Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Complexity 22 (2006) 29–49

Journal of COMPLEXITY

www.elsevier.com/locate/jco

Optimal approximation of elliptic problems by linear and nonlinear mappings I

Stephan Dahlke[a,1], Erich Novak[b,*], Winfried Sickel[b]

[a] *Philipps-Universität Marburg, FB12 Mathematik und Informatik, Hans-Meerwein Straße, Lahnberge, 35032 Marburg, Germany*

[b] *Friedrich-Schiller-Universität Jena, Mathematisches Institut, Ernst-Abbe-Platz 2, 07743 Jena, Germany*

Received 24 October 2004; accepted 24 June 2005
Available online 8 September 2005

Abstract

We study the optimal approximation of the solution of an operator equation $\mathscr{A}(u) = f$ by linear mappings of rank n and compare this with the best n-term approximation with respect to an optimal Riesz basis. We consider worst case errors, where f is an element of the unit ball of a Hilbert space. We apply our results to boundary value problems for elliptic PDEs that are given by an isomorphism $\mathscr{A} : H_0^s(\Omega) \to H^{-s}(\Omega)$, where $s > 0$ and Ω is an arbitrary bounded Lipschitz domain in $\mathbb{R}^d$. We prove that approximation by linear mappings is as good as the best n-term approximation with respect to an optimal Riesz basis. We discuss why nonlinear approximation still is important for the approximation of elliptic problems.

MSC: 41A25; 41A46; 41A65; 42C40; 65C99

Keywords: Elliptic operator equations; Worst case error; Linear and nonlinear approximation methods; Best n-term approximation; Bernstein widths; Manifold widths

* Corresponding author.

E-mail addresses: dahlke@mathematik.uni-marburg.de (S. Dahlke), novak@math.uni-jena.de (E. Novak), sickel@math.uni-jena.de (W. Sickel)

URLs: http://www.mathematik.uni-marburg.de/~dahlke/ (S. Dahlke), http://www.minet.uni-jena.de/~novak/ (E. Novak), http://www.minet.uni-jena.de/~sickel/ (W. Sickel).

[1] 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.

0885-064X/$ - see front matter
doi:10.1016/j.jco.2005.06.005

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, for example, 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. For this situation we take $H = H_0^s(\Omega)$ 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 $\mathcal{A}(u) = f$.

Let F be a specified normed (or quasi-normed) subspace of G. We use linear and nonlinear mappings S_n for approximating the solution $u = \mathcal{A}^{-1}(f)$ for $f \in F$. Let us consider the worst case error

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

For a given basis $\mathcal{B} = \{h_i \mid i \in \mathbb{N}\}$ 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 the basis $\mathcal{B}$ to be chosen in a nearly arbitrary way. Then the nonlinear widths $e_{n,C}^{\mathrm{non}}(S, F, H)$ are given by

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

Here $\mathcal{B}_C$ denotes a set of Riesz bases for H, where C indicates the stability of the basis, see Section 2.1 for details. These numbers are the main topic of our analysis. We compare nonlinear approximations with linear approximations. Here we consider the class $\mathcal{L}_n$ of all continuous linear mappings $S_n : F \to H$,

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

with arbitrary $\tilde{h}_i \in H$. The worst case error of optimal linear mappings is given by

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

The third class of approximation methods that we study in this paper is the class of 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

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

where $S_n = \varphi_n \circ N_n$. These numbers, or slightly different numbers, were studied by different authors, cf. [7,8,10,24]. Sometimes the e_n^{cont} are called manifold widths of S, see [8].

Remark 1. (i) A purpose of this paper is to compare the numbers $e_{n,C}^{\mathrm{non}}(S, F, H)$ with the numbers $e_n^{\mathrm{lin}}(S, F, H)$, where $S : F \to H$ is the restriction of $\mathcal{A}^{-1} : G \to H$ to $F \subset G$. In this sense we compare optimal linear approximation of S (i.e., by linear mappings of rank n) with the best n-term approximation with respect to an optimal Riesz basis.

(ii) To avoid possible misunderstandings, it is important to clarify the following point. In the realm of approximation theory, very often the term "linear approximation" is used for an approximation scheme that comes from a sequence of linear spaces that are uniformly refined, see, e.g., [6] for a detailed discussion. However, in our definition of $e_n^{\mathrm{lin}}(S, F, H)$ we allow arbitrary linear S_n, not only those that are based on uniformly refined subspaces. In this paper, the latter will be denoted by uniform approximation scheme.

For reader's convenience, we finish this section by briefly summarizing the main results of this paper.

Theorem 1. *Assume that $F \subset G$ is quasi-normed. Then*

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

holds for all $m \geqslant 4\,C^2\,n$, where $b_n(S, F, H)$ denotes the nth Bernstein width of the operator S, see Section 2.2 for details.

Theorem 2 and Corollary 1. *Assume that $F \subset G$ is a Hilbert space and*

$$b_{2n}(S, F, H) \asymp b_n(S, F, H).$$

Then

$$e_n^{\mathrm{lin}}(S, F, H) = e_n^{\mathrm{cont}}(S, F, H) \asymp e_{n,C}^{\mathrm{non}}(S, F, H).$$

In this sense, approximation by linear mappings is as good as approximation by nonlinear mappings. In this paper, '$a \asymp b$' always means that both quantities can be uniformly bounded by a constant multiple of each other. Likewise, '$\lesssim$' indicates inequality up to constant factors.

Theorem 4. *Assume that $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism, with no further assumptions. Then we have for all $C \geqslant 1$*

$$e_n^{\mathrm{lin}}(S, H^{-s+t}, H^s) \asymp e_{n,C}^{\mathrm{non}}(S, H^{-s+t}, H^s) \asymp n^{-t/d}.$$

In this sense, approximation by linear mappings is as good as approximation by nonlinear mappings.

Theorem 5. *If we allow only function evaluations instead of general linear information, then the order of convergence drops down from $n^{-t/d}$ to $n^{(s-t)/d}$, where $t > s + d/2$.*

- In Theorems 6 and 7 we study the Poisson equation and the best n-term wavelet approximation. Theorem 6 shows that best n-term wavelet approximation might be suboptimal in general. Theorem 7, however, shows that for a polygonal domain in $\mathbb{R}^2$ best n-term wavelet approximation is almost optimal.

Some of these results (Corollary 1, Theorem 4) might be surprising since there is a widespread believe that nonlinear approximation is better than approximation by linear operators. Therefore we want to make the following remarks concerning our setting:

- We allow arbitrary linear operators S_n with rank n, not only those that are based on a uniform refinement.
- We consider the worst case error with respect to the unit ball of a Hilbert space.
- Our results are concerned with approximations, not with their numerical realization. For instance, the construction of an optimal linear method might require the precomputation of a suitable basis (depending on $\mathcal{A}$), which is usually a prohibitive task. See also Remark 10, where we discuss in more detail why nonlinear approximation is very important for the approximation of elliptic problems.
- In another paper (in progress) we continue this work under the assumption that F is a general Besov space. Then it turns out that for some parameters nonlinear approximation is essentially better than linear approximation.

2. Basic concepts of optimality

2.1. Classes of admissible mappings

2.1.1. Nonlinear mappings S_n

We will study certain approximations of S based on Riesz bases, cf., e.g., Meyer [26, p. 21].

Definition 1. Let H be a Hilbert space. Then a sequence $h_1, h_2, \ldots$ 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 $\alpha_i \neq 0$ for only finitely many i, we have

$$A \left(\sum_k |\alpha_k|^2 \right)^{1/2} \leqslant \left\| \sum_k \alpha_k h_k \right\|_H \leqslant B \left(\sum_k |\alpha_k|^2 \right)^{1/2} \tag{4}$$

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

Remark 2. 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 [1, Sections 2.6, 2.12].

In what follows

$$\mathcal{B} = \{h_i \mid i \in \mathbb{N}\} \tag{5}$$

will always denote a Riesz basis of H and A and B will be the corresponding optimal constants in (4). We study optimal approximations S_n of $S = \mathcal{A}^{-1}$ of the form

$$S_n(f) = u_n = \sum_{k=1}^{n} c_k\, h_{i_k}, \tag{6}$$

where $f = \mathcal{A}(u)$. Assuming that we can choose $\mathcal{B}$, we want to choose an optimal basis $\mathcal{B}$. What is the error of such an approximation S_n and in which sense can we say that $\mathcal{B}$ and S_n are optimal?

It is important to note that optimality of S_n does not make sense for a single u: we simply can take a $\mathcal{B}$ where h_1 is a multiple of u, and hence we can write the exact solution u as $u_1 = c_1 h_1$, i.e., with $n = 1$. To define optimality of an approximation S_n we need a suitable subset of G. We consider the worst case error

$$e(S_n, F, H) := \sup_{\|f\|_F \leqslant 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H, \tag{7}$$

where F is a normed (or quasi-normed) space, $F \subset G$. For a given basis $\mathcal{B}$ we consider the class $\mathcal{N}_n(\mathcal{B})$ of all (linear or nonlinear) mappings of the form

$$S_n(f) = \sum_{k=1}^{n} c_k\, h_{i_k}, \tag{8}$$

where the c_k and the i_k depend in an arbitrary way on f. Optimality is expressed by the quantity

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

This reflects the best n-term approximation of $\mathcal{A}^{-1}(f)$. This subject is widely studied, see the surveys [6,37]. Since S_n is arbitrary, one immediately obtains

$$\begin{aligned}\inf_{S_n \in \mathcal{N}_n(\mathcal{B})} \sup_{\|f\|_F \leqslant 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H &= \sup_{\|f\|_F \leqslant 1} \inf_{S_n \in \mathcal{N}_n(\mathcal{B})} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H \\ &= \sup_{\|f\|_F \leqslant 1} \sigma_n(\mathcal{A}^{-1} f, \mathcal{B})_H .\end{aligned}$$

We allow the basis $\mathcal{B}$ to be chosen in a nearly arbitrary way. It is natural to assume some common stability of the bases under consideration. For a real number $C \geqslant 1$ we define

$$\mathcal{B}_C := \left\{\mathcal{B} :\ B/A \leqslant C\right\}. \tag{9}$$

We define the nonlinear widths $e^{\rm non}_{n,C}(S, F, H)$ as

$$e^{\rm non}_{n,C}(S, F, H) = \inf_{\mathcal{B}\in\mathcal{B}_C} \inf_{S_n\in\mathcal{N}_n(\mathcal{B})} e(S_n, F, H). \tag{10}$$

These numbers are the main topic of our analysis. They could be called the *errors of the best n-term approximation* (with respect to the collection $\mathcal{B}_C$ of Riesz basis of H), for brevity we call them *nonlinear widths*. In this paper we investigate the numbers $e^{\rm non}_{n,C}(S, F, H)$ only when H is a Hilbert space. More general concepts are introduced and investigated in [37].

Remark 3. 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 computational cost. Since we are interested in lower bounds, our results are strengthened by considering such a large class of approximations.

Remark 4. It is obvious from the definition (10) that $S^*_n \in \mathcal{N}_n(\mathcal{B})$ can be optimal for a given basis $\mathcal{B}$ in the sense that

$$e(S^*_n, F, H) \approx \inf_{S_n\in\mathcal{N}_n(\mathcal{B})} e(S_n, F, H),$$

although the number $e^{\rm non}_{n,C}(S, F, H)$ is much smaller, since the given $\mathcal{B}$ is far from being optimal. See also Remark 10.

2.1.2. Linear mappings S_n

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

$$S_n(f) = \sum_{i=1}^{n} L_i(f) \cdot \tilde{h}_i \tag{11}$$

with arbitrary $\tilde{h}_i \in H$. For each S_n we define $e(S_n, F, H)$ by (7) and hence we can define the worst case error of optimal linear mappings by

$$e^{\rm lin}_n(S, F, H) = \inf_{S_n\in\mathcal{L}_n} e(S_n, F, H). \tag{12}$$

The numbers $e^{\rm lin}_n(S, F, H)$ (or slightly different numbers) are usually called *approximation numbers* or *linear widths* of $S : F \to H$, cf. [24,33,34,38].

If F is a space of functions on a set Ω such that function evaluation $f \mapsto f(x)$ is continuous, then one can define the *linear sampling numbers*

$$g^{\rm lin}_n(S, F, H) = \inf_{S_n\in\mathcal{L}^{\rm std}_n} e(S_n, F, H), \tag{13}$$

where $\mathcal{L}^{\rm std}_n \subset \mathcal{L}_n$ contains only those S_n that are of the form

$$S_n(f) = \sum_{i=1}^{n} f(x_i) \cdot \tilde{h}_i \tag{14}$$

with $x_i \in \Omega$. For the numbers $g^{\rm lin}_n$ we only allow *standard information*, i.e., function values of the right-hand side. The inequality $g^{\rm lin}_n(S, F, H) \geqslant e^{\rm lin}_n(S, F, H)$ is trivial. One also might allow nonlinear $S_n = \varphi_n \circ N_n$ with (linear) standard information $N_n(f) = (f(x_1), \ldots, f(x_n))$ and arbitrary $\varphi_n : \mathbb{R}^n \to H$. This leads to the *sampling numbers* $g_n(S, F, H)$.

2.1.3. 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

$$e_n^{\mathrm{cont}}(S, F, H) = \inf_{S_n \in \mathcal{C}_n} e(S_n, F, H). \tag{15}$$

These numbers, or variants of same, were studied by different authors, cf. [7,8,10,24]. Sometimes these numbers are called manifold widths of *S*, see [8]. The inequalities

$$e_{n,C}^{\mathrm{non}}(S, F, H) \leqslant e_n^{\mathrm{lin}}(S, F, H) \tag{16}$$

and

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

are, of course, trivial.

2.2. Relations to Bernstein widths

The following quantities are useful for the understanding of e_n^{cont} and e_n^{non}:

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 $S(\{\|f\|_F \leqslant 1\})$.

Remark 5. In the literature there are used different definitions of Bernstein widths. E.g. in Pietsch [32] 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 $b_n(S, F, H) = \widetilde{b}_{n+1}(S, F, H)$.

As it is well-known, Bernstein widths are useful for the proof of lower bounds, see [7,10,34]. The next lemma is certainly known. Since we could not find a reference, we include it with a proof.

Lemma 1. *Let $n \in \mathbb{N}$ and assume that $F \subset G$ is quasi-normed. Then the inequality*

$$b_n(S, F, H) \leqslant e_n^{\mathrm{cont}}(S, F, H) \tag{18}$$

holds for all n.

Proof. We assume that $S(\{\|f\|_F \leqslant 1\})$ contains an $(n+1)$-dimensional ball $B \subset H$ of radius r. We may assume that the center is in the origin. Let $N_n : F \to \mathbb{R}^n$ be continuous. Since $S^{-1}(B)$ is an $(n+1)$-dimensional bounded and symmetric neighborhood of 0, it follows from the Borsuk Antipodality Theorem, see [5, par. 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 $\|f\|_F = 1$. Since $\|S(f) - S(-f)\| = 2r$ and $S_n(f) = S_n(-f)$, we find that the maximal error of S_n on $\{\pm f\}$ is at least r. This proves

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

We will see that the b_n can also be used to prove lower bounds for the $e_{n,C}^{\text{non}}$. As usual, c_0 denotes the Banach space of all sequences $x = (x_j)_{j=1}^{\infty}$ of real numbers such that $\lim_{j\to\infty} x_j = 0$ and equipped with the norm of ℓ_∞.

Lemma 2 below has a long history since it is central in the theory of s-numbers. See [33, Lemma 2.9.6], where also its use for proving a result as Lemma 3 is exhibited.

Lemma 2. *Let V denote an n-dimensional subspace of c_0. Then there exists an element $x \in V$ such that $\|x\|_\infty = 1$ and at least n coordinates of $x = (x_1, x_2, \ldots)$ have absolute value* 1.

Lemma 3. *Let V_n be an n-dimensional subspace of the Hilbert space H. Let $\mathcal{B}$ be a Riesz basis with Riesz constants $0 < A \leqslant B < \infty$. Then there is a nontrivial element $x \in V_n$ such that $x = \sum_{j=1}^{\infty} x_j h_j$ and*

$$A\sqrt{n}\, \|(x_j)_j\|_\infty \leqslant \|x\|_H.$$

Proof. Associated with any $x \in H$ there is a sequence $(x_j)_j$ of coefficients with respect to $\mathcal{B}$ that belongs to c_0. In the same way, we associate with $V_n \subset H$ a subspace $X_n \subset c_0$, also of dimension n. As a consequence of Lemma 2, we find an element $(x_j)_j \in X_n$ such that

$$0 < |x_{j_1}| = \cdots = |x_{j_n}| = \|(x_j)_j\|_\infty < \infty.$$

This implies

$$\|x\|_H \geqslant A \left(\sum_{l=1}^{n} |x_{j_l}|^2 \right)^{1/2} = A\sqrt{n}\, \|(x_j)_j\|_\infty. \qquad \square$$

Theorem 1. *Assume that $F \subset G$ is quasi-normed. Then*

$$e_{n,C}^{\text{non}}(S, F, H) \geqslant \frac{1}{2C}\, b_m(S, F, H) \tag{19}$$

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

Proof. Let $\mathcal{B}$ be a Riesz basis with Riesz constants A and B and let $m > n$. Assume that $S(\{\|f\| \leqslant 1\})$ contains an m-dimensional ball with radius ε. Using Lemma 3, there exists an $x \in S(\{\|f\| \leqslant 1\})$ such that $x = \sum_i x_i h_i$, $\|x\| = \varepsilon$ and $|x_i| \leqslant A^{-1} m^{-1/2} \varepsilon$ for all i. Let $x_1, \ldots x_n$ be the n largest components (with respect to the absolute value) of x. Now, consider $y = \sum_i y_i h_i$ such that at most n coefficients are nonvanishing. Then

$$\|x - y\|_H \geqslant A \|(x_i - y_i)_i\|_2$$

and the optimal choice of y (with respect to the right-hand side) is given by y^0, where $y_1^0 = x_1, \ldots, y_n^0 = x_n$. Now we continue our estimate

$$\begin{aligned} A\|(x_i - y_i)_i\|_2 &\geqslant A\,(\|(x_i)_i\|_2 - \|(y_i)_i\|_2) \\ &\geqslant A\left(\frac{\varepsilon}{B} - \frac{1}{A}\,\varepsilon\sqrt{\frac{n}{m}}\right) = \varepsilon\left(\frac{A}{B} - \sqrt{\frac{n}{m}}\right). \end{aligned} \tag{20}$$

The right-hand side is at least $\varepsilon\, A/(2B)$ if $m \geqslant 4B^2 n/A^2$. □

Remark 6. Probably the constant $1/(2C)$ is not optimal. But it is obvious from (20) that for m tending to infinity the constant is approaching A/B.

2.3. The case of a Hilbert space

Now let us assume, in addition to the assumptions of the previous subsections, that $F \subset G$ is a Hilbert space. The following result is well known, see [32] and Remark 5.

Theorem 2. *Assume that F is a Hilbert space. Then*

$$e_n^{\mathrm{lin}}(S, F, H) = e_n^{\mathrm{cont}}(S, F, H) = b_n(S, F, H). \tag{21}$$

In many applications one studies problems with "finite smoothness" and then, as a rule, one has the estimate

$$b_{2n}(S, F, H) \asymp b_n(S, F, H). \tag{22}$$

Formula (22) especially holds for the operator equations that we study in Section 3. Then we conclude that approximation by optimal linear mappings yields the same order of convergence as the best n-term approximation.

Corollary 1. *Assume that $S : F \to H$ with Hilbert spaces F and H, with* (22) *holding. Then*

$$e_n^{\mathrm{lin}}(S, F, H) = e_n^{\mathrm{cont}}(S, F, H) \asymp e_{n,C}^{\mathrm{non}}(S, F, H). \tag{23}$$

3. Elliptic problems

In this section, we study the more special case where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and $\mathcal{A} = S^{-1} : H_0^s(\Omega) \to H^{-s}(\Omega)$ is an isomorphism, where $s > 0$. The first step is to recall the definition of the smoothness spaces that are needed for our analysis.

3.1. Function spaces

If m is a natural number, we let $H^m(\Omega)$ denote the set of all functions $u \in L_2(\Omega)$ such that the (distributional) derivatives $D^\alpha u$ of order $|\alpha| \leqslant m$ also belong to $L_2(\Omega)$. This set, equipped with the norm

$$\| u \|_{H^m(\Omega)} := \sum_{|\alpha| \leqslant m} \| D^\alpha u \|_{L_2(\Omega)},$$

becomes a Hilbert space. For a positive noninteger s, we define $H^s(\Omega)$ as specific Besov spaces. If $h \in \mathbb{R}^d$, we let Ω_h denote the set of all $x \in \Omega$ such that the line segment $[x, x+h]$ is contained in Ω. The modulus of smoothness $\omega_r(u,t)_{L_p(\Omega)}$ of a function $u \in L_p(\Omega)$, where $0 < p \leqslant \infty$, is defined by

$$\omega_r(u,t)_{L_p(\Omega)} := \sup_{|h| \leqslant t} \|\Delta_h^r(u, \cdot)\|_{L_p(\Omega_{rh})}, \quad t > 0,$$

with Δ_h^r the rth difference with step h. For $s > 0$ and $0 < q, p \leqslant \infty$, the *Besov space* $B_q^s(L_p(\Omega))$ is defined as the space of all functions $u \in L_p(\Omega)$ for which

$$|u|_{B_q^s(L_p(\Omega))} := \begin{cases} \left(\int_0^\infty [t^{-s} \omega_r(u,t)_{L_p(\Omega)}]^q \, dt/t \right)^{1/q}, & 0 < q < \infty, \\ \sup_{t \geqslant 0} t^{-s} \omega_r(u,t)_{L_p(\Omega)}, & q = \infty, \end{cases} \tag{24}$$

is finite with $r \in \mathbb{N}$, $s < r \leqslant s+1$, see, e.g., [40] for details. It turns out that (24) is a (quasi-)semi-norm for $B_q^s(L_p(\Omega))$. If we add $\|u\|_{L_p(\Omega)}$ to (24), we obtain a (quasi-)norm for $B_q^s(L_p(\Omega))$. Then, for positive noninteger s, we define

$$H^s(\Omega) := B_2^s(L_2(\Omega)).$$

It is known that this definition coincides up to equivalent norms with other definitions based, e.g., on complex or real interpolation, cf. Dispa [9], Lions and Magenes [23, Volume 1] and Triebel [41].

For all $s > 0$ we let $H_0^s(\Omega)$ denote the closure of the test functions $\mathcal{D}(\Omega)$ in $H^s(\Omega)$. Finally, we put

$$H^{-s}(\Omega) := (H_0^s(\Omega))', \quad s > 0 \;\; s \neq \tfrac{1}{2} + k,$$

where $k \in \mathbb{N}_0$. Alternatively (and this is done e.g. in [19] and will play a role in Subsection 3.5) one could use the following approach: define for $s > 0$

$$\widetilde{H}^s(\Omega) := \left\{ u \in L_2(\Omega) : \text{ there exists } g \in H^s(\mathbb{R}^d) \text{ with } g_{|\Omega} = u \text{ and } \operatorname{supp} g \subset \overline{\Omega} \right\}$$

equipped with the induced norm. Then, for all $s > 0$, $s \neq \frac{1}{2} + k$, $k \in \mathbb{N}$

$$H_0^s(\Omega) = \widetilde{H}^s(\Omega),$$

in the sense of equivalent norms, cf. Grisvard [14, Corollary 1.4.4.5]. If $0 < s = \frac{1}{2} + k$, $k \in \mathbb{N}$, then we put

$$H^{-s}(\Omega) := \left(\widetilde{H}^s(\Omega) \right)'. \tag{25}$$

Observe, that by the previous remark this could be used as definition for all values of $s > 0$ (up to equivalent norms).
Since we have Hilbert spaces, linear mappings are (almost) optimal approximations, i.e., Corollary 1 holds. We want to say more about the structure of an optimal linear S_n for the approximation of $S = \mathcal{A}^{-1}$. For this, the notion of a "regular problem" is useful.

3.2. Regular problems

The notion of regularity is very important for the theory and the numerical treatment of operator equations, see [16]. We use the following definition and assume that $t > 0$.

Definition 3. Let $s > 0$. An isomorphism $\mathcal{A} : H_0^s(\Omega) \to H^{-s}(\Omega)$ is H^{s+t}-*regular* if also

$$\mathcal{A} : H_0^s(\Omega) \cap H^{s+t}(\Omega) \to H^{-s+t}(\Omega) \tag{26}$$

is an isomorphism.

A classical example is the Poisson equation in a C^∞-domain: this yields an operator that is H^{1+t}-regular for every $t > 0$. We refer, e.g., to [16] for further information and examples. It is known that in this situation we obtain the optimal rate

$$e_n^{\text{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d} \tag{27}$$

of linear methods. This is a classical result, at least for $t, s \in \mathbb{N}$ and for special domains. We refer to the books [11,31,43] that contain hundreds of references.

We prove that the rate (27) is true for arbitrary $s, t > 0$, and for arbitrary bounded (nonempty, of course) Lipschitz domains. The optimal rate can be obtained by using Galerkin spaces that do not depend on the particular operator $\mathcal{A}$. With nonlinear approximations we cannot obtain a better rate of convergence.

Theorem 3. *Assume that the problem is H^{s+t}-regular. Then for all $C \geqslant 1$, we have*

$$e_n^{\text{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp e_{n,C}^{\text{non}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d} \tag{28}$$

and the optimal order can be obtained by subspaces of H^s that do not depend on the operator $S = \mathcal{A}^{-1}$.

Proof. Consider first the identity (embedding) $I : H^{s+t}(\Omega) \to H^s(\Omega)$. It is known that

$$e_n^{\text{lin}}(I, H^{s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d}.$$

This is a classical result (going back to Kolmogorov (1936), see [22]) for $s, t \in \mathbb{N}$, see also [34]. For the general case ($s, t > 0$ and arbitrary bounded Lipschitz domains) see [12,41]. We obtain the same order for $I : H^{s+t}(\Omega) \cap H_0^s(\Omega) \to H^s(\Omega)$.

We assume (26), and hence $S : H^{-s+t}(\Omega) \to H^{s+t}(\Omega) \cap H_0^s(\Omega)$ is an isomorphism. Hence we obtain the same order of the e_n^{lin} for I and for $I \circ S_{|H^{-s+t}(\Omega)}$. Together with Corollary 1 this proves (28).

Assume that the linear mapping

$$\sum_{i=1}^{n} g_i \, L_i(f)$$

is good for the mapping $I : H^{s+t}(\Omega) \cap H_0^s(\Omega) \to H^s(\Omega)$, i.e., we consider a sequence of such approximations with the optimal rate $n^{-t/d}$. Then the linear mappings

$$\sum_{i=1}^{n} g_i \, L_i(Sf)$$

achieve the optimal rate $n^{-t/d}$ for the mapping $S : H^{-s+t}(\Omega) \to H^{s+t}(\Omega) \subset H^s(\Omega)$. $\square$

Remark 7. The same g_i are good for all $H^{s+t}(\Omega)$-regular problems on $H^{-s+t}(\Omega)$; only the linear functionals, given by $L_i \circ S_{|H^{-s+k}}$, depend on the operator $\mathcal{A}$. For the numerical realization we can use the Galerkin method with the space V_n generated by $g_1, \dots, g_n$. It is known that for V_n one can take spaces that are based on uniform refinement, e.g., constructed by uniform grids or uniform finite elements schemes. Indeed, if we consider a sequence V_n of uniformly refined spaces with dimension n, then, under natural conditions, the following characterization holds:

$$u \in H^{t+s}(\Omega) \Longleftrightarrow \sum_{n=1}^{\infty} [n^{t/d} E_n(u)]^2 \frac{1}{n} < \infty, \quad \text{where } E_n(u) := \inf_{g \in V_n} \|u - g\|_{H^s}, \tag{29}$$

see, e.g, [3,30] and the references therein.

Remark 8. Observe that the assumptions of Theorem 3 are rather restrictive. Formally, we assumed that Ω is an arbitrary bounded Lipschitz domain and that $\mathcal{A}$ is H^{s+t}-regular. In practice, however, problems tend to be regular only if Ω has a smooth boundary.

3.3. Nonregular problems

The next result shows that linear approximations also give the optimal rate $n^{-t/d}$ in the nonregular case. An important difference, however, is the fact that now the Galerkin space must depend on the operator $\mathcal{A}$. Related results can be found in the literature, see [21,25,42]. Again we allow arbitrary s and $t > 0$ and arbitrary bounded Lipschitz domains. We also prove that nonlinear approximation methods do not yield a better rate of convergence.

Theorem 4. *Assume that $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism, with no further assumptions. Here $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Then we have for all $C \geqslant 1$*

$$e_n^{\text{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp e_{n,C}^{\text{non}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d}. \tag{30}$$

Proof. Consider first the identity (or embedding) $I : H^{-s+t}(\Omega) \to H^{-s}(\Omega)$. It is known that

$$e_n^{\mathrm{lin}}(I, H^{-s+t}(\Omega), H^{-s}(\Omega)) \asymp n^{-t/d}.$$

Again this is a classical result, for the general case (with $s, t > 0$ and Ω an arbitrary bounded Lipschitz domain), see [41].

We assume that $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism, so that e_n^{lin} have the same order for I and for $S \circ I$. Together with Theorem 1 and Corollary 1, this proves (30).

Assume that the linear mapping

$$\sum_{i=1}^{n} g_i \, L_i(f)$$

is good for the mapping $I : H^{-s+t} \to H^{-s}$, i.e., we consider a sequence of such approximations with the optimal rate $n^{-t/d}$. Then the linear mappings

$$\sum_{i=1}^{n} S(g_i) \, L_i(f) \tag{31}$$

achieve the optimal rate $n^{-t/d}$ for the mapping $S : H^{-s+t}(\Omega) \to H^s(\Omega)$. $\square$

Remark 9. It is well-known that uniform methods can be quite bad for problems that are not regular. Indeed, the general characterization (29) implies that the approximation order of uniform methods is determined by the Sobolev regularity of the solution u. Therefore, if the problem is nonregular, i.e., if the solution u lacks Sobolev smoothness, then the order of convergence of uniform methods drops down.

Remark 10. For nonregular problems, we use linear combinations of $S(g_i)$. The g_i do not depend on S, but of course the $S(g_i)$ do depend on S. This has important practical consequences: if we want to realize good approximations of the form (31) then we need to know the $S(g_i)$. Observe also that in this case we need good knowledge about the approximation of the embedding $I : H^{-s+t}(\Omega) \to H^{-s}(\Omega)$. For $s > 0$, this embedding is not often studied in numerical analysis.

Hence, we see an important difference between regular and arbitrary operator equations: Yes, the order of optimal linear approximations is the same in both cases and also nonlinear (best n-term) approximations cannot be better. But to construct good linear methods in the general case we have to know or to precompute the $S(g_i)$, which is usually almost impossible in practice or at least much too expensive.

This leads us to the following problem: Can we find a $\mathcal{B} \in \mathcal{B}_C$ (here we think about a wavelet basis, but we do not want to exclude other cases) that depends only on t, s, and Ω such that

$$\inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d} \tag{32}$$

for many different operator equations, given by an isomorphism $S = \mathcal{A}^{-1} : H^{-s}(\Omega) \to H_0^s(\Omega)$?

We certainly cannot expect that a single basis $\mathcal{B}$ is optimal for all reasonable operator equations, but the results in Section 3.5 indicate that wavelet methods seem to have some potential in this direction. In any case it is important to distinguish between "an approximation S_n is optimal with respect to the given basis $\mathcal{B}$" and "S_n is optimal with respect to the optimal basis $\mathcal{B}$". See also [27,29].

3.4. Function values

Now we study the numbers $g_n(S, H^{t-s}(\Omega), H^s(\Omega)) = g_n^{\mathrm{lin}}(S, H^{t-s}(\Omega), H^s(\Omega))$ under similar conditions as we had in Theorem 4. In particular we do not assume that the problem is regular. However we have to assume $t > s + d/2$ so that function values will continuously depend on $f \in H^{t-s}(\Omega)$.

Consider first the embedding $I : H^t(\Omega) \to L_2(\Omega)$, where Ω is a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$. We want to use function values of $f \in H^t(\Omega)$ and hence have to assume that $t > d/2$. It is known that

$$e_n^{\mathrm{lin}}(I, H^t(\Omega), L_2(\Omega)) \asymp g_n^{\mathrm{lin}}(I, H^t(\Omega), L_2(\Omega)) \asymp n^{-t/d}, \tag{33}$$

see [28]. This means that arbitrary linear functionals do not yield a better order of convergence than function values. Observe that we always have $g_n = g_n^{\mathrm{lin}}$, since we consider mappings between Hilbert spaces and hence the linear spline algorithm is always optimal, see [39, 4.5.7].

It is interesting that for $s > 0$ arbitrary linear information is superior to function evaluation. In the theorem that follows, we make no smoothness or regularity assumptions.

Theorem 5. *Assume that $S : H^{-s}(\Omega) \to H_0^s(\Omega)$ is an isomorphism, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Then*

$$g_n(S, H^{-s+t}(\Omega), H^s(\Omega)) = g_n^{\mathrm{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{(s-t)/d}, \tag{34}$$

for $t > s + d/2$.

Proof. As in the proof of Theorem 4, it is enough to prove that

$$g_n(I, H^{-s+t}(\Omega), H^{-s}(\Omega)) \asymp n^{(s-t)/d}. \tag{35}$$

To prove the upper and the lower bound for (35), we use several auxiliary problems and start with the upper bound. It is known from [28] that

$$g_n(I, H^{-s+t}(\Omega), L_2(\Omega)) \asymp n^{(s-t)/d}.$$

provided that $t - s > d/2$. From this we obtain the upper bound

$$g_n(I, H^{-s+t}(\Omega), H^{-s}(\Omega)) \leqslant c \cdot n^{(s-t)/d}$$

by embedding.

For the lower bound we use the bound

$$g_n(I, H^{-s+t}(\Omega), L_1(\Omega)) \asymp n^{(s-t)/d}, \tag{36}$$

again from [28]. The lower bound in (36) is proved by the technique of bump functions: Given $x_1, \dots, x_n \in \Omega$, one can construct a function $f \in H^{-s+t}(\Omega)$ with norm one such that $f(x_1) = \cdots = f(x_n) = 0$ and

$$\|f\|_{L_1} \geqslant c \cdot n^{(s-t)/d}, \tag{37}$$

where $c > 0$ does not depend on the x_i or on n. The same technique can be used to prove lower bounds for integration problems. We consider an integration problem

$$\mathrm{Int}(f) = \int_\Omega f\sigma\, dx, \tag{38}$$

where $\sigma \geqslant 0$ is a smooth (and nonzero) function on Ω with compact support. Then this technique gives: Given $x_1, \dots, x_n \in \Omega$, one can construct a function $f \in H^{-s+t}(\Omega)$ with norm one such that $f(x_1) = \cdots = f(x_n) = 0$ and

$$\mathrm{Int}(f) \geqslant c \cdot n^{(s-t)/d}, \tag{39}$$

where $c > 0$ does not depend on the x_i or on n. Since we assumed that σ is smooth with compact support, we have

$$\|f\|_{H^{-s}} \geqslant c \cdot |\mathrm{Int}(f)|$$

and hence we may replace in (39) $\mathrm{Int} f$ by $\|f\|_{H^{-s}}$, hence

$$g_n(I, H^{-s+t}(\Omega), H^{-s}(\Omega)) \geqslant c \cdot n^{(s-t)/d}. \qquad \square$$

3.5. The Poisson equation

Finally, we discuss our results for the specific case of the Poisson equation

$$\begin{aligned} -\triangle u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{40}$$

on a bounded Lipschitz domain Ω contained in $\mathbb{R}^d$, $d \geqslant 2$. Here, as always in this paper, we understand Lipschitz domain in the sense of Steins notion of domains with minimal smooth boundary, cf. Stein [36, VI.3].

It is well-known that (40) fits into our setting with $s = 1$. Indeed, if we consider the weak formulation of this problem, it can be checked that (40) induces a boundedly invertible operator $\mathcal{A} = \triangle : H_0^1(\Omega) \longrightarrow H^{-1}(\Omega)$, see again [16, Chapter 7.2] for details.

In this section, we shall especially focus on *wavelet* bases $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\}$. The indices $\lambda \in \mathcal{J}$ typically encode several types of information, namely the *scale* (often denoted $|\lambda|$), the spatial location and also the type of the wavelet. Recall that in a classical setting, a tensor product construction yields $2^d - 1$ types of wavelets [26]. For instance, on the real line λ can be identified with (j, k), where $j = |\lambda|$ denotes the dyadic refinement level and $2^{-j}k$ signifies the location of the wavelet. We will not discuss at this point any technical description of the basis Ψ. Instead we assume that the domain Ω under consideration enables us to construct a wavelet basis Ψ with the following properties:

- the wavelets are *local* in the sense that

$$\operatorname{diam}(\operatorname{supp}\psi_\lambda) \asymp 2^{-|\lambda|}, \quad \lambda \in \mathcal{J};$$

- the wavelets satisfy the *cancellation property*

$$|\langle v, \psi_\lambda\rangle| \lesssim 2^{-|\lambda|\widetilde{m}} \|v\|_{H^{\widetilde{m}}(\operatorname{supp}\psi_\lambda)},$$

where $\widetilde{m}$ denotes some suitable parameter, and
- the wavelet basis induces characterizations of Besov spaces of the form

$$\|f\|_{B^s_q(L_p(\Omega))} \asymp \left(\sum_{|\lambda|=j_0}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \left(\sum_{\lambda\in\mathcal{J},|\lambda|=j} |\langle f, \tilde{\psi}_\lambda\rangle|^p \right)^{q/p} \right)^{1/q}, \tag{41}$$

where $s > d\left(\frac{1}{p}-1\right)_+$ and $\tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \mathcal{J}\}$ denotes the *dual basis*

$$\langle \psi_\lambda, \tilde{\psi}_\nu\rangle = \delta_{\lambda,\nu}, \quad \lambda, \nu \in \mathcal{J}.$$

For the applications we have in mind, especially the case $p = q$, $p \leqslant 2$, $1/p \leqslant s/d + 1/2$ is important, see, e.g., Theorem 6 for details.

By exploiting the norm equivalence (41) and using the fact that $B^s_2(L_2(\Omega)) = H^s(\Omega)$, a simple rescaling immediately yields a Riesz basis for H^s. We shall also assume that the Dirichlet boundary conditions can be included, so that a characterization of the form (41) also holds for $H^s_0(\Omega)$. We refer to [1] for a detailed discussion. In this setting, the following theorem holds.

Theorem 6. *Let S denote the solution operator for the problem* (40). *Then, for sufficiently large C, best n-term wavelet approximation S_n yields*

$$e^{\mathrm{non}}_{n,C}(S, H^{t-1}(\Omega), H^1(\Omega)) \leqslant e(S_n, H^{t-1}(\Omega), H^1(\Omega)) \leqslant c \begin{cases} n^{-\frac{t}{d}+\varepsilon} & \text{if } 0 < t \leqslant 1/2, \\ n^{-\frac{(t+1)}{3d}+\varepsilon} & \text{if } \frac{1}{2} < t \leqslant \frac{d+2}{2(d-1)}, \\ n^{-\frac{1}{2(d-1)}+\varepsilon} & \text{if } \frac{d+2}{2(d-1)} < t, \end{cases}$$

where $\varepsilon > 0$ is arbitrary and c does not depend on $n \in \mathbb{N}$.

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

$$\| u - S_n(f) \|_{H^1} \leqslant c\, |u|_{B^\alpha_{\tau*}(L_{\tau*}(\Omega))}\, n^{-(\alpha-1)/d}, \quad \frac{1}{\tau^*} = \frac{(\alpha-1)}{d} + \frac{1}{2}, \tag{42}$$

see, e.g., [3] for details. We therefore have to estimate the Besov norm $B^\alpha_{\tau*}(L_{\tau*}(\Omega))$.

Step 2: Besov regularity of u.

First of all, we estimate the Besov norm of u in the specific scale

$$B_\tau^s(L_\tau(\Omega)) \quad \text{where} \quad \frac{1}{\tau} = \frac{s}{d} + \frac{1}{2}. \tag{43}$$

Regularity estimates in the scale (43) have already been performed in [4]. We write the solution u to (40) as

$$u = \tilde{u} + v,$$

where $\tilde{u}$ solves $-\triangle\tilde{u} = \tilde{f}$ on a smooth domain $\widetilde{\Omega} \supset \Omega$. Here $\tilde{f} = \mathcal{E}(f)$ where $\mathcal{E}$ denotes some suitable extension operator with respect to Ω. Furthermore, v is the solution to the additional homogeneous Dirichlet problem

$$\begin{aligned} \triangle v &= 0 \quad \text{in } \Omega \\ v &= g = -\mathrm{Tr}(\tilde{u}) \quad \text{on } \partial\Omega. \end{aligned} \tag{44}$$

Substep 2.1: Regularity of $\tilde{u}$. Let $t > 0$. Let $\mathcal{E}$ be a bounded linear extension operator from $B_2^{t-1}(L_2(\Omega)) \to B_2^{t-1}(L_2(\mathbb{R}^d))$, cf. [35]. Then, by classical elliptic regularity on smooth domains, cf. e.g. [19, Theorem 0.3], it follows from $\mathcal{E}f \in B_2^{t-1}(L_2(\widetilde{\Omega}))$ that $\tilde{u} \in B_2^{t+1}(L_2(\widetilde{\Omega}))$ and

$$\| \tilde{u} \|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leqslant c_1 \, \|\mathcal{E}\| \, \| f \|_{B_2^{t-1}(L_2(\Omega))} .$$

Known embeddings of Besov spaces yield

$$\| \tilde{u} \|_{B_q^{t+1-\varepsilon}(L_q(\widetilde{\Omega}))} \leqslant c_2 \| \tilde{u} \|_{B_2^{t+1}(L_q(\widetilde{\Omega}))} \leqslant c_3 \| \tilde{u} \|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leqslant c_4 \| f \|_{B_2^{t-1}(L_2(\Omega))}, \tag{45}$$

where $0 < q \leqslant 2$ and $\varepsilon > 0$ are arbitrary.

Substep 2.2: The regularity of v. An important theorem of Jerison and Kenig, see [17,18] and also [19, Theorem 5.1] (for $d \geqslant 3$), reads as

$$\| v \|_{B_2^{\varrho}(L_2(\Omega))} \leqslant c_5 \| g \|_{B_2^{\varrho-1/2}(L_2(\partial\Omega))} \quad \text{if } 1/2 < \varrho < 3/2. \tag{46}$$

Trace problems for Lipschitz boundaries are investigated in [20]. We refer to this monograph and to [19] also for the exact meaning of Tr and $B_2^{\varrho}(L_2(\partial\Omega))$, respectively. Theorem 2 on page 209 in [20] and (46) yield

$$\| v \|_{B_2^{\varrho}(L_2(\Omega))} \leqslant c_5 \| \mathrm{Tr} \| \, \| \tilde{u} \|_{B_2^{\varrho}(L_2(\widetilde{\Omega}))} \leqslant c_6 \| \tilde{u} \|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leqslant c_7 \| f \|_{B_2^{t-1}(L_2(\Omega))},$$

if $1/2 < \varrho < 3/2$ and $\varrho \leqslant t + 1$. Consequently $v \in B_2^{\vartheta}(L_2(\Omega))$ with $\vartheta < \min(3/2, t+1)$. A harmonic function in a bounded Lipschitz domain has a higher Besov regularity than Sobolev regularity. More precisely,

$$v \in B_\tau^s(L_\tau(\Omega)), \quad 0 < s < \frac{\vartheta d}{d-1}, \quad \frac{1}{\tau} = \frac{s}{d} + \frac{1}{2},$$

and

$$\| v \|_{B^s_\tau(L_\tau(\Omega))} \leqslant c_8 \| v \|_{B^\vartheta_2(L_2(\Omega))},$$

see [4]. Combining this with (45) we arrive at

$$u \in B^s_\tau(L_\tau(\Omega)), \quad 0 < s < \min\left(\frac{\vartheta d}{d-1}, t+1\right), \quad \frac{1}{\tau} = \frac{s}{d} + \frac{1}{2}, \tag{47}$$

together with the estimate

$$\|u\|_{B^s_\tau(L_\tau(\Omega))} \leqslant \|\tilde{u}\|_{B^s_\tau(L_\tau(\Omega))} + \|v\|_{B^s_\tau(L_\tau(\Omega))} \leqslant c_9 \|f\|_{B^{t-1}_2(L_2(\Omega))}.$$

Substep 2.3: An interpolation argument. Another theorem of Jerison and Kenig, see [19, Theorem 0.5], yields

$$u \in B^s_2(L_2(\Omega)), \quad s < \min(3/2, t+1), \quad \text{where } \|u\|_{B^s_2(L_2(\Omega))} \leqslant c_{10} \|f\|_{B^{t-1}_2(L_2(\Omega))}. \tag{48}$$

Thanks to the real interpolation formula

$$\left(B^{s_0}_{p_0}(L_{p_0}(\Omega)), B^{s_1}_{p_1}(L_{p_1}(\Omega))\right)_{\Theta, p} = B^s_p(L_p(\Omega)) \qquad \text{(equivalent quasi-norms)},$$

$$0 < \Theta < 1, \quad s = (1-\Theta)\, s_0 + \Theta\, s_1, \quad \frac{1}{p} = \frac{1-\Theta}{p_0} + \frac{\Theta}{p_1},$$

valid for all $s_0, s_1 \in \mathbb{R}$ and all $0 < p_0, p_1 < \infty$, cf. [41], we can combine these two different assertions (47), (48) about the regularity of u. Let e.g. $1/2 \leqslant t \leqslant (d+2)/(2d-2)$. Then we use the interpolation formula with

$$s_0 = 3/2 - \varepsilon, \quad s_1 = t+1 \quad \text{and} \quad \Theta = 1/3,$$

and find that $u \in B^s_{\tau^*}(L_{\tau^*}(\Omega))$, $s = 1 + (t+1)/3 - \varepsilon'$, $1/\tau^* = (s-1)/d + 1/2$, where

$$\|u\|_{B^s_{\tau^*}(L_{\tau^*}(\Omega))} \leqslant c_{11} \|f\|_{B^{t-1}_2(L_2(\Omega))},$$

and ε' can be made as small as we want. In summary, we have

$$\sup_{\|f\|_{B^{t-1}_2(L_2(\Omega))} \leqslant 1} \|u - S_n(f)\|_{H^1} \leqslant c_{12}\, n^{-(\frac{t+1}{3d} + \varepsilon)}.$$

The other two cases can be treated in an analoguous way. We omit details. □

Theorem 6 shows that best n-term wavelet approximation might be suboptimal in general. However, for more specific domains, i.e., for polygonal domains, much more can be said. Let Ω denote a simply connected polygonal domain contained in $\mathbb{R}^2$, the segments of $\partial\Omega$ are denoted by $\overline{\Gamma}_l$, Γ_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. Then the following theorem holds:

Theorem 7. *Let S denote the solution operator for the problem* (40) *in a polygonal domain in* $\mathbb{R}^2$. *Let k be a nonnegative integer such that*

$$k \neq \frac{m\pi}{\omega_l} \quad \text{for all } m \in \mathbb{N}, \quad l = 1, \ldots, N.$$

Then, for sufficiently large C, best n-term wavelet approximation S_n *yields*

$$e^{\text{non}}_{n,C}(S, H^{k-1}(\Omega), H^1(\Omega)) \leqslant e(S_n, H^{k-1}(\Omega), H^1(\Omega)) \leqslant c\, n^{-k/2+\varepsilon}, \tag{49}$$

where $\varepsilon > 0$ *is arbitrary and c does not depend on n.*

Proof. The proof is based on 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 $u_R \in B^{k+1}_2(L_2(\Omega))$ and u_S only depends on the shape of the domain and can be computed explicitly. This result was established by Grisvard, see [13] or [14, Chapters 4, 5], and [15, Section 2.7] for details. We introduce polar coordinates (r_l, θ_l) in the vicinity of each vertex Υ_l and introduce the functions

$$\mathcal{S}_{l,m}(r_l, \theta_l) = \zeta_l(r_l) r_l^{\lambda_{l,m}} \sin(m\pi\theta_l/\omega_l),$$

when $\lambda_{l,m} := m\pi/\omega_l$ is not an integer and

$$\mathcal{S}_{l,m}(r_l, \theta_l) = \zeta_l(r_l) r_l^{\lambda_{l,m}} [\log r_l \, \sin(m\pi\theta_l/\omega_l) + \theta_l \, \cos(m\pi\theta_l/\omega_l)]$$

otherwise, $m \in \mathbb{N}$, $l = 1, \ldots N$. Here ζ_l denotes a suitable C^∞ truncation function. Then for $f \in H^{k-1}(\Omega)$ one has

$$u_S = \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k} c_{l,m}\, \mathcal{S}_{l,m}, \tag{50}$$

provided that no $\lambda_{l,m}$ is equal to k. 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 (42), we have to estimate the Besov regularity of both, u_S and u_R, in the specific scale

$$B^\alpha_{\tau^*}(L_{\tau^*}(\Omega)) \quad \frac{1}{\tau^*} = \frac{(\alpha - 1)}{d} + \frac{1}{2}.$$

Since $u_R \in B^{k+1}_2(L_2(\Omega))$, classical embeddings of Besov spaces imply that

$$u_R \in B^{k+1-\varepsilon'}_{\tau*}(L_{\tau*}(\Omega)), \quad \frac{1}{\tau^*} = \frac{(k - \varepsilon')}{d} + \frac{1}{2} \quad \text{for arbitrary small } \varepsilon' > 0. \tag{51}$$

Moreover, it has been shown in [2] that the functions $\mathcal{S}_{l,m}$ defined above satisfy

$$\mathcal{S}_{l,m}(r_l, \theta_l) \in B^\alpha_{\tau*}(L_{\tau*}(\Omega)), \quad \frac{1}{\tau^*} = \frac{(\alpha - 1)}{d} + \frac{1}{2} \quad \text{for all } \alpha > 0. \tag{52}$$

By combining (51) and (52) we see that

$$u \in B^{k+1-\varepsilon'}_{\tau*}(L_{\tau*}(\Omega)), \quad \frac{1}{\tau^*} = \frac{(k - \varepsilon')}{d} + \frac{1}{2} \quad \text{for arbitrary small } \varepsilon' > 0.$$

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

$$\mathcal{N} := \text{span}\left\{\mathcal{S}_{l,m}(r_l, \theta_l) : \quad 0 < \lambda_{m,l} < k,\ l = 1, \dots, N\right\}.$$

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

$$H := \left\{u \in H^{k+1}(\Omega) : \quad \gamma_l u = 0,\ l = 1, \dots, N\right\} + \mathcal{N}$$

onto $H^{k-1}(\Omega)$, cf. [14, Theorem 5.1.3.5]. This mapping is also injective, see [14, Lemma 4.4.3.1, Remark 5.1.3.6]. We equip the space H with the norm

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

see (50). Then it becomes a Banach space. Furthermore, Δ is continuous. Banach's continuous inverse theorem implies that the solution operator is continuous considered as a mapping from $H^{k-1}(\Omega)$ onto H. Observe

$$\| u_R + u_S \|_{B^{k+1-\varepsilon'}_{\tau*}(L_{\tau*}(\Omega))} \leqslant c \left(\| u_R \|_{B^{k+1}_2(L_2(\Omega))} + \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k} |c_{l,m}| \right)$$

with some constant c independent of u. $\square$

Acknowledgments

We thank Stefan Heinrich, Aicke Hinrichs, Hans Triebel, Art Werschulz and two referees for their valuable remarks and comments.

References

[1] A. Cohen, Numerical Analysis of Wavelet Methods, Elsevier Science, Amsterdam, 2003.
[2] S. Dahlke, Besov regularity for elliptic boundary value problems in polygonal domains, Appl. Math. Lett. 12 (6) (1999) 31–38.
[3] S. Dahlke, W. Dahmen, R. DeVore, Nonlinear approximation and adaptive techniques for solving elliptic operator equations, in: W. Dahmen, A. Kurdila, P. Oswald (Eds.), Multiscale Wavelet Methods for Partial Differential Equations, Academic Press, San Diego, 1997, pp. 237–283.
[4] S. Dahlke, R. DeVore, Besov regularity for elliptic boundary value problems, Commun. Partial Differential Equations 22 (1&2) (1997) 1–16.
[5] K. Deimling, Nonlinear Functional Analysis, Springer-Verlag, Berlin, 1985.
[6] R.A. DeVore, Nonlinear approximation, Acta Numerica 7 (1998) 51–150.
[7] R.A. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, Manuscripta Math. 63 (1989) 469–478.
[8] R.A. DeVore, G. Kyriazis, D. Leviatan, V.M. Tikhomirov, Wavelet compression and nonlinear n-widths, Adv. Comput. Math. 1 (1993) 197–214.
[9] S. Dispa, Intrinsic characterizations of Besov spaces on Lipschitz domains, Math. Nachr. 260 (2002) 21–33.
[10] D. Dung, V.Q. Thanh, On nonlinear n-widths, Proc. of Amer. Math. Soc. 124 (1996) 2757–2765.
[11] E.G. D'yakonov, Optimization in Solving Elliptic Problems, CRC Press, Boca Raton, FL, 1996.

[12] D.E. Edmunds, H. Triebel, Function Spaces, Entropy Numbers, Differential Operators, Cambridge University Press, 1996.
[13] P. Grisvard, Behavior of solutions of elliptic boundary value problems in a polygonal or polyhedral domain, in: B. Hubbard (Ed.), Symposium on Numerical Solutions of Partial Differential Equations, vol. III, Academic Press, New York, 1975, pp. 207–274.
[14] P. Grisvard, Elliptic Problems in Nonsmooth Domains, Pitman, Boston, 1985.
[15] P. Grisvard, Singularites in Boundary Value Problems, Research Notes in Applied Mathematics, vol. 22, Springer, Berlin, 1992.
[16] W. Hackbusch, Elliptic Differential Equations: Theory and Numerical Treatment, Springer, Berlin, 1992.
[17] D. Jerison, C.E. Kenig, The Neumann problem in Lipschitz domains, Bull. Amer. Math. Soc. 4 (1981) 203–207.
[18] D. Jerison, C.E. Kenig, Boundary value problems on Lipschitz domains, in: Studies in PDE, MAA Stud. Math. 23 (1982) 1–68.
[19] D. Jerison, C.E. Kenig, The inhomogeneous Dirichlet problem in Lipschitz domains, J. Funct. Anal. 130 (1995) 161–219.
[20] A. Jonsson, H. Wallin, Function spaces on subsets of $\mathbb{R}^d$, Math. Reports, Harwood Acad. Publ., 1984.
[21] R.B. Kellogg, M. Stynes, n-widths and singularly perturbed boundary value problems, SIAM J. Numer. Anal. 36 (1999) 1604–1620.
[22] A. Kolmogoroff, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, Ann. Math. 37 (1936) 107–110.
[23] J.L. Lions, E. Magenes, Non-homogeneous Boundary Value Problems and Applications, vol. I, Springer, Berlin, 1972.
[24] P. Mathé, s-Numbers in information-based complexity, J. Complexity 6 (1990) 41–66.
[25] J.M. Melenk, On n-widths for elliptic problems, J. Math. Anal. Appl. 247 (2000) 272–289.
[26] Y. Meyer, Wavelets and Operators, Cambridge University Press, 1992.
[27] P.-A. Nitsche, Best N term approximation spaces for sparse grids, Research Report No. 2003-11, Seminar für Angewandte Mathematik, ETH Zürich, 2003.
[28] E. Novak, H. Triebel, Function spaces in Lipschitz domains and optimal rates of convergence for sampling, preprint, 2004.
[29] E. Novak, H. Woźniakowski, Complexity of linear problems with a fixed output basis, J. Complexity 16 (2000) 333–362.
[30] P. Oswald, Multilevel Finite Element Approximation, Theory and Applications, B.G. Teubner, Stuttgart, 1991.
[31] S.V. Pereverzev, Optimization of Methods for Approximate Solution of Operator Equations, Nova Science Publishers, New York, 1996.
[32] A. Pietsch, s-Numbers of operators in Banach spaces, Studia Math. 51 (1974) 201–223.
[33] A. Pietsch, Eigenvalues and s-Numbers, Geest und Portig, Leipzig, 1987.
[34] A. Pinkus, n-Widths in Approximation Theory, Springer, Berlin, 1985.
[35] V.S. Rychkov, On restrictions and extensions of Besov and Triebel–Lizorkin spaces with respect to Lipschitz domains, J. London Math. Soc. 60 (1999) 237–257.
[36] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, 1970.
[37] V.N. Temlyakov, Nonlinear methods of approximation, Found. Comput. Math. 3 (2003) 33–107.
[38] V.M. Tikhomirov, Approximation Theory, in: Encyclopaedia of Mathematical Sciences, vol. 14, Analysis II, Springer, Berlin, 1990.
[39] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, New York, 1988.
[40] H. Triebel, Theory of Function Spaces, vol. II, Birkhäuser, Basel, 1992.
[41] H. Triebel, Function spaces in Lipschitz domains and on Lipschitz manifolds, Characteristic functions as pointwise multipliers, Revista Matemática Complutense 15 (2002) 475–524.
[42] A.G. Werschulz, Finite element methods are not always optimal, Adv. in Appl. Math. 8 (1987) 354–375.
[43] A.G. Werschulz, The Computational Complexity of Differential and Integral Equations, Oxford Science Publications, Oxford, 1996.