. We equip the Besov space with the equivalent quasi-norm (111). Observe,

$$\sigma_n(f,\mathcal{B}^*)_{B^s_{p_1,q_1}(\Omega)} \le c\,\sigma_n((< f,\widetilde{\psi}_{j,\lambda} >)_{j,\lambda},\mathcal{B})_{b^s_{p_1,q_1}(\nabla)}\,,$$

where c is one of the constants in (110). By means of Theorem 7 and Remark 2(iii) this implies the estimate from above. The estimate from below follows by combining Theorem 1 and Theorem 10. □

The simple arguments used in the proof of Theorem 11 allow to carry over Remark 23 to the sequence space level, see Remark 17, and vice versa, Theorem 8 to the level of function spaces.

Theorem 12. *Let* Ω *and* $\mathcal{B}^*$ *be as above. Let* $0 < p_0, p_1, q_0, q_1 \le \infty$, $s \in \mathbb{R}$ *and* $t > 0$ *such that* (54) *holds. Then we have*

$$\sup\Big\{\sigma_n(u,\mathcal{B}^*)_{B^s_{q_1}(L_{p_1}(\Omega))}\ :\quad \|\,u|B^{s+t}_{q_0}(L_{p_0}(\Omega))\| \le 1\Big\} \asymp n^{-\frac{t}{d}}\,.$$

Remark 24. *i) For earlier results in this direction we refer to Kashin* [51]*, Oswald* [65]*, Donoho* [37] *and DeVore, Petrova and Temlyakov* [32].

ii) Not all orthonormal systems are of the same quality, see Donoho [37]*. Let us mention the following result of DeVore and Temlyakov* [35]*. Let* $\mathcal{B}^{\#}$ *denote the trigonometric system in* $\mathbb{R}^d$*. By* $B^s_q(L_p(\mathbb{T}^d))$ *we mean the periodic Besov spaces defined on the d-dimensional torus* $\mathbb{T}^d$*. Then we put*

$$t(p_0,p_1) := \begin{cases} d\Big(\frac{1}{p_0} - \frac{1}{p_1}\Big)_+ & \textit{if}\quad 0 < p_0 \le p_1 \le 2 \ \textit{or}\ 1 \le p_1 \le p_0 \le \infty\,, \\ d\,\max\Big(\frac{1}{p_0},\frac{1}{2}\Big) & \textit{otherwise}\,. \end{cases}$$

If $1 \le p_1 \le \infty$, $0 < p_0, q_0 \le \infty$, *and* $t > t(p_0,p_1)$, *then*

$$\sup\Big\{\sigma_n(u,\mathcal{B}^{\#})_{L_{p_1}(\mathbb{T}^d)} : \qquad \|\, u\,|B^t_{q_0}(L_{p_0}(\mathbb{T}^d))\| \le 1\Big\}$$
$$\asymp \begin{cases} n^{-\frac{t}{d}} & \textit{if} \quad p_0 \ge \max(p_1,2)\,, \\ n^{-\frac{t}{d}+\frac{1}{p_0}-\frac{1}{2}} & \textit{if} \quad p_0 \le \max(p_1,2) = 2\,, \\ n^{-\frac{t}{d}+\frac{1}{p_0}-\frac{1}{p_1}} & \textit{if} \quad p_0 \le \max(p_1,2) = p_1\,. \end{cases}$$

4.3.3 The Approximation Numbers of the Identity

Theorem 13. *Let* Ω *be a bounded Lipschitz domain. Let* $0 < p \le \infty$, $0 < q \le \infty$, *and* $s \in \mathbb{R}$. *Suppose that*

$$t > d\Big(\frac{1}{p} - \frac{1}{2}\Big)_+$$

holds. Then we have

$$e^{\mathrm{lin}}_n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp \begin{cases} n^{-\frac{t}{d}} & \textit{if} \quad 0 < p \le 2\,, \\ n^{-\frac{t}{d}+\frac{1}{p}-\frac{1}{2}} & \textit{if} \quad 2 < p \le \infty\,. \end{cases}$$

Proof. The statement becomes a consequence of Theorem 7(ii), Proposition 6, (100) and (101). □

Remark 25. *(i) The proof is constructive. An in order optimal linear approximation is obtained by taking an appropriate partial sum of the wavelet series of* $\mathcal{E}f$, *where* $\mathcal{E}$ *is the linear universal extension operator from Proposition* 6, *cf. Remark* 20 *for the discrete case.*

(ii) This result is well-known. It can be derived from [84] *and* [42, 3.3.2]. *There and in* [7] *also information can be found about what is known for the general situation, i.e.,* $B^s_2(L_2(\Omega))$ *replaced by* $B^s_{q_1}(L_{p_1}(\Omega))$. *However, let us mention that there are many references which had dealt with this problem before, we refer to* [76, Thm. 1.4.2] *and* [78, Thm. 9, p.193] *and the comments given there.*

4.3.4 The Gelfand Widths of the Identity

Theorem 14. *Let* $\Omega \subset \mathbb{R}^d$ *be a bounded Lipschitz domain and let* $0 < q \le \infty$.
(i) *Let* $1 \le p < 2$ *and suppose* $t > d/2$. *Then*

$$d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-t/d}\,.$$

(ii) *Let* $2 < p \le \infty$ *and suppose* $t > 0$. *Then*

$$d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-\frac{t}{d}}\,.$$

(iii) *Let* $0 < p < 1$ *and suppose*

$$t > d\Big(\frac{1}{p} - \frac{1}{2}\Big)\,.$$

Then there exists two constants c_1 *and* c_2 *such that*

$$c_1\, n^{-\frac{t}{d}} \le d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \le c_2 n^{-\frac{t}{d}-1+\frac{1}{p}}\,.$$

Proof. Consider the diagram

$$\begin{array}{ccc} B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{I_1} & B^s_2(L_2(\Omega)) \\ {\scriptstyle T}\Big\downarrow & & \Big\uparrow{\scriptstyle T^{-1}} \\ b^{s+t}_{p_0,q_0}(\nabla) & \xrightarrow{I_2} & b^s_{2,2}(\nabla)\,, \end{array}$$

where T and T^{-1} are defined as in proof of Theorem 10. In view of $I_1 = T^{-1} \circ I_2 \circ T$ it is enough to combine the multiplicativity of the Gelfand numbers and Theorem 9 to derive the estimates from above. For the estimates from below one uses the diagram

$$\begin{array}{ccc} b^{s+t}_{p_0,q_0}(\nabla^*) & \xrightarrow{I_1} & b^s_{2,2}(\nabla^*) \\ {\scriptstyle T}\Big\downarrow & & \Big\uparrow{\scriptstyle T^{-1}} \\ B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{I_2} & B^s_2(L_2(\Omega))\,, \end{array}$$

where ∇^* is defined as in proof of Theorem 10. This completes the proof. □

Remark 26. *Partial results concerning Gelfand numbers of embedding operators may be found in the monographs Pinkus* [70, Chapt. VII, Thm. 1.1], *Tikhomirov* [78, Thm. 39, p. 206], *and Triebel* [81, 4.10.2]. *Let* T *be a compact operator in* $\mathcal{L}(F, E)$, *where* F, E *are arbitrary Banach spaces and let* $d_n(T, F, E)$ *denote the Kolmogorov numbers. Then*

$$d^n(T') = d_n(T)\,, \qquad n \in \mathbb{N}\,,$$

holds, cf. Pietsch [68] *or* [10, Prop. 2.5.6]. *For Kolmogorov numbers the asymptotic behaviour is also known in certain situations, cf.* [70, Chapt. VII, Thm. 1.1], *Tikhomirov* [78, Thm. 10, p. 193], *Triebel* [81, 4.10.2], *and Temlyakov* [76].

4.4 Proofs of Theorems 2, 4, and 5

4.4.1 Proof of Theorem 2

For $s > 0$ we have $H^{-s}(\Omega) = B^{-s}_2(L_2(\Omega))$. Hence, Theorem 13 yields

$$e^{\text{lin}}_n(I, B^{-s+t}_q(L_p(\Omega)), H^{-s}(\Omega)) \asymp \begin{cases} n^{-\frac{t}{d}} & \text{if} \quad 0 < p \le 2\,, \\ n^{-\frac{t}{d}+\frac{1}{p}-\frac{1}{2}} & \text{if} \quad 2 < p \le \infty\,. \end{cases}$$

Since $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism the multiplicativity of the approximation numbers, cf. (27), imply the desired result.

4.4.2 Proof of Theorem 4

Because of $H^{-s}(\Omega) = B_2^{-s}(L_2(\Omega))$ Theorem 11 yields

$$e_{n,C}^{\mathrm{non}}(I, B_q^{-s+t}(L_p(\Omega)), H^{-s}(\Omega)) \asymp n^{-\frac{t}{d}}$$

Since $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism Lemma 3(ii) implies the desired result.

4.4.3 Proof of Theorem 5

All what we need from the wavelet basis is the following estimate for the best n-term approximation in the H^1-norm:

$$(86) \qquad \| u - S_n(f) \|_{H^1(\Omega)} \le c \, \| u \, | B_\tau^{t+1}(L_\tau(\Omega)) \| \, n^{-t/2}, \qquad \frac{1}{\tau} = \frac{t}{2} + \frac{1}{2},$$

see, e.g., [19] (however we could use as well Theorem 12). We therefore have to estimate the Besov norm $B_\tau^\alpha(L_\tau(\Omega))$. Because of $1 < p \le 2$ the embedding $B_p^{k-1}(L_p(\Omega)) \hookrightarrow W_p^{k-1}(\Omega)$ holds, cf. e.g. [82, 2.3.2, 2.5.6]. Hence our right–hand side f is contained in the Sobolev space $W_p^{k-1}(\Omega)$. Therefore we may employ the fact that u can be decomposed into a regular part u_R and a singular part u_S, $u = u_R + u_S$, where $u_R \in W_p^{k+1}(L_p(\Omega))$ and u_S only depends on the shape of the domain and can be computed explicitly, cf. Grisvard [48, Thm. 2.4.3]. We introduce polar coordinates (r_l, θ_l) in the vicinity of each vertex Υ_l and introduce the functions

$$\mathcal{S}_{l,m}(r_l,\theta_l) := \begin{cases} \zeta_l(r_l) r_l^{\lambda_{l,m}} \sin(m\pi\theta_l/\omega_l) & \text{if} \quad \lambda_{l.m} := m\pi/\omega_l \ne \text{integer}\,, \\ \\ \zeta_l(r_l) r_l^{\lambda_{l,m}} [\log r_l \sin(m\pi\theta_l/\omega_l) + \theta_l \cos(m\pi\theta_l/\omega_l)] & \text{otherwise}\,. \end{cases}$$

Here ζ_l, $l = 1, \ldots, N$ denote suitable C^∞ truncation functions and m is a natural number. Then for $f \in W_p^{k-1}(\Omega)$, one has

$$(87) \qquad u_S = \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} c_{l,m} \, \mathcal{S}_{l.m}\,,$$

provided that no $\lambda_{l,m}$ is equal to $k + 1 - 2/p$. This means that the finite number of singularity functions that is needed depends on the scale of spaces we are interested in, i.e., on the smoothness parameter k. According to (86), we have to estimate the Besov regularity of both, u_S and u_R, in the specific scale

$$B_\tau^{t+1}(L_\tau(\Omega)), \qquad \frac{1}{\tau} = \frac{t}{2} + \frac{1}{2}\,.$$

Since $u_R \in W_p^{k+1}(\Omega)$, the boundedness of Ω implies the embedding

$$W_p^{k+1}(\Omega) \hookrightarrow B_q^{k+1-\delta}(L_q(\Omega)), \qquad \delta > 0\,, \quad 0 < q \le p\,, \quad k+1 > 2\Big(\frac{1}{q} - \frac{1}{2}\Big)\,.$$

Hence

$$u_R \in B_\tau^{k+1-\delta}(L_\tau(\Omega)), \qquad \frac{1}{\tau} = \frac{(k-\delta)}{2} + \frac{1}{2} \qquad \text{for arbitrary small} \quad \delta > 0\,. \tag{88}$$

Moreover, it has been shown in [15](see also Remark 28) that the functions $\mathcal{S}_{l,m}$ defined above satisfy

$$\mathcal{S}_{l,m}(r_l, \theta_l) \in B_q^{\frac{1}{2}+\frac{2}{q}}(L_q(\Omega)), \qquad \text{for all} \quad 0 < q < \infty\,. \tag{89}$$

By combining (88) and (89) we see that

$$u \in B_\tau^{k+1-\delta}(L_\tau(\Omega)) \qquad \frac{1}{\tau} = \frac{(k-\delta)}{2} + \frac{1}{2} \qquad \text{for arbitrary small} \quad \delta > 0.$$

To derive an estimate uniformly with respect to the unit ball in $B_p^{k-1}(L_p(\Omega))$ we argue as follows. We put

$$\mathcal{N} := \operatorname{span}\Big\{\mathcal{S}_{l,m}(r_l, \theta_l) : \quad 0 < \lambda_{m,l} < k+1-2/p\,,\ l = 1, \ldots, N\Big\}\,.$$

Let γ_l be the trace operator with respect to the segment Γ_l. Grisvard has shown that Δ maps

$$H := \Big\{u \in W_p^{k+1}(\Omega) : \quad \gamma_l u = 0\,,\ l = 1, \ldots, N\Big\} + \mathcal{N}$$

onto $W_p^{k-1}(\Omega)$, cf. [47, Thm. 5.1.3.5]. This mapping is also injective, see [47, Lemma 4.4.3.1, Rem. 5.1.3.6]. We equip the space H with the norm

$$\|\, u \,\|_H := \|\, u_R + u_S \,\|_H = \|\, u_R \,\|_{W_p^{k+1}(\Omega)} + \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} |c_{l,m}|\,,$$

see (87). Then it becomes a Banach space. Furthermore, $\Delta : H \to W_p^{k-1}(\Omega)$ is continuous. Banach's continuous inverse theorem implies that the solution operator is continuous considered as a mapping from $W_p^{k-1}(\Omega)$ onto H. Observe

$$\|\, u_R + u_S \,\|_{B_\tau^{k+1-\delta}(L_\tau(\Omega))} \le C\Big(\|\, u_R \,\|_{W_p^{k+1}(\Omega)} + \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} |c_{l,m}|\Big)$$

with some constant C independent of u. □

5 Appendix – Besov spaces

Here we collect some properties of Besov spaces which have been used in the text before. Detailed references will be given. For general information on Besov spaces we refer to the monographs [59, 60, 66, 71, 82, 83].

5.1 Besov Spaces on $\mathbb{R}^d$ and Differences

Nowadays Besov spaces are widely used in several branches of mathematics. Probably the most common way to introduce these classes makes use of differences. For $M \in \mathbb{N}$, $h \in \mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{C}$ we define

$$\Delta_h^M f(x) := \sum_{j=0}^{M} \binom{M}{j} (-1)^{M-j} f(x + jh).$$

Let $0 < p \leq \infty$. The corresponding modulus of smoothness is then given by

$$\omega^M(t, f)_p := \sup_{|h|<t} \| \Delta_h^M f \|_{L_p(\mathbb{R}^d)} , \qquad t > 0 .$$

One approach to introduce Besov spaces is the following.

Definition 4. *Let $s > 0$ and $0 < p, q \leq \infty$. Let M be a natural number satisfying $M > s$. Then $\Lambda_q^s(L_p(\mathbb{R}^d))$ is the collection of all functions $f \in L_p(\mathbb{R}^d)$ such that*

$$| f |_{\Lambda_q^s(L_p(\mathbb{R}^d))} := \left(\int_0^\infty \left[t^{-s} \, \omega^M(t, f)_p \right]^q \frac{dt}{t} \right)^{1/q} < \infty$$

if $q < \infty$ and

$$| f |_{\Lambda_\infty^s(L_p(\mathbb{R}^d))} := \sup_{t>0} t^{-s} \, \omega^M(t, f)_p < \infty$$

if $q = \infty$. These classes are equipped with a quasi-norm by taking

$$\| f \|_{\Lambda_q^s(L_p(\mathbb{R}^d))} := \| f \|_{L_p(\mathbb{R}^d)} + | f |_{\Lambda_q^s(L_p(\mathbb{R}^d))} .$$

Remark 27. *It turns out that these classes do not depend on M, cf.* [34].

Remark 28. *Let $\varrho \in C_0^\infty(\mathbb{R}^d)$ be a function such that $\varrho(0) \neq 0$. By means of the above definition it is not complicated to show that a function*

$$f_\alpha(x) := |x|^\alpha \, \varrho(x) , \qquad x \in \mathbb{R}^d , \quad \alpha > 0 ,$$

belongs to $\Lambda_\infty^{\alpha+d/p}(L_p(\mathbb{R}^d))$ and this is best possible (if α is not an even natural number), cf. [71, 2.3.1] *for details. A minor modification yields*

$$f_{\alpha,\beta}(x) := |x|^\alpha \, (\log |x|)^\beta \, \varrho(x) , \qquad x \in \mathbb{R}^d , \quad \alpha, \, \beta > 0 ,$$

belongs to $\Lambda_\infty^{\alpha+d/p-\varepsilon}(L_p(\mathbb{R}^d))$ for all ε, $0 < \varepsilon < \alpha + d/p$.

5.2 Besov Spaces on $\mathbb{R}^d$ and Littlewood-Paley Characterizations

Since we are using also spaces with negative smoothness $s < 0$ and/or $p, q < 1$ we shall give a further definition, which relies on Fourier analysis. We use it here for introductory purposes. This approach makes use of smooth dyadic decompositions of unity. Let $\varphi \in C_0^\infty(\mathbb{R}^d)$ be a function such that $\varphi(x) = 1$ if $|x| \le 1$ and $\varphi(x) = 0$ if $|x| \ge 2$. Then we put

$$\varphi_0(x) := \varphi(x), \qquad \varphi_j(x) := \varphi(2^{-j}x) - \varphi(2^{-j+1}x)\,, \quad j \in \mathbb{N}\,. \tag{90}$$

It follows

$$\sum_{j=0}^{\infty} \varphi_j(x) = 1\,, \qquad x \in \mathbb{R}^d\,,$$

and

$$\mathrm{supp}\ \varphi_j \subset \Big\{ x \in \mathbb{R}^d : \quad 2^{j-2} \le |x| \le 2^{j+1} \Big\}\,, \qquad j = 1, 2, \ldots\,.$$

Let $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, both defined on $\mathcal{S}'(\mathbb{R}^d)$. For $f \in \mathcal{S}'(\mathbb{R}^d)$ we consider the sequence $\mathcal{F}^{-1}[\varphi_j(\xi)\,\mathcal{F}f(\xi)](x)$, $j \in \mathbb{N}_0$, of entire analytic functions. By means of these functions we define the Besov classes.

Definition 5. *Let $s \in \mathbb{R}$ and $0 < p, q \le \infty$. Then $B^s_q(L_p(\mathbb{R}^d))$ is the collection of all tempered distributions f such that*

$$\|\, f \,|B^s_q(L_p(\mathbb{R}^d))\| = \Big(\sum_{j=0}^{\infty} 2^{sjq}\, \|\mathcal{F}^{-1}[\varphi_j(\xi)\,\mathcal{F}f(\xi)](\,\cdot\,)\,|L_p(\mathbb{R}^d)\|^q \Big)^{1/q} < \infty$$

if $q < \infty$ and

$$\|\, f|B^s_\infty(L_p(\mathbb{R}^d))\| = \sup_{j=0,1,\ldots}\ 2^{sj}\, \|\mathcal{F}^{-1}[\varphi_j(\xi)\,\mathcal{F}f(\xi)](\,\cdot\,)\,|L_p(\mathbb{R}^d)\| < \infty$$

if $q = \infty$.

Remark 29. *i) If no confusion is possible we drop $\mathbb{R}^d$ in notations.*

ii) These classes are quasi-Banach spaces. They do not depend on the chosen function φ (up to equivalent quasi-norms). If $t = \min(1, p, q)$, then

$$\|\, f + g\,|B^s_q(L_p)\|^t \le \|\, f\,|B^s_q(L_p)\|^t + \|\, g\,|B^s_q(L_p)\|^t$$

holds for all $f, g \in B^s_q(L_p)$.

Proposition 4. [82, 2.5.12]. *Let $0 < p, q \le \infty$ and $s > d\,\max(0, 1/p - 1)$. Then we have coincidence of $\Lambda^s_q(L_p)$ and $B^s_q(L_p)$ in the sense of equivalent quasi-norms.*

Remark 30. *i) For $s \le d \max(0, 1/p-1)$ we have $\Lambda^s_q(L_p) \neq B^s_q(L_p)$. E.g., the Dirac distribution δ belongs to $B^{d(\frac{1}{p}-1)}_\infty(L_p)$, cf.* [71, 2.3.1].

ii) Smooth cut-off functions are pointwise multipliers for all Besov spaces. More exactly, let $\psi \in \mathcal{D}$. Then the product $\psi\, f$ belongs to $B^s_q(L_p)$ for any $f \in B^s_q(L_p)$ and there exists a constant c such that

$$\| \psi\, f \, | B^s_q(L_p)\| \le c \| \, f \, | B^s_q(L_p)\|$$

holds, see e.g. [82, 2.8], [71, 4.7].

5.3 Wavelet Characterizations

For the construction of biorthogonal wavelet bases as considered below we refer to the recent monograph of Cohen [11, Chapt. 2]. Let φ be a compactly supported scaling function of sufficiently high regularity and let ψ_i, $i = 1, \ldots 2^d - 1$ be corresponding wavelets. More exactly, we suppose for some $N > 0$ and $r \in \mathbb{N}$

$$\begin{aligned} &\text{supp}\, \varphi\,,\ \text{supp}\, \psi_i \quad \subset \quad [-N,N]^d\,, \qquad i = 1, \ldots, 2^d-1\,, \\ &\varphi, \psi_i \in C^r(\mathbb{R}^d)\,, \qquad i = 1, \ldots, 2^d-1\,, \\ &\int x^\alpha\, \psi_i(x)\, dx = 0 \qquad \text{for all} \quad |\alpha| \le r\,, \qquad i = 1, \ldots, 2^d-1\,, \end{aligned}$$

and

$$\varphi(x-k),\ 2^{jd/2}\, \psi_i(2^j x - k)\,, \qquad j \in \mathbb{N}_0\,, \quad k \in \mathbb{Z}^d\,,$$

is a Riesz basis in $L_2(\mathbb{R}^d)$. We shall use the standard abbreviations

$$\psi_{i,j,k}(x) = 2^{jd/2}\, \psi_i(2^j x - k) \qquad \text{and} \qquad \varphi_k(x) = \varphi(x-k)\,.$$

Further, the dual Riesz basis should fulfill the same requirements, i.e. there exist functions $\widetilde{\varphi}$ and $\widetilde{\psi}_i$, $i = 1, \ldots, 2^d - 1$, such that

$$\begin{aligned} \langle \widetilde{\varphi}_k, \psi_{i,j,k} \rangle &= \langle \widetilde{\psi}_{i,j,k}, \varphi_k \rangle = 0\,, \\ \langle \widetilde{\varphi}_k, \varphi_\ell \rangle &= \delta_{k,\ell} \qquad \text{(Kronecker symbol)}\,, \\ \langle \widetilde{\psi}_{i,j,k}, \psi_{u,v,\ell} \rangle &= \delta_{i,u}\, \delta_{j,v}\, \delta_{k,\ell}\,, \\ \text{supp}\, \widetilde{\varphi}\,, &\qquad \text{supp}\, \widetilde{\psi}_i \quad \subset \quad [-N,N]^d\,, \qquad i = 1, \ldots, 2^d-1\,, \\ \widetilde{\varphi}, \widetilde{\psi}_i &\in C^r(\mathbb{R}^d)\,, \qquad i = 1, \ldots, 2^d-1\,, \\ \int x^\alpha\, \widetilde{\psi}_i(x)\, dx &= 0 \qquad \text{for all} \quad |\alpha| \le r\,, \qquad i = 1, \ldots, 2^d-1\,. \end{aligned}$$

For $f \in \mathcal{S}'(\mathbb{R}^d)$ we put

$$\langle f, \psi_{i,j,k} \rangle = f(\overline{\psi_{i,j,k}}) \qquad \text{and} \qquad \langle f, \varphi_k \rangle = f(\overline{\varphi_k})\,, \tag{91}$$

whenever this makes sense.

Proposition 5. *Let $s \in \mathbb{R}$ and $0 < p, q \leq \infty$. Suppose*

$$r > \max\Big(s, \frac{2d}{p} + \frac{d}{2} - s\Big)\,. \tag{92}$$

Then $B^s_q(L_p)$ is the collection of all tempered distributions f such that f is representable as

$$f = \sum_{k\in\mathbb{Z}^d} a_k\, \varphi_k + \sum_{i=1}^{2^d-1} \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}^d} a_{i,j,k}\, \psi_{i,j,k} \qquad (\text{convergence in} \quad \mathcal{S}')$$

with

$$\| f \,|B^s_q(L_p)\|^* := \Big(\sum_{k\in\mathbb{Z}^d} |a_k|^p\Big)^{1/p} + \Big(\sum_{i=1}^{2^d-1} \sum_{j=0}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \Big(\sum_{k\in\mathbb{Z}^d} |a_{i,j,k}|^p\Big)^{q/p}\Big)^{1/q} < \infty\,,$$

if $q < \infty$ and

$$\| f \,|B^s_\infty(L_p)\|^* := \Big(\sum_{k\in\mathbb{Z}^d} |a_k|^p\Big)^{1/p} + \sup_{i=1,\ldots,2^d-1}\ \sup_{j=0,\ldots}\ 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))} \Big(\sum_{k\in\mathbb{Z}^d} |a_{i,j,k}|^p\Big)^{1/p} < \infty\,.$$

The representation is unique and

$$a_{i,j,k} = \langle f, \widetilde{\psi}_{i,j,k}\rangle \qquad \text{and} \qquad a_k = \langle f, \widetilde{\varphi}_k\rangle$$

hold. Further $I : f \mapsto \{\langle f, \widetilde{\varphi}_k\rangle,\ \langle f, \widetilde{\psi}_{i,j,k}\rangle\}$ is an isomorphic map of $B^s_q(L_p(\mathbb{R}^d))$ onto the sequence space (equipped with the quasi-norm $\|\cdot|B^s_q(L_p)\|^*$), i.e. $\|\cdot|B^s_q(L_p)\|^*$ may serve as an equivalent quasi-norm on $B^s_q(L_p)$.

Remark 31. *i) The restriction* (92) *is guaranteeing that* (91) *makes sense for all $f \in B^s_q(L_p)$.*

ii) It is immediate from this proposition that the functions $\varphi_k, \psi_{i,j,k}, k \in \mathbb{Z}^d, 1 \leq i \leq 2^d - 1, j \in \mathbb{N}_0$ form a basis for $B^s_q(L_p)$ if $\max(p,q) < \infty$. *By the same reasonings the functions*

$$\varphi_k, \quad 2^{-js}\, \psi_{i,j,k}, \qquad k \in \mathbb{Z}^d, \quad 1 \leq i \leq 2^d - 1, \quad j \in \mathbb{N}_0\,,$$

form a Riesz basis for $B^s_2(L_2)$.

iii) If the wavelet basis is orthonormal (in L_2), then this proposition is proved in Triebel [85]. *But the comments made in Subsection* 3.4 *of the quoted paper make clear that this extends to the situation considered in Proposition* 5. *A different proof, but restricted to $s > d(\frac{1}{p} - 1)_+$, is given in* [11, Thm. 3.7.7]. *However, there are many forerunners with some restrictions concerning s, p and q. We refer to* [6] *and* [59].

5.4 Besov Spaces on Domains – the Approach via Restrictions

There are at least two different approaches to define function spaces on domains. One approach uses restrictions to Ω of functions defined on $\mathbb{R}^d$. So, all calculations are done on $\mathbb{R}^d$. The other approach introduces theses spaces by means of local quantities defined only in Ω. For numerical purposes the second approach is more promising whereas for analytic investigations the first one looks more elegant. Here we discuss both, since both were used.

Let $\Omega \subset \mathbb{R}^d$ be an bounded open nonempty set. Then we define $B^s_q(L_p(\Omega))$ to be the collection of all distributions $f \in \mathcal{D}'(\Omega)$ such that there exists a tempered distribution $g \in B^s_q(L_p(\mathbb{R}^d))$ satisfying

$$f(\varphi) = g(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\Omega)\,,$$

i.e. $g|_\Omega = f$ in $\mathcal{D}'(\Omega)$. We put

$$\|\, f\,|B^s_q(L_p(\Omega))\| := \inf \|\, g\,|B^s_q(L_p(\mathbb{R}^d))\|\,,$$

where the infimum is taken with respect to all distributions g as above.

Let $\operatorname{diam}\Omega$ be the diameter of the set Ω and let x^0 be a point with the property

$$\Omega \subset \Big\{y:\ |x^0 - y| \le \operatorname{diam}\Omega\Big\}\,.$$

Such a point we shall call a center of Ω. Since smooth cut-off functions are pointwise multipliers, cf. Remark 30, we can associate to any $f \in B^s_q(L_p(\Omega))$ a tempered distribution $g \in B^s_q(L_p)$ such that $g|_\Omega = f$ in $\mathcal{D}'(\Omega)$,

$$C\,\|\, g\,|B^s_q(L_p)\| \ \le\ \|\, f\,|B^s_q(L_p(\Omega))\| \le \|\, g\,|B^s_q(L_p)\| \tag{93}$$

$$\operatorname{supp} g \ \subset\ \{x \in \mathbb{R}^d:\quad |x - x^0| \le 2\operatorname{diam}\Omega\}\,. \tag{94}$$

Here $0 < C < 1$ does not depend on f (but on Ω, s, p, q).

Now we turn to decompositions by means of wavelets. We use the notation from the preceeding subsection. Define

$$\Lambda_j := \Big\{k \in \mathbb{Z}^d:\quad |k_i - x^0_i| \le 2^j \operatorname{diam}\Omega + N\,, i = 1, \ldots, d\Big\}\,, \qquad j = 0, 1, \ldots\,. \tag{95}$$

Then, given f and taking g as above we find

$$g = \sum_{k\in\Lambda_0} \langle g, \widetilde{\varphi}_k\rangle\, \varphi_k + \sum_{i=1}^{2^d-1}\sum_{j=0}^{\infty}\sum_{k\in\Lambda_j} \langle g, \widetilde{\psi}_{i,j,k}\rangle\, \psi_{i,j,k} \qquad (\text{convergence in}\quad \mathcal{S}') \tag{96}$$

and

$$\| g \,|B^s_q(L_p)\| \asymp \Big(\sum_{k\in\Lambda_0} |\langle g, \widetilde{\varphi}_k\rangle|^p \Big)^{1/p} + \Big(\sum_{i=1}^{2^d-1} \sum_{j=0}^{\infty} 2^{jq(s+d(\frac{1}{2}-\frac{1}{p}))} \Big(\sum_{k\in\Lambda_j} |\langle g, \widetilde{\psi}_{i,j,k}\rangle|^p \Big)^{q/p} \Big)^{1/q} < \infty \, . \tag{97}$$

The following more handy notation is also used. We put

$$\nabla_{-1} := \Lambda_0 \tag{98}$$

$$\nabla_j := \Big\{ (i,k) : \quad 1 \le i \le 2^d - 1 \, , \quad k \in \Lambda_j \Big\} \, , \quad j = 0, 1, \ldots \, , \tag{99}$$

$\psi_{j,\lambda} := \psi_{i,j,k}$, if $\lambda = (i,k) \in \nabla_j$, $j \in \mathbb{N}_0$, and $\psi_{j,\lambda} := \varphi_k$ if $\lambda = k \in \nabla_{-1}$. Similarly in case of the dual basis. Then (96), (97) read as

$$g = \sum_{j=-1}^{\infty} \sum_{\lambda\in\nabla_j} \langle g, \widetilde{\psi}_{j,\lambda}\rangle \, \psi_{j,\lambda} \qquad \text{(convergence in } \quad \mathcal{S}'\text{)} \tag{100}$$

and

$$\| g \,|B^s_q(L_p)\| \asymp \Big(\sum_{j=-1}^{\infty} 2^{jq(s+d(\frac{1}{2}-\frac{1}{p}))} \Big(\sum_{\lambda\in\nabla_j} |\langle g, \widetilde{\psi}_{j,\lambda}\rangle|^p \Big)^{q/p} \Big)^{1/q} < \infty \, . \tag{101}$$

5.5 Lipschitz Domains, Embeddings, and Interpolation

We call a domain Ω a *special Lipschitz domain* (see Stein [74]), if Ω is an open set in $\mathbb{R}^d$ and if there exists a function $\omega : \mathbb{R}^{d-1} \to \mathbb{R}$ such that

$$\Omega = \Big\{ (x', x_d) \in \mathbb{R}^d : \, x_d > \omega(x') \Big\}$$

and

$$|\, \omega(x') - \omega(y') \,| \le C \, |x' - y'| \qquad \text{for all} \quad x', \, y' \in \mathbb{R}^{d-1} \, ,$$

and some constant $C > 0$. We call a domain Ω a *bounded Lipschitz domain* if Ω is bounded and its boundary $\partial\Omega$ can be covered by a finite number of open balls B_k, so that, possibly after a proper rotation, $\partial\Omega \cap B_k$ for each k is a part of the graph of a Lipschitz function.

Proposition 6. *Let $\Omega \in \mathbb{R}^d$ be a bounded Lipschitz domain with center x^0. Then there exists a universal bounded linear extension operator $\mathcal{E}$ for all values of s, p, and q, i.e.*

$$(\mathcal{E} f)|_\Omega = f \qquad \textit{for all} \quad f \in B^s_q(L_p(\Omega)) \, ,$$

and

$$\| \mathcal{E} \, : \, B^s_q(L_p(\Omega)) \to B^s_q(L_p(\mathbb{R}^d)) \| < \infty \, .$$

In addition we may assume

$$\text{supp } \mathcal{E}f \subset \{x \in \mathbb{R}^d : \quad |x - x^0| \le 2 \operatorname{diam} \Omega\} \, . \tag{102}$$

Remark 32. *Proposition* 6 *has been proved by Rychkov* [72]*. Property* (102) *follows from Remark* 30*.*

Let us now discuss some embedding properties of Besov spaces that are needed for our purposes.

Proposition 7. *Let $\Omega \subset \mathbb{R}^d$ be an bounded open set. Let $0 < p_0, p_1, q_0, q_1 \le \infty$ and let $s, t \in \mathbb{R}$. Then the embedding*

$$I : B^{s+t}_{q_0}(L_{p_0}(\Omega)) \to B^s_{q_1}(L_{p_1}(\Omega))$$

is compact if and only if

$$t > d \Big(\frac{1}{p_0} - \frac{1}{p_1}\Big)_+ \, . \tag{103}$$

Remark 33. *Sufficiency is proved e.g. in* [42]*. The necessity of the given restrictions is almost obvious, but see Lemma 4 and* [54] *for details.*

Sometimes Besov spaces or Sobolev spaces of fractional order are introduced by means of interpolation (real and/or complex). Here we state following, cf. [84]. As usual, $(\,\cdot\,,\,\cdot\,)_{\Theta,q}$ and $[\,\cdot\,,\,\cdot\,]_\Theta$ denote the real and the complex interpolation functor, respectively.

Proposition 8. *Let Ω be a bounded Lipschitz domain. Let $0 < q_0, q_1 \le \infty$ and let $s_0, s_1 \in \mathbb{R}$. Let $0 < \Theta < 1$.*
(i) *Let $0 < p, q \le \infty$. Suppose $s_0 \neq s_1$ and put $s = (1-\Theta)\, s_0 + \Theta\, s_1$. Then*

$$\Big(B^{s_0}_{q_0}(L_p(\Omega)), B^{s_1}_{q_1}(L_p(\Omega))\Big)_{\Theta,q} = B^s_q(L_p(\Omega)) \qquad \textit{(equivalent quasi-norms)}.$$

(ii) *Let $0 < p_0, p_1 \le \infty$. We put $s = (1-\Theta)\, s_0 + \Theta\, s_1$,*

$$\frac{1}{p} = \frac{1-\Theta}{p_0} + \frac{\Theta}{p_1} \qquad \textit{and} \qquad \frac{1}{q} = \frac{1-\Theta}{q_0} + \frac{\Theta}{q_1} \, .$$

Then

$$\Big[B^{s_0}_{q_0}(L_{p_0}(\Omega)), B^{s_1}_{q_1}(L_{p_1}(\Omega))\Big]_\Theta = B^s_q(L_p(\Omega)) \qquad \textit{(equivalent quasi-norms)}.$$

5.6 Besov Spaces on Domains – Intrinsic Descriptions

For $M \in \mathbb{N}$, $h \in \mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{C}$ we define

$$\Delta_h^M f(x) := \begin{cases} \sum_{j=0}^{M} \binom{M}{j} (-1)^{M-j} f(x+jh) & \text{if} \quad x, x+h, \ldots, x+Mh \in \Omega\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

The corresponding modulus of smoothness is then given by

$$\omega^M(t,f)_p := \sup_{|h|<t} \| \Delta_h^M f \|_{L_p(\Omega)}\,, \qquad t > 0\,.$$

The approach by differences coincides with that using restrictions as can be seen by the recent result of Dispa [36].

Proposition 9. *Let Ω be a bounded Lipschitz domain. Let $M \in \mathbb{N}$. Let $0 < p, q \le \infty$ and $d \max(0, \frac{1}{p} - 1) < s < M$. Then*

$$\begin{aligned} B_q^s(L_p(\Omega)) \quad = \quad & \Big\{ f \in L_{\max(p,1)}(\Omega) : \\ & \|f\|^{\square} := \|f\|_{L_p(\Omega)} + \left(\int_0^1 \Big[t^{-s}\, \omega^M(t,f)_p \Big]^q \frac{dt}{t} \right)^{1/q} < \infty \Big\} \end{aligned}$$

in the sense of equivalent quasi-norms.

5.7 Sobolev Spaces on Domains

Let Ω be a bounded Lipschitz domain. Let $m \in \mathbb{N}$. As usual $H^m(\Omega)$ denotes the collection of all functions f such that the distributional derivatives $D^\alpha f$ of order $|\alpha| \le m$ belong to $L_2(\Omega)$. The norm is defined as

$$\| f \,|H^m(\Omega)\| := \sum_{|\alpha| \le m} \| D^\alpha f \,|L_2(\Omega)\|\,.$$

It is well-known that $H^m(\mathbb{R}^d) = B_2^m(L_2(\mathbb{R}^d))$ in the sense of equivalent norms, cf. e.g. [82]. As a consequence of the existence of a bounded linear extension operator for Sobolev spaces on bounded Lipschitz domains, cf. [74, p. 181], it follows

$$H^m(\Omega) = B_2^m(L_2(\Omega)) \qquad \text{(equivalent norms)}\,,$$

for such domains. For fractional $s > 0$ we introduce the classes by complex interpolation. Let $0 < s < m$, $s \notin \mathbb{N}$. Then, following [56, 9.1], we define

$$H^s(\Omega) := \Big[H^m(\Omega), L_2(\Omega) \Big]_\Theta\,, \qquad \Theta = 1 - \frac{s}{m}\,.$$

This definition does not depend on m in the sense of equivalent norms. This follows immediately from

$$\Big[H^m(\Omega), L_2(\Omega)\Big]_\Theta = \Big[B_2^m(L_2(\Omega)), B_2^0(L_2(\Omega))\Big]_\Theta = B_2^s(L_2(\Omega))\,, \qquad \Theta = 1 - \frac{s}{m}\,.$$

(all in the sense of equivalent norms), cf. Proposition 8.

5.8 Function Spaces on Domains and Boundary Conditions

We concentrate on homogeneous boundary conditions. Here it makes sense to introduce two further scales of function spaces (distribution spaces).

Definition 6. *Let $\Omega \subset \mathbb{R}^d$ be an open nontrivial set. Let $s \in \mathbb{R}$ and $0 < p, q \le \infty$.*
(i) *Then $\mathring{B}^s_q(L_p(\Omega))$ denotes the closure of $\mathcal{D}(\Omega)$ in $B^s_q(L_p(\Omega))$, equipped with the quasi-norm of $B^s_q(L_p(\Omega))$.*
(ii) *Let $s \ge 0$. Then $H^s_0(\Omega)$ denotes the closure of $\mathcal{D}(\Omega)$ in $H^s(\Omega)$, equipped with the norm of $H^s(\Omega)$.*
(iii) *By $\widetilde{B}^s_q(L_p(\Omega))$ we denote the collection of all $f \in \mathcal{D}'(\Omega)$ such that there is a $g \in B^s_q(L_p(\mathbb{R}^d))$ with*

$$g_{|\Omega} = f \qquad \text{and} \qquad \operatorname{supp} g \subset \overline{\Omega}\,, \tag{104}$$

equipped with the quasi-norm

$$\|\, f \,|\widetilde{B}^s_q(L_p(\Omega))\| = \inf \|\, g \,|B^s_q(L_p(\mathbb{R}^d))\|\,,$$

where the infimum is taken over all such distributions g as in (104).

Remark 34. *For a bounded Lipschitz domain it holds $\mathring{B}^s_q(L_p(\Omega)) = \widetilde{B}^s_q(L_p(\Omega)) = B^s_q(L_p(\Omega))$ if*

$$0 < p, q < \infty\,, \quad \max\Big(\frac{1}{p} - 1,\, d\Big(\frac{1}{p} - 1\Big)\Big) < s < \frac{1}{p}\,,$$

cf. [47, Cor. 1.4.4.5] *and* [84]. *Hence,*

$$H^s_0(\Omega) = \mathring{B}^s_2(L_2(\Omega)) = \widetilde{B}^s_2(L_2(\Omega)) = B^s_2(L_2(\Omega)) = H^s(\Omega)$$

if $0 \le s < 1/2$.

Often it is more convenient to work with a scale $\overline{B}^s_q(L_p(\Omega))$, originally introduced in [84].

Definition 7. *Let $\Omega \subset \mathbb{R}^d$ be an open nontrivial set. Let $s \in \mathbb{R}$ and $0 < p, q \leq \infty$. Then we put*

$$\overline{B}^s_q(L_p(\Omega)) := \begin{cases} B^s_q(L_p(\Omega) & \text{if} \quad s < 1/p\,, \\ \widetilde{B}^s_q(L_p(\Omega)) & \text{if} \quad s \geq 1/p\,. \end{cases}$$

This scale $\overline{B}^s_q(L_p(\Omega))$ is well-behaved under interpolation and duality, cf. [84].

Proposition 10. *Let Ω be a bounded Lipschitz domain. Let $1 < p, p_0, p_1, q, q_0, q_1 < \infty$ and let $s_0, s_1 \in \mathbb{R}$. Let $0 < \Theta < 1$.*
(i) *Suppose $s_0 \neq s_1$ and put $s = (1-\Theta)\, s_0 + \Theta\, s_1$. Then*

$$\Big(\overline{B}^{s_0}_{q_0}(L_p(\Omega)), \overline{B}^{s_1}_{q_1}(L_p(\Omega))\Big)_{\Theta,q} = \overline{B}^s_q(L_p(\Omega)) \qquad \textit{(equivalent quasi-norms)}.$$

(ii) *We put $s = (1-\Theta)\, s_0 + \Theta\, s_1$,*

$$\frac{1}{p} = \frac{1-\Theta}{p_0} + \frac{\Theta}{p_1} \qquad \text{and} \qquad \frac{1}{q} = \frac{1-\Theta}{q_0} + \frac{\Theta}{q_1}\,.$$

Then

$$\Big[\overline{B}^{s_0}_{q_0}(L_{p_0}(\Omega)), \overline{B}^{s_1}_{q_1}(L_{p_1}(\Omega))\Big]_{\Theta} = \overline{B}^s_q(L_p(\Omega)) \qquad \textit{(equivalent quasi-norms)}.$$

(iii) *With $s \in \mathbb{R}$ and*

$$1 = \frac{1}{p} + \frac{1}{p'} \qquad \text{and} \qquad 1 = \frac{1}{q} + \frac{1}{q'}$$

we find

$$\Big(\overline{B}^s_q(L_p(\Omega))\Big)' = \overline{B}^{-s}_{q'}(L_{p'}(\Omega))\,.$$

Here the duality must be understood in the framework of the dual pairing $(\mathcal{D}(\Omega), \mathcal{D}'(\Omega)$.

5.9 Sobolev Spaces with Negative Smoothness

Definition 8. *For $s > 0$ we define*

$$H^{-s}(\Omega) := \begin{cases} \Big(H^s_0(\Omega)\Big)' & \text{if} \quad s - \frac{1}{2} \neq \text{integer}\,, \\ & \\ \Big(\widetilde{B}^s_2(L_2(\Omega))\Big)' & \text{otherwise}\,. \end{cases}$$

Remark 35. *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Then*

$$H^s_0(\Omega) = \widetilde{B}^s_2(L_2(\Omega))\,, \qquad s > 0\,, \quad s - \frac{1}{2} \neq \text{integer}\,,$$

cf. [47, Cor. 1.4.4.5] *and Proposition* 9*. From Remark* 34 *and Proposition* 10 *we conclude the identity*

$$H^{-s}(\Omega) = B_2^{-s}(L_2(\Omega)) , \qquad s > 0 , \tag{105}$$

to be understood in the sense of equivalent norms.

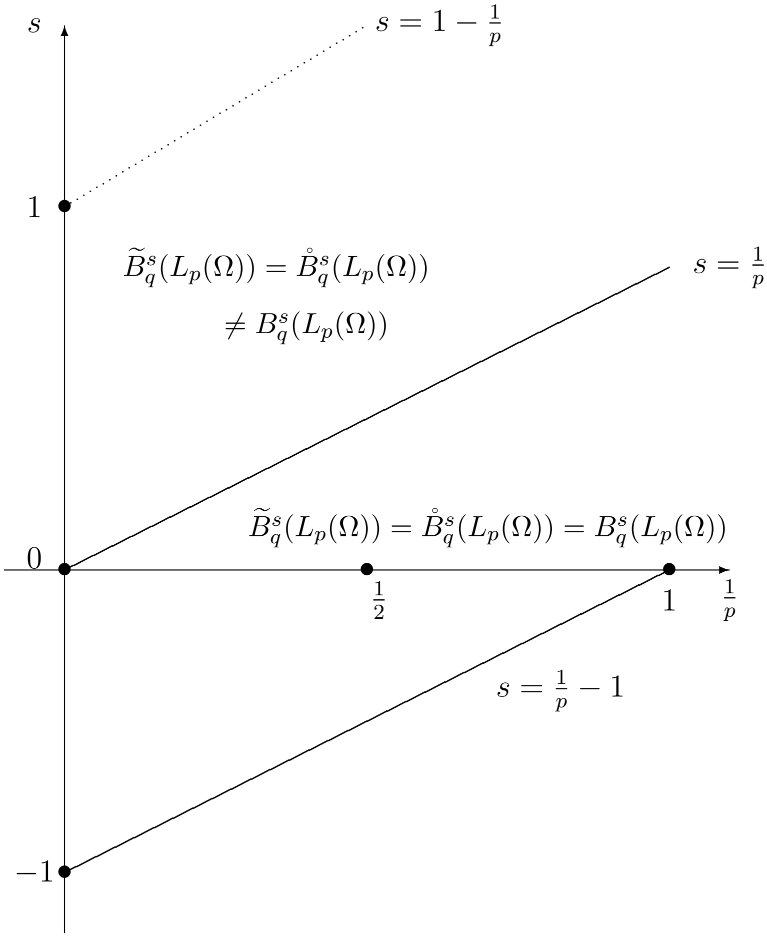

Fig. 1

Remark 36. [81, 4.3.2]. *Let* Ω *be a bounded open set with a smooth boundary. Then* $\mathring{B}^s_q(L_p(\Omega)) = \widetilde{B}^s_q(L_p(\Omega))$ *holds if*

$$1 < p, q < \infty , \quad \frac{1}{p} - 1 < s < \infty , \quad s - \frac{1}{p} \neq \textit{integer} .$$

5.10 Wavelet Characterization of Besov Spaces on Domains

It is a difficult task to construct wavelet bases on domains, see [11, 2.12] and the references given there. Under certain conditions on the domain Ω such constructions with properties similar to (100), (101) are known in the literature, see Remark 11 above.

Let Ω be a bounded open set in $\mathbb{R}^d$. Let p, q and s be fixed such that $s > d\max(0, \frac{1}{p} - 1)$. We suppose that there exist sets $\nabla_j \subset \{1, 2, \dots, 2^d - 1\} \times \mathbb{Z}^d$,

$$0 < \inf_{j=-1,0,\dots} 2^{-jd}\,|\nabla_j| \le \sup_{j=-1,0,\dots} 2^{-jd}\,|\nabla_j| < \infty\,, \tag{106}$$

and functions $\psi_{j,\lambda}$, $\widetilde{\psi}_{j,\lambda}$, $\lambda \in \nabla_j$, $j = -1, 0, 1, \dots$, such that

$$\operatorname{supp}\psi_{j,\lambda}, \quad \operatorname{supp}\widetilde{\psi}_{j,\lambda} \subset \Omega\,, \quad \lambda \in \nabla_j\,, \tag{107}$$

$$\langle \widetilde{\psi}_{i,j,k}, \psi_{u,v,\ell} \rangle = \delta_{i,u}\,\delta_{j,v}\,\delta_{k,\ell}\,, \tag{108}$$

and such that $f \in B^s_q(L_p(\Omega))$ if and only if

$$f = \sum_{j=-1}^{\infty} \sum_{\lambda\in\nabla_j} \langle f, \widetilde{\psi}_{j,\lambda}\rangle\, \psi_{j,\lambda} \qquad \text{(convergence in } \mathcal{D}')\,, \tag{109}$$

and

$$\|\, f \,\|^{\clubsuit}_{B^s_q(L_p(\Omega))} \asymp \|\, f \,\|_{B^s_q(L_p(\Omega))}\,. \tag{110}$$

where

$$\|\, f \,\|^{\clubsuit}_{B^s_q(L_p(\Omega))} := \left(\sum_{j=-1}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \left(\sum_{\lambda\in\nabla_j} |\langle f, \widetilde{\psi}_{j,\lambda}\rangle|^p \right)^{q/p} \right)^{1/q} < \infty\,. \tag{111}$$

Acknowledgment. We thank Hans Triebel for many valuable remarks and comments.

References

[1] I. Babuška, W.C. Rheinboldt (1978): A posteriori error estimates for finite element methods. *Int. J. Numer. Math. Engrg.* **12**, 1597–1615.

[2] N.S. Bakhvalov (1971): On the optimality of linear methods for operator approximation in convex classes of functions. *USSR Comput. Math. and Math. Phys.* **11**, 244–249.

[3] R.E. Bank, A. Weiser (1985): Some a posteriori error estimators for elliptic partial differential equations. *Math. Comput.* **44**, 283–301.

[4] R. Becker, C. Johnson, R. Rannacher (1995): Adaptive error control for multigrid finite element methods. *Computing* **55** , 271–288.

[5] F. Bornemann, B. Erdmann, R. Kornhuber (1996): A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM J. Numer. Anal.* **33** (1996), 1188–1204.

[6] G. Bourdaud (1995): Ondelletes et des espaces de Besov. *Revista Mat. Iberoam.* **11**, 477–512.

[7] A.M. Caetano (1998): About approximation numbers in function spaces. *J. Approx. Theory* **94**, 383–395.

[8] C. Canuto, A. Tabacco, K. Urban (1999): The wavelet element method, Part I: construction and analysis. *Appl. Comp. Harm. Anal.* **6**, 1–52.

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

[10] B. Carl, I. Stephani (1990): *Entropy, Compactness and the Approximation of Operators*. Cambridge Univ. Press, Cambridge.

[11] A. Cohen (2003): *Numerical Analysis of Wavelet Methods*. Elsevier Science, Amsterdam.

[12] A. Cohen, W. Dahmen, R. DeVore (2001): Adaptive wavelet methods for elliptic operator equations – convergence rates. *Math. Comp.* **70** (2001), 22–75.

[13] A. Cohen, W. Dahmen, R. DeVore (2002): Adaptive wavelet methods II – beyond the elliptic case. *Found. Comput. Math.* **2**, 203–245.

[14] A. Cohen, W. Dahmen, R. DeVore (2003): Adaptive methods for nonlinear variational problems. *SIAM J. Numer. Anal.* **41(5)**, 1785–1823.

[15] S. Dahlke (1999): Besov regularity for elliptic boundary value problems in polygonal domains. *Appl. Math. Lett.* **12(6)**, 31–38.

[16] S. Dahlke (1998): Besov regularity for elliptic boundary value problems with variable coefficients. *Manuscripta Math.* **95**, 59–77.

[17] S. Dahlke (1999): Besov regularity for interface problems. *Z. Angew. Math. Mech.* **79**, 383–388.

[18] S. Dahlke (1999): Besov regularity for the Stokes problem. In: Advances in Multivariate Approximation, (W. Haussmann, K. Jetter, M. Reimer, Eds.), Wiley VCH, Mathematical Research 107, Berlin, 129–138.

[19] S. Dahlke, W. Dahmen, R. DeVore (1997): Nonlinear approximation and adaptive techniques for solving elliptic operator equations, in: Multicale Wavelet Methods for Partial Differential Equations, (W. Dahmen, A. Kurdila, P. Oswald, Eds.), Academic Press, San Diego, 237–283.

[20] S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider (1997): Stable multiscale bases and local error estimation for elliptic problems. *Appl. Numer. Math.* **23**, 21–48.

[21] S. Dahlke, W. Dahmen, K. Urban (2002): Adaptive wavelet methods for saddle point problems – optimal convergence rates, *SIAM J. Numer. Anal.* **40(4)**, 1230–1262.

[22] S. Dahlke, R. DeVore (1997): Besov regularity for elliptic boundary value problems. *Comm. Partial Differential Equations* **22(1&2)**, 1–16.

[23] S. Dahlke, E. Novak, W. Sickel (2004): Optimal approximation of elliptic problems by linear and nonlinear mappings I. Preprint, Marburg, Jena, 18 pp, to appear in: *J. Complexity.*

[24] W. Dahmen, R. Schneider (1998): Wavelets with complementary boundary conditions - function spaces on the cube. *Results in Math.* **34**, 255–293.

[25] W. Dahmen, R. Schneider (1999): Composite wavelet bases for operator equations. *Math. Comp.* **68**, 1533–1567.

[26] W. Dahmen, R. Schneider (1999): Wavelets on manifolds I: Construction and domain decomposition. *SIAM J. Math. Anal.* **31**, 184–230.

[27] K. Deimling (1985): *Nonlinear Functional Analysis.* Springer-Verlag, Berlin.

[28] R.A. DeVore (1998): Nonlinear Approximation. *Acta Numerica* **7**, 51–150.

[29] R.A. DeVore, R. Howard, C. Micchelli (1989): Optimal nonlinear approximation. *Manuscripta Math.* **63**, 469–478.

[30] R.A. DeVore, G. Kyriazis, D. Leviatan, V.M. Tikhomirov (1993): Wavelet compression and nonlinear n-widths. *Adv. Comput. Math.* **1**, 197–214.

[31] R.A. DeVore, B. Jawerth, V. Popov (1992): Compression of wavelet decompositions. *Amer. J. Math.* **114**, 737–785.

[32] R.A. DeVore, G. Petrova, V. Temlyakov (2003): Best basis selection for approximation in L_p. *Found. Comput. Math.* **3**, 161-185.

[33] R.A. DeVore, V. Popov (1988): Interpolation spaces and nonlinear approximation. In: Function spaces and approximation, Lect. notes in Math. **1302**, 191–205.

[34] R.A. DeVore, R.C. Sharpley (1993): Besov spaces on domains in $\mathbb{R}^d$. *Trans. Amer. Math. Soc.* **335**, 843–864.

[35] R.A. DeVore, V.N. Temlyakov (1995): Nonlinear approximation by trigonometric sums. *J. Fourier Anal. Appl.* **2**, 29–48.

[36] S. Dispa (2002): Intrinsic characterizations of Besov spaces on Lipschitz domains. *Math. Nachr.* **260**, 21-33.

[37] D. Donoho (1993): Unconditional bases are optimal for data compression and for statistical etsimation. Appl. Comput. Harmon. Anal. **1**, 100-115.

[38] W. Dörfler (1996): A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.* **33**, 737–785.

[39] D. Dung, V.Q. Thanh (1996): On nonlinear n-widths. *Proc. of the AMS* **124**, 2757–2765.

[40] D. Dung (2000): Continuous algorithms in n-term approximation and nonlinear widths. *J. Approx. Th.* **102**, 217–242.

[41] E.G. D'yakonov (1996): *Optimization in Solving Elliptic Problems.* CRC Press, Boca Raton.

[42] D.E. Edmunds, H. Triebel (1996): *Function Spaces, Entropy Numbers, Differential Operators.* Cambridge University Press, Cambridge.

[43] S. Gal, C.A. Micchelli (1980): Optimal sequential and non-sequential procedures for evaluating a functional. *Appl. Anal.* **10**, 105–120.

[44] E.D. Gluskin (1981): On some finite dimensional problems of the theory of diameters. *Vestnik Leningr. Univ.* **13**, 5–10.

[45] E.D. Gluskin (1983): Norms of random matrices and diameters of finite dimensional sets. *Mat. Sb.* **120**, 180–189.

[46] P. Grisvard (1975): Behavior of solutions of elliptic boundary value problems in a polygonal or polyhedral domain. In: Symposium on Numerical Solutions of Partial Differential Equations III, (B. Hubbard, Ed.), Academic Press, New York 207–274.

[47] P. Grisvard (1985): *Elliptic Problems in Nonsmooth Domains.* Pitman, Boston.

[48] P. Grisvard (1992): *Singularites in Boundary Value Problems.* Research Notes in Applied Mathematics **22**, Springer, Berlin.

[49] W. Hackbusch (1992): *Elliptic Differential Equations: Theory and Numerical Treatment.* Springer, Berlin.

[50] D. Jerison, C.E. Kenig (1995): The inhomogeneous Dirichlet problem in Lipschitz domains. *J. Funct. Anal.* **130**, 161–219.

[51] B. Kashin (1985): Approximation properties of complete orthonormal systems. Trudy Mat. Inst. Steklov **172**, 187–191.

[52] R.B. Kellogg, M. Stynes (1999): n-widths and singularly perturbed boundary value problems. *SIAM J. Numer. Anal.* **36**, 1604–1620.

[53] T. Kühn, H.-G. Leopold, W. Sickel, L. Skrzypczak (2003): Entropy numbers of embeddings of weighted Besov spaces. II. Proc. Edinburgh Math. Soc. (to appear).

[54] H.-G. Leopold (1999): Embeddings for general weighted sequence spaces and entropy numbers. In: Function spaces, differential operators, and nonlinear analysis, Academy of Sciences of the Czech Republic, Praha 2000, 170–186.

[55] R. Linde (1985): s-numbers of diagonal operators and Besov embeddings In: Proc. 13.th Winter School on Abstract Analysis, Rend. Circ. Mat. Palermo, II. Ser. Suppl. **10**, 83–110.

[56] J.L. Lions, E. Magenes (1972): *Non-Homogeneous Boundary Value Problems and Applications I.* Springer, Berlin.

[57] P. Mathé (1990): s-Numbers in information-based complexity. *J. Complexity* **6**, 41–66.

[58] J.M. Melenk (2000): On n-widths for elliptic problems. *J. Math. Anal. Appl.* **247**, 272–289.

[59] Y. Meyer (1992): *Wavelets and Operators.* Cambridge Univ. Press.

[60] S.M. Nikol'skij (1975): *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer, Berlin.

[61] E. Novak (1995): The real number model in numerical analysis. *J. Complexity* **11**, 57–73.

[62] E. Novak (1996): On the power of adaption. *J. Complexity* **12**, 199–237.

[63] E. Novak, H. Woźniakowski (1999): On the cost of uniform and nonuniform algorithms. *Theor. Comp. Sci.* **219**, 301–318.

[64] E. Novak, H. Woźniakowski (2000): Complexity of linear problems with a fixed output basis. *J. Complexity* **16**, 333–362.

[65] P. Oswald (1990): On the degree of nonlinear spline approximation in Besov-Sobolev spaces. *J. Approximation Theory* **61**, 131-157.

[66] J. Peetre (1976): *New Thoughts on Besov Spaces.* Duke Univ. Math. Series, Durham.

[67] S.V. Pereverzev (1996): *Optimization of Methods for Approximate Solution of Operator Equations.* Nova Science Publishers, New York.

[68] A. Pietsch (1974): s-numbers of operators in Banach spaces. *Studia Math.* **51**, 201–223.

[69] A. Pietsch (1987): *Eigenvalues and s-Numbers.* Geest und Portig, Leipzig.

[70] A. Pinkus (1985): *n-Widths in Approximation Theory.* Springer-Verlag, Berlin.

[71] T. Runst, W. Sickel (1996): *Sobolev Spaces of Fractional Order, Nemytskij Operators and Nonlinear Partial Differential Equations.* de Gruyter, Berlin.

[72] V.S. Rychkov (1999): On restrictions and extensions of the Besov and Triebel-Lizorkin spaces with respect to Lipschitz domains. *J. London Math. Soc.* **60**, 237–257.

[73] A. Seeger (1989): A note on Triebel-Lizorkin spaces. *Banach Center Publ.* **22**, 391–400.

[74] E.M. Stein (1970): *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press, Princeton.

[75] M.I. Stesin (1974): Aleksandrov diameters of finite-dimensional sets and of classes of smooth functions. *Dokl. Akad. Nauk SSSR* **220**, 1278–1281.

[76] V.N. Temlyakov (1993): Approximation of periodic functions. Nova Science, New York.

[77] V.N. Temlyakov (2003): Nonlinear methods of approximation. *Found. Comput. Math.* **3**, 33-107.

[78] V.M. Tikhomirov (1990): *Approximation Theory.* In Encyclopaedia of Math. Sciences **14**, Analysis II, Springer, Berlin.

[79] J.F. Traub, H. Woźniakowski (1980): *A General Theory of Optimal Algorithms.* Academic Press.

[80] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski (1988): *Information–Based Complexity.* Academic Press.

[81] H. Triebel (1978): *Interpolation Theory, Function Spaces, Differential Operators.* VEB Deutscher Verlag der Wissenschaften.

[82] H. Triebel (1983): *Theory of Function Spaces.* Birkhäuser, Basel.

[83] H. Triebel (1992): *Theory of Function Spaces. II.* Birkhäuser, Basel.

[84] H. Triebel (2002): Function spaces in Lipschitz domains and on Lipschitz manifolds. Characteristic functions as pointwise multipliers. *Revista Matemática Complutense* **15**, 475–524.

[85] H. Triebel (2004): A note on wavelet bases in function spaces. Proc. Orlicz Centenary Conf. Function Spaces 7. Banach Center Publ. **64**, 193–206, Polish Acad. Sci.

[86] R. Verfürth (1994): A posteriori error estimation and adaptive mesh–refinement techniques. *J. Comp. Appl. Math.* **50**, 67–83.

[87] A.G. Werschulz (1996): *The Computational Complexity of Differential and Integral Equations.* Oxford Science Publications.

Stephan Dahlke
Philipps-Universität Marburg
FB12 Mathematik und Informatik
Hans-Meerwein Straße
Lahnberge
35032 Marburg
Germany
e–mail: dahlke@mathematik.uni-marburg.de
WWW: http://www.mathematik.uni-marburg.de/∼dahlke/

Erich Novak, Winfried Sickel
Friedrich-Schiller-Universität Jena
Mathematisches Institut
Ernst-Abbe-Platz 2
07743 Jena
Germany
e-mail: {novak, sickel}@math.uni-jena.de
WWW: http://www.minet.uni-jena.de/∼{novak,sickel}