Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings

Stephan Dahlke*, Erich Novak, Winfried Sickel

October 22, 2004

Abstract

We study the optimal approximation of the solution of an operator equation $\mathcal{A}(u) = f$ 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 $\mathcal{A} : H_0^s(\Omega) \to H^{-s}(\Omega)$, where $s > 0$ and Ω is an arbitrary bounded Lipschitz domain in $\mathbb{R}^d$. Here 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 might be very important for the approximation of elliptic problems. Our results are concerned with approximations and their errors, not with their numerical realization.

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, Bernstein widths, manifold widths.

*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.

1 Introduction

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

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

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

$$(2) \qquad \mathcal{A} : H \to G$$

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

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

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

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

where F is a normed (or quasi-normed) space, $F \subset G$. 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 $e_{n,C}^{\mathrm{non}}(S, F, H)$ 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, 20]. Sometimes these numbers 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

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

  holds for all $m \geq 4\, C^2\, n$, where $b_n(S, F, H)$ denotes the n-th 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 $C \geq 1$

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

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

- In Theorem 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 compared to 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.

We plan to continue this work with the assumption that F is a general Besov space. In this case the results (and the proofs) are much more difficult.

This paper is organized as follows. In Section 2 we discuss the basic concepts of optimality. In Subsection 2.1, we introduce in detail the linear, nonlinear and manifold widths and discuss their various relationships. Then, in Subsection 2.2, we study the relationships of these concepts to the well-known Bernstein widths. The main result in this section is Theorem 1 already mentioned above. In Section 3, we apply the general concepts to the more specific case of elliptic operator equations. After briefly introducing the basic function spaces that are needed, we first discuss regular problems. It turns out that in this case linear and nonlinear methods provide the same order of convergence, and uniform discretization schemes are sufficient. Then, in Subsection 3.3, we treat nonregular problems and state and prove the fundamental Theorem 4 mentioned above. In Subsection 3.4 we discuss the case that the linear functionals applied to the right-hand side are not arbitrary but given by function evaluations. In this case, the order of approximation decreases significantly. This means that arbitrary linear information gives a better rate of convergence compared to function evaluations. Finally, in Subsection 3.5, we apply the whole machinery to the case of best n-term wavelet approximation for the solution of the Poisson equation. There we state and prove the Theorems 6 and 7 discussed above.

2 Basic Concepts of Optimality

2.1 Classes of Admissible Mappings

Nonlinear Mappings S_n

We will study certain approximations of S based on Riesz bases, cf., e.g., Meyer [22, 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 $\alpha_1, \alpha_2, \dots$ with $\alpha_i \neq 0$ for only finitely many i, we have*

$$A\Big(\sum_k |\alpha_k|^2\Big)^{1/2} \le \Big\| \sum_k \alpha_k\, h_k \Big\|_H \le B\Big(\sum_k |\alpha_k|^2\Big)^{1/2} \tag{4}$$

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, Sect. 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 are 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 $f = \mathcal{A}(u)$. We assume that we can choose $\mathcal{B}$ and of course we have in mind 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

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

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] and [31]. By the arbitrariness of S_n one obtains immediately

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

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

$$\mathcal{B}_C := \Big\{ \mathcal{B} : \; B/A \le C \Big\}. \tag{9}$$

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

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

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). We call these numbers *nonlinear widths*, but this name will also be used for the numbers e_n^{cont} that we discuss below. In this paper we investigate the numbers $e_{n,C}^{\mathrm{non}}(S,F,H)$ only in cases where H is a Hilbert space. More general concepts are introduced and investigated in [31].

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 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.*

Remark 4. *It is obvious from the definition* (10) *that $S_n^* \in \mathcal{N}_n(B)$ can be (almost) optimal (for the given $\mathcal{B}$) 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 $e_{n,C}^{\mathrm{non}}(S,F,H)$ is much smaller, since the given $\mathcal{B}$ is far from being optimal. See also Remark 10.

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

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

The numbers $e_n^{\mathrm{lin}}(S,F,H)$ (or slightly different numbers) are usually called *approximation numbers or linear widths* of $S : F \to H$, cf. [20, 29, 30, 32].

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*

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

where $\mathcal{L}_n^{\text{std}} \subset \mathcal{L}_n$ 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_n^{lin} we only allow *standard information*, i.e., function values of the right-hand side. The inequality $g_n^{\text{lin}}(S, F, H) \geq e_n^{\text{lin}}(S, F, H)$ 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* $g_n(S, F, H)$.

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

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

These numbers, or slightly different numbers, were studied by different authors, cf. [7, 8, 10, 20]. Sometimes these numbers are called manifold widths of S, see [8]. The inequalities

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

and

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

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}. The number $b_n(S, F, H)$, called 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 $S(\{\|f\|_F \leq 1\})$. As it is well-known, Bernstein widths are useful for the proof of lower bounds, see [7, 10, 30].

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

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

holds for all n.

Proof. We assume that $S(\{\|f\|_F \leq 1\})$ contains an $(n+1)$-dimensional ball $B \subset H$ of radius r. We may assume that the center is in the origine. 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 $\|f\|_F = 1$. 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

$$b_n(S, F, H) \leq e_n^{\mathrm{cont}}(S, F, H)\,.$$

$\square$

We will see that the b_n can also be used to prove lower bounds for the $e_{n,C}^{\mathrm{non}}$. First we need some lemmata. 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. *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.*

Proof. Let

$$V = \Big\{ \sum_{i=1}^{n} \lambda_i v^i \; : \quad \lambda_i \in \mathbb{R} \Big\}\,,$$

where the v^i are linearly independent elements in H. We argue by contradiction. To this end we assume that there only exist a natural number $m < n$ and an element $x^* \in V$ such that

$$1 = |x_1^*| = |x_2^*| = \;\ldots\; = |x_m^*| > |x_j^*|$$

holds for all $j > m$. We put

$$\mathcal{V} = \Big\{ x \in V \; : \quad x_j = x_j^*, \quad j = 1, \ldots, m \Big\}\,.$$

Of course $x^* \in \mathcal{V}$. Let

$$V_0 = \Big\{ x \in V \ : \quad x_1 = x_2 = \ \ldots \ = x_m = 0 \Big\} \, ,$$

Then elementary linear algebra yields $\dim V_0 = n - m \ge 1$. Selecting $v \in V_0$ such that $\|v\|_\infty = 1$ we obtain

$$\Big\{ x^* + \lambda v \ : \quad \lambda \in \mathbb{R} \Big\} \subset \mathcal{V} \, .$$

Define

$$g(\lambda) := \sup_{j=m+1,\ldots} |x^*_j + \lambda\, v_j| \, , \qquad \lambda \in \mathbb{R} \, .$$

Then $g(\lambda) \to \infty$ if $|\lambda| \to \infty$. Because of $x^*, v \in c_0$ we have

$$\sup_{j=m+1,\ldots} |x^*_j| = |x^*_{j_0}| < 1 \qquad \text{and} \qquad \sup_{j=m+1,\ldots} |v_j| = |v_{j_1}| \le 1 \, .$$

We choose $\lambda > 0$ such that $|x_{j_0}| + \lambda\, |v_{j_1}| < 1$. Then $g(\lambda) < 1$ follows. The function g is continuous. Hence there exists a number $\lambda_0 > 0$ and an index j_2 such that

$$1 = g(\lambda_0) = \sup_{j=m+1,\ldots} |x^*_j + \lambda_0\, v_j| = |x^*_{j_2} + \lambda_0\, v_{j_2}| \, .$$

With $x^* + \lambda_0\, v$ we arrive at a contradiction. □

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 $0 < A \le B < \infty$. Then there is an nontrivial element $x \in V_n$ such that $x = \sum_{j=1}^\infty x_j\, h_j$ and*

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

Proof. Associated to any $x \in H$ there is a sequence $(x_j)_j$ of coefficients with respect to $\mathcal{B}$ which belongs to c_0. In the same way we associate to $V_n \subset H$ 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

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

This implies

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

□

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

$$e^{\rm non}_{n,C}(S,F,H) \ \ge \ \frac{1}{2C}\, b_m(S,F,H) \tag{19}$$

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

Proof. Let $\mathcal{B}$ be a Riesz basis with Riesz constants A and B and let $m > n$. Assume that $S(\{\|f\| \leq 1\})$ contains an m-dimensional ball with radius ε. Then, according to Lemma 3, there exists an $x \in S(\{\|f\| \leq 1\})$ such that $x = \sum_i x_i h_i$, $\|x\| = \varepsilon$ and $|x_i| \leq A^{-1} m^{-1/2} \varepsilon$ 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

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

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

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

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

Remark 5. *We do not believe that the constant $1/(2C)$ is 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 [28].

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

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

Remark 6. *There are different definitions of s-numbers in the literature. In particular, the definition used in the monograph [29] does not coincide with that one from the article [28]. In the sense of [29] the Bernstein numbers are not s-numbers, but they are s-numbers in the sense of [28].*

Arbitrary continuous mappings cannot be better than linear mappings. This is a general result for Hilbert spaces. 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 optimal linear mappings have 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). Then*

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

Proof. This follows from Theorem 1 and Theorem 2, together with (22). ☐

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 denote by $H^m(\Omega)$ the set of all functions $u \in L_2(\Omega)$ such that the (distributional) derivatives $D^\alpha u$ of order $|\alpha| \le m$ also belong to $L_2(\Omega)$. This set equipped with the norm

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

becomes a Hilbert space. For positive noninteger s we define $H^s(\Omega)$ as specific Besov spaces. If $h \in \mathbb{R}^d$, we denote by Ω_h 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 $u \in L_p(\Omega)$, $0 < p \le \infty$, is defined by

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

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

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

is finite with $r := [s]+1$. 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 definition based, e.g., on complex interpolation as outlined in Lions and Magenes [19, Vol. 1]. We refer to [9, 33] for details.

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

$$H^{-s}(\Omega) := (H_0^s(\Omega))' .$$

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 $S = \mathcal{A}^{-1}$. Then 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 [15]. We use the following definition and assume that $t > 0$.

Definition 2. *Let $s > 0$. An operator $\mathcal{A} : H_0^s(\Omega) \to H^{-s}(\Omega)$ is H^{s+t}-*regular *if in addition to* (3) *also*

$$\mathcal{A} : H_0^s(\Omega) \cap H^{s+t}(\Omega) \to H^{-s+t}(\Omega) \tag{25}$$

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 [15] for further information and examples. It is known that in this situation we obtain the optimal rate

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

of linear methods. This is a classical result, at least for $t, s \in \mathbb{N}$ and for special domains, such as $\Omega = [0,1]^d$. We refer to the books [11, 27, 35] that contain hundreds of references.

We prove that this result 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 can not obtain a better rate of convergence.

Theorem 3. *Assume that the problem is H^{s+t}-regular, as in* (25)*. Then for all $C \geq 1$*

$$e_n^{\mathrm{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp e_{n,C}^{\mathrm{non}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d} \tag{26}$$

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 (or 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 [18]) for $s, t \in \mathbb{N}$, see also [30]. For the general case ($s, t > 0$ and arbitrary bounded Lipschitz domains) see [12] and [33]. We obtain the same order for $I : H^{s+t}(\Omega) \cap H_0^s(\Omega) \to H^s(\Omega)$.

We assume (25) 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 Corollary 1 this proves (26).

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 $S : H^{-s+t}(\Omega) \to H^{s+t}(\Omega) \subset H^s(\Omega)$. □

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 $g_1, \ldots, g_n$. 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:*

$$(27) \quad u \in H^{t+s}(\Omega) \Longleftrightarrow \sum_{n=1}^{\infty} [n^{t/d} E_n(u)]^2 \frac{1}{n} < \infty, \quad \textit{where} \quad E_n(u) := \inf_{g \in V_n} \|u - g\|_{H^s},$$

see, e.g, [3, 26] *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 [17, 21, 34]. 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 still have for all $C \geq 1$*

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

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 ($s, t > 0$, and arbitrary bounded Lipschitz domains) see [33].

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

Assume that the linear mapping

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

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

$$\sum_{i=1}^{n} S(g_i) \, L_i(f) \tag{29}$$

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

Remark 9. *It is well-known that uniform methods can be quite bad for problems that are not regular. Indeed, the general characterization* (27) *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* (29) *then we need to know the $S(g_i)$. Observe also that in this case we need a 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

$$\inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{-t/d} \tag{30}$$

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 [23] *and* [25].

3.4 Function Values

Now we study the numbers $g_n(S, H^{t-s}(\Omega), H^s(\Omega)) = g_n^{\text{lin}}(S, H^{t-s}(\Omega), H^s(\Omega))$ 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, in order to have function values that continuously depend on $f \in H^{t-s}(\Omega)$, that $t > s + d/2$.

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

$$e_n^{\text{lin}}(I, H^t(\Omega), L_2(\Omega)) \asymp g_n^{\text{lin}}(I, H^t(\Omega), L_2(\Omega)) \asymp n^{-t/d}, \tag{31}$$

see [24]. This means that arbitrary linear functionals do not yield a better order of convergence than function values. Furthermore we have $g_n = g_n^{\text{lin}}$, since we have Hilbert spaces.

It is interesting that for $s > 0$ arbitrary linear information is better compared to function evaluation.

Theorem 5. *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*

$$g_n(S, H^{-s+t}(\Omega), H^s(\Omega)) = g_n^{\mathrm{lin}}(S, H^{-s+t}(\Omega), H^s(\Omega)) \asymp n^{(s-t)/d}, \tag{32}$$

for $t > s + d/2$.

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

$$g_n(I, H^{-s+t}(\Omega), H^{-s}(\Omega)) \asymp n^{(s-t)/d}. \tag{33}$$

To prove the upper and the lower bound for (33), we use several auxiliary problems and start with the upper bound. It is known from [24] that

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

From this we obtain the upper bound

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

by embedding.

For the lower bound we use the bound

$$g_n(I, H^{-s+t}(\Omega), L_1(\Omega)) \asymp n^{(s-t)/d}, \tag{34}$$

again from [24]. The lower bound in (34) is proved by the technique of bump functions: Given $x_1, \ldots, x_n \in \Omega$, 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

$$\|f\|_{L_1} \geq c \cdot n^{(s-t)/d}, \tag{35}$$

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

$$\mathrm{Int}(f) = \int_\Omega f \sigma \, dx, \tag{36}$$

where $\sigma \geq 0$ is a smooth (and nonzero) function on Ω with compact support. Then this technique gives: Given $x_1, \ldots, x_n \in \Omega$, 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

$$\mathrm{Int}(f) \geq c \cdot n^{(s-t)/d}, \tag{37}$$

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

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

and hence we may replace in (37) $\mathrm{Int} f$ by $\|f\|_{H^{-s}}$, hence

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

□

3.5 The Poisson Equation

Finally we discuss our results for the specific case of the Poisson equation on a bounded Lipschitz domain Ω contained in $\mathbb{R}^d$

$$\begin{aligned} -\triangle u &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega. \end{aligned} \tag{38}$$

It is well-known that (38) fits into our setting with $s = 1$. Indeed, if we consider this problem in the weak formulation, it can be checked that (38) induces a boundedly invertible operator $\mathcal{A} = \triangle : \ H_0^1(\Omega) \longrightarrow H^{-1}(\Omega)$, see again [15] for details. Here we meet the problem of existence of a Riesz basis for $H_0^1(\Omega)$. 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 [22]. 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

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

- the wavelets satisfy the *cancellation property*

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

  where $\widetilde{m}$ denotes some suitable parameter;

- the wavelet basis induces characterizations of Besov spaces of the form

$$\|f\|_{B^s_q(L_p(\Omega))} \asymp \left(\sum_{|\lambda|=j_0}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \left(\sum_{\lambda \in \mathcal{J}, |\lambda|=j} |\langle f, \tilde{\psi}_\lambda \rangle|^p \right)^{q/p} \right)^{1/q}, \tag{39}$$

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

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

By exploiting the norm equivalence (39) 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 the characterization (39) also carries over to $H^s_0(\Omega)$. We refer to [1] for a detailed discussion. In this setting, the following theorem holds.

Theorem 6. *For the problem* (38), *best n-term wavelet approximation produces the worst case error estimate:*

$$e(S_n, H^{t-1}(\Omega), H^1(\Omega)) \leq C\, n^{-(\frac{(t+1)}{3} - \varrho)/d} \quad \textit{for all} \quad \varrho > 0, \tag{40}$$

provided that $\frac{1}{2} < t \leq \frac{3d}{2(d-1)} - 1$.

Proof. It is well-known that

$$\|u - S_n(f)\|_{H^1} \leq C|u|_{B^\alpha_{\tau^*}(L_{\tau^*}(\Omega))} n^{(\alpha-1)/d}, \qquad \frac{1}{\tau^*} = \frac{(\alpha-1)}{d} + \frac{1}{2}, \tag{41}$$

see, e.g., [3] for details. We therefore have to estimate the Besov norm $B^\alpha_{\tau^*}(L_{\tau^*}(\Omega))$. We do this is in two steps. First of all, we estimate the Besov norm of u in the specific scale

$$B^s_\tau(L_\tau(\Omega)), \quad \frac{1}{\tau} = \frac{s}{d} + \frac{1}{2}. \tag{42}$$

Regularity estimates in the scale (42) have already been performed in [4]. We write the solution u to (38) 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. Furthermore, v is the solution to the additional Dirichlet problem

$$\begin{aligned} \triangle v &= 0 \\ v &= g = \mathrm{Tr}(\tilde{u}). \end{aligned} \tag{43}$$

Then, by classical elliptic regularity on smooth domains we observe that

$$\tilde{u} \in B_2^{t+1}(L_2(\widetilde{\Omega})), \quad \|\tilde{u}\|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leq C\|\mathcal{E}\|\|f\|_{B_2^{t-1}(L_2(\Omega))}$$

and hence by embeddings of Besov spaces

$$\|\tilde{u}\|_{B_\tau^{t+1-\varepsilon}(L_\tau(\widetilde{\Omega}))} \leq C\|\tilde{u}\|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leq C\|\mathcal{E}\|\|f\|_{B_2^{t-1}(L_2(\Omega))}.$$

Now let t be chosen in such a way that $t > 1/2$. By construction,

$$\begin{aligned} \|g\|_{B_2^\varrho(L_2(\partial\Omega))} &= \|\mathrm{Tr}(\tilde{u})\|_{B_2^\varrho(L_2(\partial\Omega))} \leq C\|\mathrm{Tr}\|\|\tilde{u}\|_{B_2^{\varrho+1/2}(L_2(\widetilde{\Omega}))} \\ &\leq C\|\mathrm{Tr}\|\|\tilde{u}\|_{B_2^{t+1}(L_2(\widetilde{\Omega}))} \leq C\|\mathrm{Tr}\|\|\mathcal{E}\|\|f\|_{B_2^{t-1}(L_2(\Omega))}, \quad \varrho < 1. \end{aligned}$$

Then a famous theorem of Jerison and Kenig [16] implies that

$$\|v\|_{B_2^{\varrho+1/2}(L_2(\Omega))} \leq C\|g\|_{B_2^\varrho(L_2(\partial\Omega))}, \quad \text{if} \quad \varrho < 1,$$

and therefore

$$\|v\|_{B_2^{\varrho+1/2}(L_2(\Omega))} \leq C\|\mathrm{Tr}\|\|\mathcal{E}\|\|f\|_{B_2^{t-1}(L_2(\Omega))}, \qquad \varrho < 1.$$

In [4], Theorem 3.2 the following fact has been shown:

$$\|v\|_{B_\tau^s(L_\tau(\Omega))} \leq C\|v\|_{B_2^{\varrho+1/2}(L_2(\Omega))}, \quad 0 < s < \frac{(\varrho+1/2)d}{d-1}.$$

Consequently, if $t+1 \leq \frac{3d}{2(d-1)}$,

$$\begin{aligned} \|u\|_{B_\tau^{t-\varepsilon}(L_\tau(\Omega))} &\leq \|\tilde{u}\|_{B_\tau^{t+1-\varepsilon}(L_\tau(\Omega))} + \|v\|_{B_\tau^{t+1-\varepsilon}(L_\tau(\Omega))} \\ &\leq C(\Omega, \mathrm{Tr}, \mathcal{E})\|f\|_{B_2^{t-1}(L_2)}. \end{aligned}$$

So far, we have shown that all the solutions u of (38) are contained in a Besov ball in the space $B_\tau^{t-\varepsilon}(L_\tau(\Omega))$. However, another theorem of Jerison and Kenig [16] implies that

$$u \in B_2^{3/2-\varepsilon}(L_2(\Omega)), \qquad \|u\|_{B_2^{3/2-\varepsilon}(L_2(\Omega))} \leq C\|f\|_{B_2^{1/2-\varepsilon}(L_2(\Omega))} \leq C\|f\|_{B_2^{t-1}(L_2(\Omega))}.$$

Then, by interpolation between the spaces $B_\tau^{t+1-\varepsilon}(L_\tau(\Omega))$ and $B_2^{3/2-\varepsilon}(L_2(\Omega))$ we conclude that

$$u \in B_{\tau^*}^{s^*-\varrho}(L_{\tau^*}(\Omega)), \quad \frac{1}{\tau^*} = \frac{s^*-\varrho-1}{d} + \frac{1}{2}, \quad s^* = \frac{t+1}{3} + 1,$$

and

$$\|u\|_{B_{\tau^*}^{s^*-\varrho}(L_{\tau^*}(\Omega))} \leq C\|f\|_{B_2^{t-1}(L_2(\Omega))},$$

see also [3] for details. In summary, we have

$$\sup_{\|f\|_{B_2^{t-1}(L_2(\Omega))}} \|u - S_n(f)\|_{H^1} \leq Cn^{-((\frac{t+1}{3})-\varrho)/d}.$$

$\square$

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 $\overline{\Gamma}_l, \Gamma_l$ 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. Then the following theorem holds:

Theorem 7. *For problem* (38) *in a polygonal domain in* $\mathbb{R}^2$*, best* n*-term wavelet approximation is almost optimal in the sense that*

$$e(S_n, H^{t-1}(\Omega), H^1(\Omega)) \leq Cn^{-(t-\varrho)/2}, \qquad \text{for all} \quad \varrho > 0. \tag{44}$$

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 $u_R \in B_2^{t+1}(L_2(\Omega))$ and u_S only depends on the shape of the domain and can be computed explicitly. This result was established by Grisvard, see [14], Chapter 2.7, and [13] 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, $m \in \mathbb{N}$. Here ζ_l denotes a suitable C^∞ truncation function. Then for $f \in H^{t-1}(\Omega)$ one has

$$u_S = \sum_{j=1}^{N} \sum_{0<\lambda_{l,m}<t} c_{l,m} \mathcal{S}_{l,m}, \tag{45}$$

provided that no $\lambda_{l,m}$ is equal to t. 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 t. According to (41), we have to estimate the Besov regularity of both, u_S and u_R, in the specific scale

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

Since $u_R \in B_2^{t+1}(L_2(\Omega))$, classical embeddings of Besov spaces imply that

$$u \in B^{t+1-\varrho}_{\tau*}(L_{\tau*}(\Omega)) \qquad \frac{1}{\tau^*} = \frac{(t-\varrho)}{d} + \frac{1}{2} \qquad \text{for all} \quad \varrho > 0. \tag{46}$$

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

$$\mathcal{S}_{l,m}(r_l, \theta_l) \in B^{\alpha}_{\tau*}(L_{\tau*}(\Omega)), \qquad \frac{1}{\tau^*} = \frac{(\alpha-1)}{d} + \frac{1}{2} \quad \text{for all} \quad \alpha > 0. \tag{47}$$

By combining (46) and (47), the result follows. □

Acknowledgment. We thank Aicke Hinrichs who helped with Theorem 2, see Remark 6.

References

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

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

[3] S. Dahlke, W. Dahmen, and 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, and P. Oswald, eds.), Academic Press, San Diego, 237–283.

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

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

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

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

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

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

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

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

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

[13] 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.

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

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

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

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

[18] A. Kolmogoroff (1936): Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Annals of Math.* **37**, 107–110.

[19] J. L. Lions and E. Magenes (1972): *Non-homogeneous boundary value problems and applications I.* Springer, Berlin.

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

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

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

[23] P.-A. Nitsche (2003): Best N term approximation spaces for sparse grids. Research Report No. 2003-11, Seminar für Angewandte Mathematik, ETH Zürich.

[24] E. Novak and H. Triebel (2004): Function spaces in Lipschitz domains and optimal rates of convergence for sampling. Preprint.

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

[26] P. Oswald (1991): *Multilevel Finite Element Approximation. Theory and applications.* B. G. Teubner, Stuttgart.

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

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

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

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

[31] V. N. Temlyakov (2003): Nonlinear methods of approximation. Preprint.

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

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

[34] A. G. Werschulz (1987): Finite element methods are not always optimal. *Adv. in Appl. Math.* **8**, 354–375.

[35] 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}/