MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

OPTIMIZED GENERAL SPARSE GRID APPROXIMATION SPACES FOR OPERATOR EQUATIONS

M. GRIEBEL AND S. KNAPEK

Abstract. This paper is concerned with the construction of optimized sparse grid approximation spaces for elliptic pseudodifferential operators of arbitrary order. Based on the framework of tensor-product biorthogonal wavelet bases and stable subspace splittings, we construct operator-adapted subspaces with a dimension smaller than that of the standard full grid spaces but which have the same approximation order as the standard full grid spaces, provided that certain additional regularity assumptions on the solution are fulfilled. Specifically for operators of positive order, their dimension is $O(2^J)$ independent of the dimension n of the problem, compared to $O(2^{Jn})$ for the full grid space. Also, for operators of negative order the overall cost is significantly in favor of the new approximation spaces. We give cost estimates for the case of continuous linear information. We show these results in a constructive manner by proposing a Galerkin method together with optimal preconditioning. The theory covers elliptic boundary value problems as well as boundary integral equations.

1. Introduction

In this paper we deal with the construction of finite element spaces for the approximate solution of elliptic problems in Sobolev spaces $\mathcal{H}^s(\Omega), s \in \mathbb{R}$. It is well-known [48, 50] that the cost of approximatively solving, e.g., Poisson's equation in n dimensions in the Sobolev space $\mathcal{H}_0^1 \cap \mathcal{H}^2$ on a bounded domain up to an accuracy of ϵ is $O(\epsilon^{-n})$, i.e., it is exponentially dependent on n. This dependence on n is called the curse of dimensionality. Hence, for higher-dimensional problems, a direct numerical solution on a regular uniform mesh is prohibitive [48]. The curse of dimensionality can be overcome if additional assumptions on the regularity of the solution of the elliptic problem are posed, i.e., if we further restrict the space from which the solution is allowed to be.[1] To this end, if the solution is in the space of functions with dominating mixed second derivative the cost reduces to $O(\epsilon^{-1})$ [5, 6, 21].

However, standard finite element algorithms that use regular *full grids* for the discretization lead to a cost of $O(\epsilon^{-n})$. Hence they are not suited for such problems. A corresponding algorithm that realizes the cost of $O(\epsilon^{-1})$ up to an additional logarithmic term has been presented in [18, 52]. It uses tensor-products of hierarchical linear splines as basis functions, and the so-called regular *sparse grid*

2000 *Mathematics Subject Classification.* Primary 41A17, 41A25, 41A30, 41A65, 45L05, 45L10, 65D99, 65N12, 65N30, 65N38, 65N55.

[1]Note that this translates to a restriction on the data of the problem, i.e., on its right hand side and/or boundary conditions.

©XXXX American Mathematical Society

space/hyperbolic cross space as approximation space in the finite element method. Together with an optimal multigrid solver for sparse grid discretizations [19, 27], the cost of this algorithm is $O(\epsilon^{-1} \cdot |\ln \epsilon|^{n-1})$. See also the subsequent papers [3, 29]. For a survey on sparse grids see [6].

This regular sparse grid scheme is well established in approximation and interpolation theory. It continues to attract significant attention and has also been used successfully in connection with other multiscale function systems, for example prewavelets, interpolets and higher order splines. It was successfully employed for the solution of partial differential equations [3, 4, 6, 17, 20, 26, 29]. Furthermore, the sparse grid approach together with prewavelets was used in [28] for the solution of boundary integral equations of order between $-\frac{1}{2}$ and $\frac{1}{2}$. For a study of the complexity of integral equations with smooth kernels we refer to [16, 32, 33, 34, 40]. The regular sparse grid approach still involves a J^{n-1}-term in its cost complexity. Therefore the curse of dimensionality is still somewhat present, albeit only for the logarithmic J-term. In practice this limits the method to problems with up to about 12 dimensions. In [5] it was shown shown how to get rid of the additional logarithmic term by the use of a *subspace* of the sparse grid space. This results in so-called energy-norm based sparse grid spaces. Then the overall cost for the solution of Poisson's equation is indeed of the order $O(\epsilon^{-1})$.

In this paper we generalize the construction of [5] to differential and pseudo-differential operators of arbitrary order $s \in \mathbb{R}$, compare also [22, 23, 24, 25, 30, 32, 33, 34]. We construct operator adapted finite-element subspaces of lower dimension than the standard full-grid spaces. These new subspaces preserve the approximation order of the standard full-grid spaces, provided that certain additional regularity assumptions are fulfilled, i.e., we assume that the solution possesses $\mathcal{H}^{t,l}_{\text{mix}}$-regularity. To this end, we analyze the approximation of the embedding $\mathcal{H}^{t,l}_{\text{mix}} \to \mathcal{H}^s$ on the n-dimensional torus, i.e., we measure the approximation error in $\mathcal{H}^s$ and estimate it from above by terms involving the $\mathcal{H}^{t,l}_{\text{mix}}$-norm of the solution. Here $\mathcal{H}^{t,l}_{\text{mix}}$ is a certain intersection of classes of functions with bounded mixed derivatives, see (2.25) below, and $\mathcal{H}^s$ is the standard Sobolev space. The parameter l in $\mathcal{H}^{t,l}_{\text{mix}}$ governs the isotropic smoothness whereas t governs the dominating mixed smoothness. We use norm equivalences to facilitate the decoupling of the subspaces arising from the tensor-product approach and to ensure the stability of the resulting subspace splittings. Hence, the analysis is reduced to diagonal mappings between Hilbert sequence spaces. It turns out that the optimal approximation space in terms of the quotient of cost versus accuracy is only dependent on the quotient $(s-l)/t$ of isotropic smoothness $s-l$ to dominating mixed smoothness t. We will identify those approximation spaces that lead to algorithms with minimal cost. Specifically we show that one can break the curse of dimension in the case $(s-l)/t > 0$ and get rid of all asymptotic dependencies[2] on n. In the case $(s-l)/t < 0$ (e.g., for operators of negative order and spaces with dominating mixed derivative) there remains a certain moderate dependence on the dimension.

[2]Note that the constant in the order notation still depends on the dimension n. In a very special case, i.e., for the parameters $s = 1, t = 2, l = 0$, for the unit cube as domain and for the hierarchical Faber basis as expansion system, it was shown that it can be bounded by $Cn^2 0.97515^n \|u\|_{\mathcal{H}^{t,l}_{\text{mix}}}$, i.e., the part of the constant related to the approximation scheme decays exponentially with n, which is a manifestation of the concentration of measure phenomenon. For details see [21].

The remainder of this paper is as follows. Section 2 introduces some notation, collects basic facts about biorthogonal wavelet bases and tensor-product spaces and gives motivations for the construction of optimized sparse grids. Section 3 contains some theory about norm equivalences in Sobolev spaces. In Section 4 the optimized spaces are defined and estimates on the dimension of the optimized spaces and their order of approximation are given. Section 5 contains remarks on the cost complexity of solving elliptic equations for the case that continuous linear information is permissible, i.e., that the stiffness matrix as well as load vector can be computed exactly. We show these results in a constructive manner by proposing a Galerkin method working with the optimized approximation spaces together with a multilevel preconditioned iterative solver. Section 6 discusses two elliptic problems as examples of our theory, the Poisson problem and the single layer potential equation. Section 7 collects further generalizations for the construction of optimized grids. Some concluding remarks close the paper.

2. Motivation and Assumptions

Let us denote by $\mathcal{H}^t(T^n), t \in \mathbb{R}$, a scale of Sobolev spaces on the n-dimensional torus, and $\mathcal{L}^2(T^n)$ the space of $\mathcal{L}^2$-integrable functions on T^n, see [1]. For ease of presentation and to avoid restrictions on the smoothness exponent t we restrict ourselves to the n-dimensional torus in the first parts of this paper.[3] Applications to non-periodic problems will be discussed in Section 6. We represent T^n by the n-dimensional cube $T := [0,1]$, $T^n = T \times T \times \cdots \times T$ where opposite faces are identified. If $t < 0$, then $\mathcal{H}^t(T^n)$ is defined as the dual of $\mathcal{H}^{-t}(T^n)$, i.e.,

$$\mathcal{H}^t(T^n) := (\mathcal{H}^{-t}(T^n))'. \tag{2.1}$$

When the meaning is clear from the context, we will write $\mathcal{H}^t$ instead of $\mathcal{H}^t(T^n)$ and we proceed analogously for other function spaces.

Consider an elliptic variational problem: Given $f \in \mathcal{H}^{-s}$, find $u \in \mathcal{H}^s$ such that

$$a(u,v) = (f,v) \ \ \forall v \in \mathcal{H}^s, \tag{2.2}$$

where a is a symmetric positive definite form satisfying[4]

$$a(v,v) \approx \|v\|^2_{\mathcal{H}^s}. \tag{2.3}$$

Here $x \approx y$ means that there exist C_1, C_2 independent of any parameters x or y may depend on, such that $C_1 \cdot y \le x \le C_2 \cdot y$.

In the rest of the paper C denotes a generic constant that may depend on the smoothness assumptions and on the dimension n of the problem, but does not depend on the number of levels J. In the following multi-indices (vectors) are written boldface, for example $\mathbf{j}$ for $(j_1, \ldots, j_n)$. Inequalities like $\mathbf{l} \le \mathbf{t}$ or $\mathbf{l} \le \mathbf{0}$ are to be understood componentwise.

Model examples for (2.2) would be the variational form of the biharmonic equation ($s = 2$)

$$\Delta^2 u = f,$$

[3]Note that the non-periodic case with, e.g., Dirichlet or Neumann boundary conditions can be basically treated in the same way. Then, basis functions whose support intersects with the boundary have to fulfil special boundary conditions.

[4]Clearly, the lower estimate $a(u,u) \ge \alpha \cdot \|u\|^2_{\mathcal{H}^s}$ in (2.3) is in general not fulfilled for problems on the torus without additional constraints ensuring uniqueness of the solution of (2.2). In the following we will assume that the solution of the variational problem (2.2) is unique. Note however that for the construction of optimized grids, we will only need the upper estimate in (2.3).

which has applications in plate bending and shell problems or the (anisotropic) Helmholtz equation ($s = 1$)

$$-\nabla \cdot K\nabla u + cu = f \text{ on } T^n, \tag{2.4}$$

where $K(x) \approx I$ and $\exists\, C > 0 : 0 \leq c(x) \leq C$, modeling for example the single phase flow in a porous medium with permeability K, or a diffusion process in a (possibly) anisotropic medium characterized by the diffusion tensor K. Other examples would be the hypersingular equation ($s = \frac{1}{2}$)

$$\frac{1}{c} \int\limits_{T^n} \frac{\partial}{\partial n_x} \frac{\partial}{\partial n_y} \left(\frac{1}{|x-y|} \right) \cdot u(y)dy = f(x),$$

Fredholm equations of the second kind ($s = 0$) $u(x) - \int\limits_{T^n} k(x,y)\ u(y)dy = f(x)$, with given kernel function k defined on $T^n \times T^n$, specifically the double layer potential equation

$$u(x) - \frac{1}{c} \int\limits_{T^n} \frac{n_y \cdot (y-x)}{|x-y|^3}\ u(y)dy = f(x),$$

arising from a reformulation of Laplace's equation via the indirect method, or the single layer potential equation ($s = -\frac{1}{2}$)

$$\frac{1}{c} \int\limits_{T^n} \frac{u(y)}{|x-y|}\ dy = f(x). \tag{2.5}$$

The Galerkin method for numerically solving problem (2.2) selects a finite dimensional subspace from $\mathcal{H}^s \cap \mathcal{L}^2$ and solves the variational problem in this subspace instead of $\mathcal{H}^s$. It is well-known that the most efficient way of solving such problems exploits the interaction of several scales of discretization. These multilevel schemes use a sequence of closed nested subspaces $S_0 \subset S_1 \subset \cdots \subset \mathcal{H}^s \cap \mathcal{L}^2$ of the basic Hilbert space $\mathcal{H}^s$, whose union is dense in $\mathcal{H}^s$. Fixing a basis of S_J finally leads to a linear system of equations

$$A_J x_J = b_J \tag{2.6}$$

of dimension $\dim(S_J)$. Here A_J is called the stiffness matrix and b_J is the load vector. Storage requirements and computation time mostly exclude the use of direct solvers, since $\dim(S_J)$ is usually very large. Specifically for full grid spaces with subdivision rate two we have $\dim(S_J) = O(2^{J \cdot n})$. That is, the dimension of S_J grows exponentially with the dimension n.

In order to iteratively solve (2.2) or (2.6), respectively, the following problems and questions arise. Accuracy requirements necessitate a fine partitioning of T^n , i.e., $\dim(S_J)$ is large. Is it possible to select S_J as a subspace of the full grid space with $\dim(S_J)$ only polynomially dependent on the dimension n, compared to an exponential dependence on n of the dimension of the full grid space? Choosing such a finite element space would require that one can identify those basis functions that add most to an accurate representation of the solution of the variational problem. For differential operators, the resulting linear systems are sparse if the basis functions have local support. However, the discretization of integral operators results in most cases in discrete systems that are dense, i.e., on a regular full grid with $O(2^{nJ})$ unknowns the discrete operator has $O(2^{2nJ})$ entries. This makes matrix-vector multiplications, as they are needed in iterative methods, prohibitively expensive

for large n and enforces the use of bases that result in nearly sparse matrices, e.g., biorthogonal wavelet bases with a sufficient number of vanishing moments. Then, most entries in these matrices are close to zero and can be replaced by zero without destroying the order of approximation (compression) [11, 13, 15, 43, 47].

Let us recall the definition of the tensor-product of two separable Hilbert-spaces H with associated bilinear form $a(\cdot,\cdot)$ and $\hat{H}$ with bilinear form $\hat{a}(\cdot,\cdot)$, see for example [49]. Let $\{e_j\}_{j=1}^m$, $\{\hat{e}_i\}_{i=1}^{\hat{m}}$ be complete orthonormal systems in H and $\hat{H}$. Then $\{e_j \otimes \hat{e}_i\}$ is a complete orthonormal system in

$$H \otimes \hat{H} := \left\{ \sum_{j,i} \gamma_{i,j} \;\; e_j \otimes \hat{e}_i : \sum_{j,i} \gamma_{i,j}^2 < \infty \right\} \tag{2.7}$$

with scalar product $a \otimes \hat{a}\left(\sum_{j,i} \gamma_{i,j}\, e_j \otimes \hat{e}_i, \sum_{k,\ell} \gamma'_{k,\ell} \;\; e_k \otimes \hat{e}_\ell\right) = \sum_{j,i} \gamma_{j,i}\gamma'_{j,i}$. We identify the tensor-product $H \otimes \hat{H}$ with a function space over the corresponding product domain via the mapping $f \otimes \hat{f} \mapsto f(x)\hat{f}(\hat{x})$. Hence, a basis in $H \otimes \hat{H}$ is given by $\{\psi_j(x) = e_{j_1}(x_1)\hat{e}_{j_2}(x_2) : 1 \le j \le (m, \hat{m})\}$. These definitions extend naturally to higher dimensions $n > 2$.

The finite element spaces considered here are tensor-products of univariate function spaces. Starting from a one-dimensional splitting $\mathcal{L}^2 = \bigoplus_{j\ge 0} S_j$ we assume that the complement spaces[5]

$$W_j = S_j \ominus S_{j-1} \tag{2.8}$$

of S_{j-1} in S_j are spanned by some $\mathcal{L}^2$-stable bases

$$W_j = \operatorname{span}\{\psi_{j,k}, k \in \tau_j\}, \tag{2.9}$$

where τ_j is some finite dimensional index set defined from the subdivision rate of successive refinement levels. Here we stick to dyadic refinement. Furthermore we assume that

$$\left\| \sum_{k\in\tau_j} C_k \psi_{j,k} \right\|_{\mathcal{L}^2} \approx \left\| \{C_k\}_k \right\|_{\ell^2(k\in\tau_j)} \quad j \in \mathbb{N}_0 \tag{2.10}$$

where as usual $\| \sum_{k\in\tau_j} C_k \psi_{j,k}\|_{\mathcal{L}^2}$ denotes the norm induced from the scalar product on $\mathcal{L}^2$ and $\|\{C_k\}_k\|^2_{\ell^2(\tau_j)} = \sum_{k\in\tau_j} |C_k|^2$.

Let there be given a biorthogonal system $\{\tilde{\psi}_{j,k}, k \in \tau_j, j \in \mathbb{N}_0\}$, i.e.,

$$\langle \psi_{j,k}, \tilde{\psi}_{j',k'} \rangle = \delta_{j,j'}\delta_{k,k'}, \;\; j, j' \in \mathbb{N}_0, k \in \tau_j, k' \in \tau_{j'}. \tag{2.11}$$

Assuming that $\{\psi_{j,k}, k \in \tau_j, j \in \mathbb{N}_0\}$ forms a Riesz basis in $\mathcal{L}^2$, i.e.,

$$\left\| \sum_{j\in\mathbb{N}_0, k\in\tau_j} C_{j,k}\psi_{j,k} \right\|_{\mathcal{L}^2} \approx \left\| \{C_{j,k}\}_{j,k} \right\|_{\ell^2(j\in\mathbb{N}_0, k\in\tau_j)}, \tag{2.12}$$

every $u \in H$ has a unique expansion

$$u = \sum_{j=0}^{\infty} \sum_{k\in\tau_j} \langle u, \tilde{\psi}_{j,k}\rangle \psi_{j,k} = \sum_{j=0}^{\infty} \sum_{k\in\tau_j} \langle u, \psi_{j,k}\rangle \tilde{\psi}_{j,k} \tag{2.13}$$

[5]Here, $S_{-1} := \{\}$. Furthermore, in the periodic setting on the torus, we have $S_0 = \operatorname{span}\{\text{const}\}$ and for homogeneous boundary conditions we even have $S_0 = \{\}$. To later allow for the case of the unit n-square and non-periodic, non-homogeneous boundary conditions, we start counting with $j = 0$ here. We will switch to start counting with $j = 1$ when we derive specific estimates for the periodic homogeneous case.

and the biorthogonal system also forms a Riesz basis in $\mathcal{L}^2$.
Let us recall the notion of vanishing moments. In one dimension, a function ψ is said to have vanishing moments of order N if

$$\int_{\mathbb{R}} x^r \psi(x) dx = 0 \quad (0 \le r \le N-1). \tag{2.14}$$

Note that due to the biorthogonality of the basis functions (i.e., due to (2.13)) the number of vanishing moments N of the biorthogonal basis $\{\tilde{\psi}_{j,k}\}$ is exactly the order of polynomial reproduction of the wavelet basis $\{\psi_{j,k}\}$ and vice versa. It is well-known [13, 43] that the number of vanishing moments governs the compression capacity of a wavelet and that the order of polynomial reproduction governs the approximation power. Estimates of the order of approximation are mainly based on the local $\mathcal{L}^2$-stability (2.3) and an inequality of Jackson type, which in turn depends on estimates of the coefficients $\langle u, \tilde{\psi}_{j,k}\rangle$, i.e., on a moment condition for the dual wavelet. For purposes of compression, one usually assumes specific decay properties of the Schwarz kernel of the pseudodifferential operator under consideration. Then estimates of the size of the entries $a(\psi_{j,k}, \psi_{l,m})$ of the Galerkin stiffness matrix are obtained by expansions of the Schwarz kernel in a polynomial basis together with the cancellation properties of the primal wavelets $\psi_{j,k}$ [10, 15].
One of the merits of biorthogonal wavelets is that the number of vanishing moments can be chosen independently of the order of polynomial exactness. We will see later on that it is the number of vanishing moments of the dual wavelets $\tilde{\psi}_{j,k}$ that governs the form of the resulting optimized grids, if we pose specific assumptions on the solution of the variational problem.
Let

$$S = \bigcup_{i=0}^{\infty} S_i \quad \text{and} \quad \tilde{S} = \bigcup_{i=0}^{\infty} \tilde{S}_i \quad \text{with } \tilde{S}_i := \bigcup_{j=0}^{i} \{\tilde{\psi}_{j,k}, k \in \tau_j\}.$$

Moreover, we assume that the $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$ are scaled and dilated versions of single scale functions (mother wavelets) ψ_0 and $\tilde{\psi}_0$, i.e.,

$$\psi_{j,k}(x) = 2^{j/2}\psi_0\Big(\frac{x-k}{2^{-j}}\Big) \quad \text{and} \quad \tilde{\psi}_{j,k}(x) = 2^{j/2}\tilde{\psi}_0\Big(\frac{x-k}{2^{-j}}\Big). \tag{2.15}$$

We assume the following conditions to hold: First, we need a direct estimate (estimate of Jackson type, approximation order m)

$$\inf_{u_j \in S_j} \|u - u_j\|_{\mathcal{L}^2} \le C 2^{-jm} \|u\|_{\mathcal{H}^m} \ \ \forall u \in \mathcal{H}^m \tag{2.16}$$

for some positive integer m, and second, we need an inverse estimate (Bernstein inequality)

$$\|u_j\|_{\mathcal{H}^r} \le C 2^{jq} \|u_j\|_{\mathcal{L}^2} \ \ \forall u_j \in S_j \tag{2.17}$$

for $q < r$ with $r \in (0, m]$. We also assume that similar relations hold for the dual system $\tilde{S}$ with parameters $\tilde{m}$ and $\tilde{r}$. Then the validity of the following norm equivalences can be inferred from (2.16) and (2.17), see [9, 38]:[6]

$$\|u\|^2_{\mathcal{H}^t} \approx \sum_{j=0}^{\infty} \|w_j\|^2_{\mathcal{H}^t} \approx \sum_{j=0}^{\infty} 2^{2tj} \|w_j\|^2_{\mathcal{L}^2} \tag{2.18}$$

[6]Here, for $t < 0$, the $\mathcal{L}^2$-convergence has to be replaced by distributional convergence.

for $t \in (-\tilde{r}, r)$ where $u = \sum_{j=0}^{\infty} w_j, w_j \in W_j$. Note that (2.18) with $t = 0$ together with the local stability (2.10) enforces the global stability

$$\|u\|_{\mathcal{L}^2} \approx \|\{\langle u, \tilde{\psi}_{j,k}\rangle\}_{j,k}\|_{\ell^2(j\in\mathbb{N}_0, k\in\tau_j)},$$

i.e., condition (2.12) holds. The two-sided estimate (2.18) allows us to characterize smoothness properties of a function from the properties of a multiscale decomposition. This estimate is a consequence of approximation theory in Sobolev spaces together with interpolation and duality arguments [9, 38]. Moreover, it states that bilinear forms $a(\cdot,\cdot)$ satisfying the two-sided estimate (2.3) are spectrally equivalent to the sum of the bilinear forms $2^{2sj}(\cdot,\cdot)_{\mathcal{L}^2}$ on $W_j \times W_j$ induced from the right hand side of (2.18). A similar result holds for the analogous construction using the dual wavelets instead of the primal ones. This leads to the range $t \in (-r, \tilde{r})$. See [10] for an overview over multiscale methods dealing with biorthogonal wavelets.
For the higher-dimensional case $n > 1$, let $\mathbf{j} \in \mathbb{N}_0^n, \mathbf{j} \equiv (j_1, \ldots, j_n)$, be given, and consider the tensor-product partition with uniform step size 2^{-j_i} into the i-th coordinate direction. By $W_{\mathbf{j}}$ we denote the corresponding function space of tensor-products of one dimensional function spaces, i.e.,

$$W_{\mathbf{j}} := W_{j_1} \otimes \cdots \otimes W_{j_n}.$$

A basis of $W_{\mathbf{j}}$ is given by

$$\bigcup_{\mathbf{k}\in\tau_{\mathbf{j}}} \{\psi_{\mathbf{j},\mathbf{k}}(\mathbf{x}) = \psi_{j_1,k_1}(x_1) \cdot \cdots \cdot \psi_{j_n,k_n}(x_n)\}$$

with $\tau_{\mathbf{j}} \equiv (\tau_{j_1}, \ldots, \tau_{j_n})$.
Given an index set[7] $I_J \subset \mathbb{N}^n, J \in \mathbb{N}$, we consider the approximation spaces

$$V_J := \sum_{\mathbf{j}\in I_J} W_{\mathbf{j}}. \tag{2.19}$$

Here, J is the maximal level in V_J, i.e., $j_i \le J$, for $i = 1, \ldots, n$ and for all $\mathbf{j} \in I_J$. Associated with rectangular index sets $I_J^{-\infty} := \{|\mathbf{j}|_\infty \le J\}$ are the full grid spaces

$$V_J^{-\infty} := \bigoplus_{|\mathbf{j}|_\infty \le J} W_{\mathbf{j}}, \quad J > 0. \tag{2.20}$$

The so-called sparse grid space

$$V_J^0 := \bigoplus_{|\mathbf{j}|_1 \le J+n-1} W_{\mathbf{j}}, \quad J > 0 \tag{2.21}$$

is associated with the index set $I_J^0 := \{|\mathbf{j}|_1 \le J+n-1\}$. The approximation spaces $V_J^{-\infty}$ and V_J^0 will later turn out to be special choices of a family of approximation spaces V_J^T that are adapted to Sobolev spaces. Specifically, V_J^0 will turn out to be the appropriate choice for $\mathcal{H}^0$. See Figure 1 for the index sets of the full grid space $V_3^{-\infty}$ and the sparse grid space V_3^0 in the two-dimensional case. The dimensions of $W_{\mathbf{j}}$, $V_J^{-\infty}$ and V_J^0 are (note that $I_J \subset \mathbb{N}^n$, that is we count only interior grid points)

$$|W_{\mathbf{j}}| = 2^{|\mathbf{j}|_1 - n}, \tag{2.22}$$

$$|V_J^{-\infty}| = (2^J - 1)^n = O(2^{Jn}) \tag{2.23}$$

[7]We now restrict ourselves to the case of homogeneous boundary conditions and consider only indices $\mathbf{j}$ with $j_i > 0$, $i = 1, \ldots, n$.

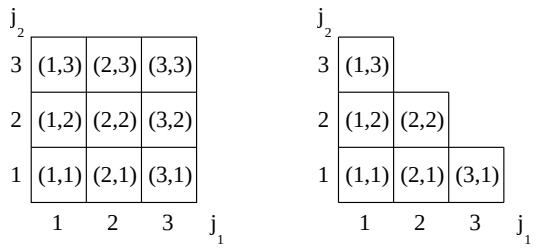

FIGURE 1. Index sets of the full grid space $V_3^{-\infty}$ (left) and of the sparse grid space V_3^0 (right), case $n = 2$.

and

$$|V_J^0| = 2^J \left(\frac{J^{n-1}}{(n-1)!} + O(J^{n-2}) \right), \tag{2.24}$$

see [3, 5, 6, 21, 25]. The estimates of $|W_{\mathbf{j}}|$ and $|V_J^{-\infty}|$ are clear. The estimate of $|V_J^0|$ is straightforward and will follow as a byproduct of the estimate of the dimensions of the spaces from a more general class of spaces in Section 4.2.
In this paper we introduce index sets that are optimized with respect to Sobolev norms and spaces with specific bounded mixed derivatives. To this end, we consider smoothness assumptions on the solution u of the variational problem or on the right hand side f (that in turn leads to smoothness assumptions on u). This leads us to the definition of more general spaces $\mathcal{H}^{t,l}_{\text{mix}}$ than the standard Sobolev spaces $\mathcal{H}^t$. They are defined as follows:

Definition 2.1. Let $t \in \mathbb{R}, l \in \mathbb{R}_0^+$. Furthermore, denote $\mathbf{1} = (1, \dots, 1)$ and let $\mathbf{e_i} = (0, \dots, 0, 1, 0, \dots, 0)$ be the i-th unit-vector in $\mathbb{R}^n$. Then

$$\mathcal{H}^{t,l}_{\text{mix}}(T^n) := \mathcal{H}^{t\mathbf{1}+l\mathbf{e_1}}_{\text{mix}}(T^n) \cap \dots \cap \mathcal{H}^{t\mathbf{1}+l\mathbf{e_n}}_{\text{mix}}(T^n), \tag{2.25}$$

where

$$\mathcal{H}^{\mathbf{k}}_{\text{mix}}(T^n) := \mathcal{H}^{k_1}(T) \otimes \dots \otimes \mathcal{H}^{k_n}(T).$$

For $l < 0$, $\mathcal{H}^{t,l}_{\text{mix}}$ is defined as the dual of $\mathcal{H}^{-t,-l}_{\text{mix}}$, i.e. we set $\mathcal{H}^{t,l}_{\text{mix}} := (\mathcal{H}^{-t,-l}_{\text{mix}})'$.

Furthermore we write

$$\mathcal{H}^{t}_{\text{mix}}(T^n) := \mathcal{H}^{t}(T) \otimes \dots \otimes \mathcal{H}^{t}(T), \quad t \geq 0. \tag{2.26}$$

These are spaces of dominating mixed derivative. For $\mathbf{t} \in \mathbb{N}^n$ the space $\mathcal{H}^{\mathbf{t}}_{\text{mix}}$ possesses the equivalent norm

$$\|u\|^2_{\mathcal{H}^{\mathbf{t}}_{\text{mix}}} \approx \sum_{\mathbf{0} \leq \mathbf{k} \leq \mathbf{t}} \|u^{(\mathbf{k})}\|^2_{\mathcal{L}^2}. \tag{2.27}$$

Here, $u^{(\mathbf{k})}$ is the generalized mixed derivative $\frac{\partial^{|\mathbf{k}|_1}}{\partial^{k_1} \dots \partial^{k_n}} u$. For example, $u^{(t,\dots,t)}$ is the $n \cdot t$-th order mixed derivative and describes the additional smoothness requirements for the space $\mathcal{H}^t_{\text{mix}}$ compared to the larger isotropic Sobolev space $\mathcal{H}^t$.
Note that the relations

$$\mathcal{H}^t_{\text{mix}} \subset \mathcal{H}^t \subset \mathcal{H}^{t/n}_{\text{mix}} \text{ for } t \geq 0 \text{ and } \mathcal{H}^{t/n}_{\text{mix}} \subset \mathcal{H}^t \subset \mathcal{H}^t_{\text{mix}} \text{ for } t \leq 0 \tag{2.28}$$

hold. See [42] for problems connected with the spaces H^t_{mix} and further references. In the general case, with $s \in \mathbb{R}$, the relation

$$\mathcal{H}^{t,l}_{\text{mix}} \subset \mathcal{H}^s \tag{2.29}$$

holds for either $s \leq t + l, t \geq 0$ or $s \leq nt + l, t \leq 0$.
The spaces $\mathcal{H}^{t,l}_{\text{mix}}$ with $l \geq 0, t \geq 0$ are special cases of the spaces

$$\mathcal{H}^{(\mathbf{t}^1,\ldots,\mathbf{t}^n)}_{\text{mix},\cap}(T^n) := \mathcal{H}^{\mathbf{t}^1}_{\text{mix}}(T^n) \cap \cdots \cap \mathcal{H}^{\mathbf{t}^n}_{\text{mix}}(T^n), \tag{2.30}$$

where $\mathbf{t}^i \in \mathbb{R}^n$ and $\mathbf{t}^i \geq 0$, where $1 \leq i \leq n$. On the other hand the standard Sobolev spaces $\mathcal{H}^t(T^n)$, as well as the spaces $\mathcal{H}^t_{\text{mix}}(T^n)$ with dominating mixed derivative, are special cases of the spaces $\mathcal{H}^{t,l}_{\text{mix}}(T^n)$. We have

$$\mathcal{H}^t(T^n) = \mathcal{H}^{0,t}_{\text{mix}}(T^n) \text{ and } \mathcal{H}^t_{\text{mix}}(T^n) = \mathcal{H}^{t,0}_{\text{mix}}(T^n). \tag{2.31}$$

Indeed, for $t \in \mathbb{R}^+_0$ we have the representation

$$\begin{aligned} \mathcal{H}^{0,t}_{\text{mix}}(T^n) &= \mathcal{H}^{(t,0,\ldots,0)}_{\text{mix}}(T^n) \cap \cdots \cap \mathcal{H}^{(0,\ldots,0,t)}_{\text{mix}}(T^n) \\ &= \mathcal{H}^{t\mathbf{e_1}}_{\text{mix}}(T^n) \cap \cdots \cap \mathcal{H}^{t\mathbf{e_n}}_{\text{mix}}(T^n) \\ &= \mathcal{H}^t(T^n) \end{aligned} \tag{2.32}$$

where

$$\mathcal{H}^{(0,\cdots,0,1,0,\cdots,0)}_{\text{mix}}(T^n) := \mathcal{L}^2(T) \otimes \cdots \otimes \mathcal{L}^2(T) \otimes \mathcal{H}^t(T) \otimes \mathcal{L}^2(T) \otimes \cdots \otimes \mathcal{L}^2(T).$$

To show the last equality in (2.32), choose an orthogonal basis of $\mathcal{H}^t(T)$ and use the definition of the tensor-product via orthonormal systems (2.7). More precisely, using periodic continuation on $\mathbb{R}$ and the fact that $\{\sin(n(2x\pi - \pi))\}$ defines a complete orthonormal system in $\mathcal{L}^2(T)$ and $\mathcal{H}^t(T)$, it is clear that every $u \in \mathcal{H}^t(T^n)$ can be represented as a Fourier sine series and (2.32) follows directly from the definition of the tensor product (2.7) and the definition of intersection of Hilbert spaces. Note that similar results hold for problems with Dirichlet or Neumann boundary conditions and certain cases of mixed boundary conditions. See [27] for more details and some examples. The rightmost equation in (2.31) is clear from the definition of $\mathcal{H}^t_{\text{mix}}(T^n)$ in (2.26). A norm on $\mathcal{H}^{t,l}_{\text{mix}}(T^n)$ can be defined directly via

$$\|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}} \approx \sum_{1 \leq i \leq n} \|u\|^2_{\mathcal{H}^{t\mathbf{1}+l\mathbf{e_i}}_{\text{mix}}}.$$

Hence, the spaces $\mathcal{H}^{t,l}_{\text{mix}}$ from (2.25) give a unified framework for the study of the special cases $\mathcal{H}^t = \mathcal{H}^{0,t}_{\text{mix}}$ and $\mathcal{H}^t_{\text{mix}} = \mathcal{H}^{t,0}_{\text{mix}}$.

3. Norm equivalences

To get norm equivalences analogous to (2.18) in $n \geq 2$ dimensions, we use the representation of $\mathcal{H}^t$ and $\mathcal{H}^t_{\text{mix}}$ above as tensor-products of 1D spaces and intersections. We use the notation $\{V; a\}$ to denote a Hilbert space V equipped with the scalar product $a(\cdot,\cdot)$. Consider a collection of Hilbert spaces H_l, where $1 \leq l \leq n$ for some $n \in \mathbb{N}$ and a collection of closed subspaces $V_{l,i} \subset H_l$ such that topologically $H_l = \sum_i V_{l,i}$. An additive subspace splitting $\{H_l; a_l\} = \sum_i \{V_{l,i}; b_{l,i}\}$ is called stable

if the norm equivalence $a_l(u,u) \approx \|u\|^2 := \inf_{u_i \in H_{l,i}: u=\sum_i u_i} (b_{l,i}(u_i,u_i))$ holds true, i.e., if the characteristic numbers

$$\lambda_{\min,l} = \min_{0\neq u\in H_l} \frac{a_l(u,u)}{\|u\|^2},\ \lambda_{\max,l} = \max_{0\neq u\in H_l} \frac{a_l(u,u)}{\|u\|^2},\ \kappa_l = \frac{\lambda_{\max,l}}{\lambda_{\min,l}}$$

are finite and positive. We have the following two Propositions.

Proposition 3.1. *If the splittings*

$$\{H_l; a_l\} = \sum_i \{V_{l,i}; b_{l,i}\}, \qquad l \in \{1,\dots,n\}, \qquad n \in \mathbb{N}$$

are stable and possess the condition numbers κ_l then the tensor-product splitting

$$\{H_1 \otimes H_2 \otimes \cdots \otimes H_n; a_1 \otimes \cdots \otimes a_n\} = \sum_{i_1} \cdots \sum_{i_n} \{V_{1,i_1} \otimes \cdots \otimes V_{n,i_n}; b_{1,i_1} \otimes \cdots \otimes b_{n,i_n}\}$$

is also stable and possesses the condition number $\prod_{l=1}^n \kappa_l$.

See [27] for a proof in the case $n = 2$. The extension to the n-dimensional case is straightforward.

Proposition 3.2. *Let there be given sequences $\{\alpha_{l,i}\}_i, l = 1,\dots,n, n \in \mathbb{N}$. Suppose that the splittings*

$$\{H_l; a_l\} = \sum_i \{V_i; \alpha_{l,i} b\}, \qquad l = 1,\dots,n,$$

are stable and that the sums are direct. Then, for all $\alpha_l > 0, l = 1,\dots,n$, the splitting

$$\{H_1 \cap \cdots \cap H_n; \alpha_1 a_1 + \cdots + \alpha_n a_n\} = \sum_i \{V_i; (\alpha_1\alpha_{1,i} + \cdots + \alpha_n \alpha_{n,i}) b\} \tag{3.1}$$

is stable with condition number

$$\kappa \le \frac{\max\{\lambda_{\max,1},\dots,\lambda_{\max,n}\}}{\min\{\lambda_{\min,1},\dots,\lambda_{\min,n}\}}.$$

Proof. See [27] for a proof in the case $n = 2$. The n-dimensional case is analogous. □

Combining the representation (2.32) of $\mathcal{H}^t(T^n)$, where $t \ge 0$, with these Propositions and the stability result (2.18) in one dimension we come up with the following norm equivalence and stable splitting of $\mathcal{H}^t(T^n)$.

Theorem 3.3. *Let $u \in \mathcal{H}^t(T^n), u = \sum_{\mathbf{j}} w_{\mathbf{j}}, w_{\mathbf{j}} \in W_{\mathbf{j}}$ (for $t < 0$ with distributional convergence) and let the assumptions* (2.16) *and* (2.17) *on the validity of a Jackson and a Bernstein inequality for the primal as well as the dual system hold. Then*

$$\|u\|^2_{\mathcal{H}^t} \approx \sum_{\mathbf{j}} 2^{2t|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \text{ for } t \in (-\tilde{r}, r), \text{ where } |\mathbf{j}|_\infty = \max_{1\le i\le n} j_i. \tag{3.2}$$

Proof. In the one-dimensional case (2.18) yields

$$\|u\|^2_{\mathcal{H}^t(T)} \approx \sum_{j=0}^{\infty} 2^{2tj} \|w_j\|^2_{\mathcal{L}^2(T)}, \quad 0 \le t < r,$$

$$u = \sum_{j=0}^{\infty} w_j, \quad w_j \in W_j, u \in \mathcal{H}^t(T)$$

and from (2.12)

$$\begin{aligned} \|u\|_{\mathcal{L}^2} &\approx \|\{\langle u, \tilde{\psi}_{j,l}\rangle\}_{j,l}\|_{\ell^2(j\in\mathbb{N}_0, l\in\tau_j)}, \\ u &= \sum_{j=0}^{\infty} w_j, \quad w_j \in W_j, \quad u \in \mathcal{L}^2(T). \end{aligned}$$

This shows the stability of the one-dimensional splittings

$$\{\mathcal{H}^t(T); \|\cdot\|^2_{\mathcal{H}^t(T)}\} = \sum_j \{W_j; 2^{2tj}\|\cdot\|^2_{\mathcal{L}^2(T)}\}$$

and

$$\{\mathcal{L}^2(T); \|\cdot\|^2_{\mathcal{L}^2(T)}\} = \sum_j \{W_j; \|\cdot\|^2_{\mathcal{L}^2(T)}\}.$$

From Proposition 3.1 we obtain the stability of the splittings

$$\begin{aligned} &\left\{\mathcal{H}^{(0,\dots,0,t,0,\dots,0)}_{\text{mix}}; (\cdot,\cdot)_{\mathcal{L}^2} \otimes \cdots \otimes (\cdot,\cdot)_{\mathcal{L}^2} \otimes a(\cdot,\cdot) \otimes (\cdot,\cdot)_{\mathcal{L}^2} \otimes \cdots \otimes (\cdot,\cdot)_{\mathcal{L}^2}\right\} \\ &= \sum_{\mathbf{j}} \left\{W_{j_1} \otimes \cdots \otimes W_{j_n}; 2^{2tj_i}(\cdot,\cdot)_{\mathcal{L}^2} \otimes \cdots \otimes (\cdot,\cdot)_{\mathcal{L}^2}\right\}. \end{aligned}$$

Now we represent $\mathcal{H}^t(T^n)$ as in (2.32) and we apply Proposition 3.2. Then we obtain the stability of the splitting

$$\{\mathcal{H}^t(T^n); \|\cdot\|^2_{\mathcal{H}^t(T^n)}\} = \sum_{\mathbf{j}} \left\{W_j; \left(\sum_{i=1}^n 2^{2tj_i}\right)\|\cdot\|^2_{\mathcal{L}^2(T^n)}\right\}$$

for nonnegative $t < r$. Because of $2^{2t|\mathbf{j}|_\infty} \le \sum_{i=1}^n 2^{2tj_i} \le n2^{2t|\mathbf{j}|_\infty}$ for $t \ge 0$ we then have (3.2) for positive t. To obtain the validity of (3.2) for $-\tilde{r} < t < 0$ note that the same reasoning as above applied to the representation of u in the dual wavelet system shows that we have a similar result for the spaces spanned by the dual wavelets for $0 \le t < \tilde{r}$. By the duality $(\mathcal{H}^t)' = \mathcal{H}^{-t}$, the assertion follows then for the range $-\tilde{r} < t < 0$ and hence for the whole range $t \in (-\tilde{r}, r)$. □

For the space H^t_{mix} the following norm equivalence holds:

Theorem 3.4. *Let $u \in \mathcal{H}^t_{\text{mix}}$ have the representation $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, where $w_{\mathbf{j}} \in W_{\mathbf{j}}$, and let the assumptions* (2.16) *and* (2.17) *on the validity of a Jackson and a Bernstein inequality for the primal as well as the dual system hold. Then*

$$\|u\|^2_{\mathcal{H}^t_{\text{mix}}} \approx \sum_{\mathbf{j}} 2^{2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \text{ for } t \in (-\tilde{r}, r). \tag{3.3}$$

Proof. The two-sided estimate (3.3) is a direct consequence of Proposition 3.1 and the definition of the space $\mathcal{H}^t_{\text{mix}}$ as tensor-product of one-dimensional Hilbert spaces. Again we use the stable one-dimensional splittings

$$\{\mathcal{H}^t(T); \|\cdot\|^2_{\mathcal{H}^t(T)}\} = \sum_j \{W_j; 2^{2tj}\|\cdot\|^2_{\mathcal{L}^2(T)}\}$$

$$\{\mathcal{L}^2(T); \|\cdot\|^2_{\mathcal{L}^2(T)}\} = \sum_j \{W_j; \|\cdot\|^2_{\mathcal{L}^2(T)}\}$$

(which we get from (2.18) and (2.12)) and Proposition 3.1 to obtain the stability of the splitting

$$\begin{aligned}
&\{\mathcal{H}^t_{\text{mix}}; a(\cdot,\cdot)\otimes\cdots\otimes a(\cdot,\cdot)\}\\
&\quad=\{\mathcal{H}^t(T)\otimes\cdots\otimes\mathcal{H}^t(T); a(\cdot,\cdot)\otimes\cdots\otimes a(\cdot,\cdot)\}\\
&\quad=\sum_{\mathbf{j}}\{W_{j_1}\otimes\cdots\otimes W_{j_n}; 2^{2tj_1}(\cdot,\cdot)_{\mathcal{L}^2}\otimes\cdots\otimes 2^{2tj_n}(\cdot,\cdot)_{\mathcal{L}^2}\}\\
&\quad=\sum_{\mathbf{j}}\{W_{j_1}\otimes\cdots\otimes W_{j_n}; 2^{2t|\mathbf{j}|_1}(\cdot,\cdot)_{\mathcal{L}^2}\otimes\cdots\otimes(\cdot,\cdot)_{\mathcal{L}^2}\}.
\end{aligned}$$

This shows (3.3). □

Note that, under the assumptions of the Theorems 3.3 and 3.4, we have similar relations for the subspace splittings induced by the dual wavelets. There, r must be replaced by $\tilde{r}$ and vice versa.

Remark 3.5. The norm equivalences in Theorems 3.3 and 3.4 are special cases of norm equivalences for the spaces $\mathcal{H}^{(\mathbf{t}^1,\ldots,\mathbf{t}^n)}_{\text{mix},\cap}$ from (2.30). Again using Propositions 3.1 and 3.2 it is straightforward to show that

$$\|u\|^2_{\mathcal{H}^{(\mathbf{t}^1,\ldots,\mathbf{t}^n)}_{\text{mix},\cap}}\approx\sum_{\mathbf{j}}\left(\sum_{i=1}^n 2^{2\langle\mathbf{t}^i,\,\mathbf{j}\rangle}\right)\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}\ \text{ for } \mathbf{t}^i\geq 0, -\tilde{r}<\mathbf{t}^i<r. \tag{3.4}$$

Furthermore, with straightforward duality arguments, we have the norm equivalency

$$\|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}}\approx\sum_{\mathbf{j}}\left(\sum_{i=1}^n 2^{2t|\mathbf{j}|_1+2lj_i}\right)\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}\approx\sum_{\mathbf{j}}2^{2t|\mathbf{j}|_1+2l|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \tag{3.5}$$

for the spaces $\mathcal{H}^{t,l}_{\text{mix}}$, where $0\leq t<r$ and $0\leq t+l<r$. Compared to (3.2) and (3.3), the additional factors $2^{2t|\mathbf{j}|_1}$ or $2^{2l|\mathbf{j}|_\infty}$ in (3.5) reflect the different smoothness requirements. Note that for $t=0$ or $l=0$ we regain (3.2) from Theorem 3.3 and (3.3) from Theorem 3.4, respectively. Analogous relations hold for the dual spaces.

Remark 3.6. For the construction of optimized approximation spaces, we will use the upper estimate from (3.2) and the lower estimate in (3.3) and (3.5).

Remark 3.7. One of the merits of the norm equivalences (3.2), (3.3) or the more general one (3.5) is the fact that they lead directly to optimal preconditioning. For example, if one chooses the scaled system $\{2^{-s|\mathbf{l}|_\infty}\psi_{\mathbf{l},\mathbf{k}} : |\mathbf{l}|_\infty\leq J, \mathbf{k}\in\tau_{\mathbf{l}}\}$ as the basis in the finite element approximation space $V_J^{-\infty}$, then the spectral condition numbers $\kappa(A_J)$ of the discretization matrices $A_J=\{2^{-s|\mathbf{l}+\mathbf{l}'|_\infty}a(\psi_{\mathbf{l},\mathbf{k}},\psi_{\mathbf{l}',\mathbf{k}'})\}_{\mathbf{l},\mathbf{l}',\mathbf{k},\mathbf{k}'}$ are bounded uniformly in J, i.e.,

$$\kappa(A_J)=O(1), \tag{3.6}$$

see [12, 31]. This leads to fast iterative methods with convergence rates independent of the number of unknowns of the approximation space. Note that this result can be trivially extended to the case of discretization matrices built from arbitrary collections of scaled basis functions.

4. Optimized approximation spaces for Sobolev spaces

Suppose a symmetric elliptic problem (2.2) and its variational formulation

$$a(u_{\mathrm{FE}}, v) = (f, v) \ \forall v \in V_{\mathrm{FE}} \tag{4.1}$$

on a finite element approximation space $V_{\mathrm{FE}} \subset \mathcal{H}^s$ are given. Using the $\mathcal{H}^s$-ellipticity condition (2.3) and Cea's Lemma, we have the estimate

$$\sqrt{a(u - u_{\mathrm{FE}}, u - u_{\mathrm{FE}})} \approx \|u - u_{\mathrm{FE}}\|_{\mathcal{H}^s} \approx \inf_{v \in V_{\mathrm{FE}}} \|u - v\|_{\mathcal{H}^s}$$

for the error $\sqrt{a(u - u_{\mathrm{FE}}, u - u_{\mathrm{FE}})}$ between the solution u of the continuous problem (2.2) and the solution u_{FE} of the approximate problem (4.1) measured in the energy norm. In this section we give bounds on the term

$$\inf_{v \in V_{\mathrm{FE}}} \|u - v\|_{\mathcal{H}^s}$$

for various choices of the approximation space V_{FE}, under the constraint

$$u \in \mathcal{H}^{t,l}_{\mathrm{mix}}, \text{ where } 0 \le t < r \text{ and } -\tilde{r} < s < t + l < r.$$

We define grids and associated approximation spaces that are adapted to the parameter s and to the constraint on the smoothness of the solution and give estimates on their dimension and the order of approximation. The definition of the grids is motivated by the results of Section 3, specifically on the norm equivalence (3.5) and the special cases (3.2) and (3.3). We are particularly interested in constructing approximation spaces that break the curse of dimensionality, that is whose dimensions are at most polynomially dependent on n.

4.1. Approximation spaces for problems with constraint on the solution.

We first deal with the cases $u \in \mathcal{H}^t$ and $u \in \mathcal{H}^t_{\mathrm{mix}}$. More general cases will be discussed at the end of Section 4.1.2, see Theorem 4.1. In this section let $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, where $w_{\mathbf{j}} \in W_{\mathbf{j}}$. Furthermore let $-\tilde{r} < s < t < r$. Then $\mathcal{H}^t \subset \mathcal{H}^s$. For notational convenience we restrict ourselves to the case $t \ge 0$. Note that the case $t < 0$ could be covered with analogous reasoning.

4.1.1. *Estimates on the order of approximation for the spaces $V_J^{-\infty}$ and V_J^0.* First, we consider the order of approximation for the full grid case. Let $u \in \mathcal{H}^s$. Applying the norm equivalence (3.2) gives us

$$\begin{aligned}
\inf_{v \in V_J^{-\infty}} \|u - v\|^2_{\mathcal{H}^s} &\le \|u - \sum_{|\mathbf{j}|_\infty \le J} w_{\mathbf{j}}\|^2_{\mathcal{H}^s} \overset{(3.2)}{\approx} \sum_{|\mathbf{j}|_\infty > J} 2^{2s|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \\
&= \sum_{|\mathbf{j}|_\infty > J} 2^{2(s-t)|\mathbf{j}|_\infty} 2^{2t|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \\
&\le \max_{|\mathbf{j}|_\infty > J} 2^{2(s-t)|\mathbf{j}|_\infty} \sum_{|\mathbf{j}|_\infty > J} 2^{2t|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}. \qquad (4.2)
\end{aligned}$$

To continue, we assume additional smoothness of the solution, i.e., $u \in \mathcal{H}^t$. Then we can apply (3.2) once more, now with $u \in \mathcal{H}^t$. This yields

$$\begin{aligned}
\max_{|\mathbf{j}|_\infty > J} 2^{2(s-t)|\mathbf{j}|_\infty} \sum_{|\mathbf{j}|_\infty > J} 2^{2t|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} &\overset{(3.2)}{\le} C \cdot \max_{|\mathbf{j}|_\infty > J} 2^{2(s-t)|\mathbf{j}|_\infty} \|u\|^2_{\mathcal{H}^t} \\
&\le C \cdot 2^{2(s-t)(J+1)} \|u\|^2_{\mathcal{H}^t}. \qquad (4.3)
\end{aligned}$$

Altogether we have the standard error estimate

$$\inf_{v\in V_J^{-\infty}} \|u-v\|^2_{\mathcal{H}^s} \le C\cdot 2^{2(s-t)}2^{2(s-t)J}\|u\|^2_{\mathcal{H}^t} \text{ for } u\in\mathcal{H}^t \text{ and } -\tilde{r}<s<t<r. \tag{4.4}$$

From the exponent on the right hand side we get $O(t-s)$ as order of approximation. It is easy to see that the order of approximation does not change when $u\in\mathcal{H}^t_{\text{mix}}\subset\mathcal{H}^t$, i.e.,[8]

$$\inf_{v\in V_J^{-\infty}} \|u-v\|^2_{\mathcal{H}^s} \le C\cdot 2^{2(s-t)}2^{2(s-t)J}\|u\|^2_{\mathcal{H}^t_{\text{mix}}} \text{ for } -\tilde{r}<s<t<r.$$

Note that we are implicitly using several times the vanishing moment condition of the dual wavelets, which is implicitly contained in the Jackson inequality (2.16). Changing from the full grid space $V_J^{-\infty}$ to the approximation space V_J^0 changes the situation significantly. Applying again the norm equivalence (3.2), we have for $u\in\mathcal{H}^s$

$$\begin{aligned}\inf_{v\in V_J^0}\|u-v\|^2_{\mathcal{H}^s} &\le \|u-\sum_{|\mathbf{j}|_1\le J+n-1} w_{\mathbf{j}}\|^2_{\mathcal{H}^s} \overset{(3.2)}{\approx} \sum_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}\\ &= \sum_{|\mathbf{j}|_1>J+n-1} 2^{2(s-t)|\mathbf{j}|_\infty}2^{2t|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}\\ &\le \max_{|\mathbf{j}|_1>J+n-1} 2^{2(s-t)|\mathbf{j}|_\infty}\sum_{|\mathbf{j}|_1>J+n-1} 2^{2t|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}.\end{aligned}$$

Now we again require u to be of higher regularity, i.e., $u\in\mathcal{H}^t$. This yields

$$\begin{aligned}&\max_{|\mathbf{j}|_1>J+n-1} 2^{2(s-t)|\mathbf{j}|_\infty}\sum_{|\mathbf{j}|_1>J+n-1} 2^{2t|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2}\\ &\overset{(3.2)}{\le} C\cdot\max_{|\mathbf{j}|_1>J+n-1} 2^{2(s-t)|\mathbf{j}|_\infty}\|u\|^2_{\mathcal{H}^t} \le C\cdot 2^{2(s-t)(J+n-1)/n}\|u\|^2_{\mathcal{H}^t}\end{aligned}$$

where we used in the penultimate step the fact that the maximum is obtained for $|\mathbf{j}|_\infty=\lceil(J+n-1)/n\rceil$. Altogether we have for $u\in\mathcal{H}^t$ and $-\tilde{r}<s<t<r$

$$\inf_{v\in V_J^0}\|u-v\|^2_{\mathcal{H}^s} \le C\cdot 2^{2(s-t)(1-1/n)}2^{2(s-t)J/n}\|u\|^2_{\mathcal{H}^t}. \tag{4.5}$$

Compared to the result for the full grid approximation space, the order of approximation deteriorates from $O(t-s)$ to $O((t-s)/n)$.
However, for the smaller space $\mathcal{H}^t_{\text{mix}}\subset\mathcal{H}^t, t\ge 0$, and operators of positive order, i.e., $s\ge 0$, no loss in the order of approximation occurs, if the full grid space is replaced by the space V_J^0. This is because we can apply norm equivalence (3.3) instead of (3.2).[9] We apply (3.2) for functions from $\mathcal{H}^s$ and (3.3) for $u\in\mathcal{H}^t_{\text{mix}}$ and

[8] For $t<0$ we would have to assume $u\in\mathcal{H}^t_{\text{mix}}\cap\mathcal{H}^s$ here.

[9] Remember the different exponents in the terms $2^{2t|\mathbf{j}|_\infty}$ and $2^{2t|\mathbf{j}|_1}$ in (3.2) and (3.3), respectively.

get

$$
\begin{aligned}
\inf_{v\in V_J^0}\|u-v\|_{\mathcal{H}^s}^2 &\le \|u-\sum_{|\mathbf{j}|_1\le J+n-1} w_{\mathbf{j}}\|_{\mathcal{H}^s}^2 \overset{(3.2)}{\approx} \sum_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \\
&= \sum_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}2^{2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \\
&\le \max_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}\sum_{|\mathbf{j}|_1>n+J-1} 2^{2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \qquad (4.6)\\
&\overset{(3.3)}{\le} C\cdot \max_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}\|u\|_{\mathcal{H}^t_{\text{mix}}}^2 \\
&\le C\cdot 2^{2s(J+1)-2t(J+n)}\|u\|_{\mathcal{H}^t_{\text{mix}}}^2 \text{ for } u\in\mathcal{H}^t_{\text{mix}},
\end{aligned}
$$

where we used in the last step the fact that the term $2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}$ takes its maximum in $(J+1,1,\dots,1)$. Altogether we have for $u\in\mathcal{H}^t_{\text{mix}}$ and $-\tilde{r}<s<t<r$ with $t\ge 0$ (compare (2.29) for $l=0$) that

$$
\inf_{v\in V_J^0}\|u-v\|_{\mathcal{H}^s}^2 \le C\cdot 2^{2(s-tn)}2^{2(s-t)J}\|u\|_{\mathcal{H}^t_{\text{mix}}}^2.
$$

That is, there appears no loss in the order of approximation compared to the result for the full grid approximation space.

For operators of negative order, i.e., $s<0$, the situation is different. Here, compared to the estimate (4.4) for $u\in\mathcal{H}^t$, the order of approximation improves when changing to the space $\mathcal{H}^t_{\text{mix}}$, but in contrast to the case $s\ge 0$, the optimal order of convergence cannot be attained. Applying (3.2) for functions from $\mathcal{H}^s$ and (3.3) we find that if $u\in\mathcal{H}^t_{\text{mix}}$, then

$$
\begin{aligned}
\inf_{v\in V_J^0}\|u-v\|_{\mathcal{H}^s}^2 &\le \|u-\sum_{|\mathbf{j}|_\infty>J+n-1} w_{\mathbf{j}}\|_{\mathcal{H}^s}^2 \overset{(3.2)}{\approx} \sum_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \\
&= \sum_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}2^{2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \\
&\le \max_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}\sum_{|\mathbf{j}|_1>n+J-1} 2^{2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \qquad (4.7)\\
&\overset{(3.3)}{\le} C\cdot \max_{|\mathbf{j}|_1>J+n-1} 2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}\|u\|_{\mathcal{H}^t_{\text{mix}}}^2 \\
&\le C\cdot 2^{2s(1-1/n)-2tn}2^{2(s/n-t)J}\|u\|_{\mathcal{H}^t_{\text{mix}}}^2,
\end{aligned}
$$

where we used in the last step that $2^{2s|\mathbf{j}|_\infty-2t|\mathbf{j}|_1}$ takes its maximum for $|\mathbf{j}|_\infty=\lceil (J+n-1)/n\rceil$ and $|\mathbf{j}|_1=J+n$ for $s<0$. That is, although the order of approximation is improved when changing from $\mathcal{H}^t$ to $\mathcal{H}^t_{\text{mix}}$ there still appears a loss in the order of approximation of $s(1/n-1)$ compared to the full grid. This fact has been described already in [28] for the case $-1\le s<0$ and prewavelets (i.e., wavelets that are $\mathcal{L}^2$-orthogonal between different subspaces $W_{\mathbf{j}}$ and build a Riesz basis in the subspaces $W_{\mathbf{j}}$). In summary we have that for operators with $s\ge 0$ the order of approximation is kept for $u\in\mathcal{H}^t_{\text{mix}}$, $s<t$, when changing from the approximation space $V_J^{-\infty}$ to the sparse grid space V_J^0. For operators of negative order a slight deterioration of the order of approximation appears.

4.1.2. *Definition and order of approximation of the approximation spaces V_J^T.* In the following we construct approximation spaces for functions from $\mathcal{H}_{\text{mix}}^{t,l}, t > 0$, $-\tilde{r} < s < t + l < r$, and operators of positive or negative order by carefully selecting subspaces of the full grid space. These subspaces are chosen so that the order of approximation of the full grid space is kept. The sparse grid space V_J^0 and the full grid space $V_J^{-\infty}$ are special cases. We start with the space $\mathcal{H}_{\text{mix}}^t = \mathcal{H}_{\text{mix}}^{t,0}$. The inequality

$$\max_{\mathbf{j}\notin I_J^0} 2^{2s|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \|u\|^2_{\mathcal{H}^t_{\text{mix}}} \leq C \cdot 2^{2(s-t)J} \|u\|^2_{\mathcal{H}^t_{\text{mix}}} \text{ for } u \in \mathcal{H}^t_{\text{mix}}, 0 \leq s < t,$$

from (4.6) reveals that for $s \geq 0$ one could discard indices from the index set I_J^0 without destroying the optimal order of approximation. Consider an index set $I_J \subset I_J^0$ such that

$$\max_{\mathbf{j}\notin I_J} 2^{2s|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \leq C \cdot 2^{2(s-t)J} \tag{4.8}$$

where $C \neq C(s, t, J)$. Then the order of approximation is kept for the approximation space defined from the index set I_J. Taking logarithms on both sides of (4.8) and dividing by $2t$ (remember that we have $t > 0$) shows that (4.8) is equivalent to

$$\mathbf{j} \in I_J \Leftrightarrow -|\mathbf{j}|_1 + \frac{s}{t}|\mathbf{j}|_\infty \geq -J + \frac{s}{t}J + c, \tag{4.9}$$

where $c \neq c(\mathbf{j}, J)$ is essentially the logarithm of the constant C on the right hand side of the asymptotic estimate (4.8). For operators of negative order we deduce from (4.7) that we have to add indices to the index set I_J^0 to keep the optimal order of approximation. Again, the order is kept if I_J is such that (4.8) and hence (4.9) holds.

Therefore we define the optimized grid as the minimal index set for which (4.9) holds. Fixing $(J, 1, \cdots, 1)$ to be the index with maximal $|\cdot|_\infty$-norm to be included into the index sets leads to $c = n - 1$ and the index sets

$$I_J^{s/t} := \left\{ \mathbf{j} \in \mathbb{N}^n : -|\mathbf{j}|_1 + \frac{s}{t}|\mathbf{j}|_\infty \geq -(n + J - 1) + \frac{s}{t}J \right\}.$$

These index sets depend on the parameter J and on the quotient s/t.

To give the results more flexibility we parametrize the index sets with a new parameter T and finally get

$$I_J^T := \left\{ \mathbf{j} \in \mathbb{N}^n : -|\mathbf{j}|_1 + T|\mathbf{j}|_\infty \geq -(n + J - 1) + TJ \right\} \tag{4.10}$$

with the related approximation spaces

$$V_J^T := \bigoplus_{\mathbf{j}\in I_J^T} W_{\mathbf{j}} = \bigoplus_{-|\mathbf{j}|_1 + T|\mathbf{j}|_\infty \geq -(n+J-1)+TJ} W_{\mathbf{j}}.$$

The new parameter T allows us to decouple the definition of the index sets and the resulting grids from the smoothness parameters s and t. This parameter T also allows us to more closely investigate into the relation between smoothness assumptions, the choice of approximation space and the order of approximation. In the following we will consider terms like

$$\inf_{v\in V_J^T} \|u - v\|^2_{\mathcal{H}^s}$$

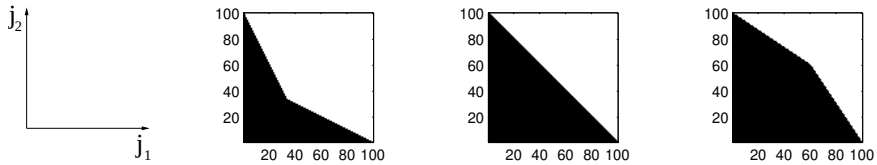

FIGURE 2. Index sets I_J^T for $T > 0$, $T = 0$ and $T < 0$.

for varying T, where we assume again that $u \in \mathcal{H}^t$ or $u \in \mathcal{H}^t_{\text{mix}}$. Definition (4.10) ensures that the optimal order of approximation is kept for $T \le s/t$ and functions from $\mathcal{H}^t_{\text{mix}}$ (compare (4.8) and (4.9)). For $T > s/t$ the order of approximation deteriorates. We discuss this point in more detail below.
Note that for $T = 0$ we have $V_J^T = V_J^0$ and for $T \to -\infty$ we have $V_J^T \to V_J^{-\infty}$, i.e., the full grid space. Furthermore we have the natural restriction to $T < 1$. Obviously the inclusions

$$V_J^1 \subset V_J^{T_1} \subset V_J^{T_2} \subset V_J^0 \subset V_J^{T_3} \subset V_J^{T_4} \subset V_J^{-\infty} \text{ for } T_4 \le T_3 \le 0 \le T_2 \le T_1 < 1$$

hold. Schematically the behavior of the index sets I_J^T is depicted in Figure 2 with varying T for the two-dimensional case. Figures 3–6 show some examples for the two-dimensional case.
We now discuss the dependence of the order of approximation of the approximation space V_J^T on the parameter T in more detail. Let us first consider the case $u \in \mathcal{H}^t$. Remember that $\mathcal{H}^t \subset \mathcal{H}^s$. Similarly to (4.5), we have

$$\begin{aligned}
\inf_{v \in V_J^T} \|u - v\|^2_{\mathcal{H}^s} &\le \|u - \sum_{\mathbf{j} \in I_J^T} w_{\mathbf{j}}\|^2_{\mathcal{H}^s} \overset{(3.2)}{\le} C \cdot \max_{\mathbf{j} \notin I_J^T} 2^{2(s-t)|\mathbf{j}|_\infty} \|u\|^2_{\mathcal{H}^t} \\
&\overset{(4.10)}{=} C \cdot \max_{T|\mathbf{j}|_\infty - |\mathbf{j}|_1 < TJ-(n+J-1)} 2^{2(s-t)|\mathbf{j}|_\infty} \|u\|^2_{\mathcal{H}^t} \qquad (4.11) \\
&= C \cdot 2^{2(s-t)((1-T)J-n+1)/(n-T)} \|u\|^2_{\mathcal{H}^t} \\
&= C \cdot 2^{2(s-t)(1-n)/(n-T)} 2^{2(s-t)((1-T)/(1-T/n))J/n} \|u\|^2_{\mathcal{H}^t}.
\end{aligned}$$

Here we used the fact that $\max_{T|\mathbf{j}|_\infty - |\mathbf{j}|_1 < TJ-(n+J-1)} 2^{2(s-t)|\mathbf{j}|_\infty}$ takes its maximum at $|\mathbf{j}|_\infty = \lceil ((1-T)J - n + 1)/(n-T) \rceil$. If we compare (4.11) for the space V_J^T with $T > 0$ to the result (4.5) for the space V_J^0, we see that the order of approximation deteriorates by the factor $(1-T)/(1-T/n)$. For $T < 0$ the order of approximation is improved by the factor $(1-T)/(1-T/n)$. If we compare (4.11) for the space V_J^T with $T < 0$ to the result (4.4) for the full grid space $V_J^{-\infty}$, we see that the order of approximation deteriorates by the factor $(1-T)/(n-T)$. Note that for $T = 0$ we regain the same approximation order as for estimate (4.5).
For $u \in \mathcal{H}^t_{\text{mix}}$ we have (compare (4.8) and remember that $\mathcal{H}^t_{\text{mix}} \subset \mathcal{H}^s$)

$$\begin{aligned}
\inf_{v \in V_J^T} \|u - v\|^2_{\mathcal{H}^s} &\le \|u - \sum_{\mathbf{j} \in I_J^T} w_{\mathbf{j}}\|^2_{\mathcal{H}^s} \overset{(3.2),(3.3)}{\le} C \cdot \max_{\mathbf{j} \notin I_J^T} 2^{2s|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \|u\|^2_{\mathcal{H}^t_{\text{mix}}} \\
&= C \cdot \max_{T|\mathbf{j}|_\infty - |\mathbf{j}|_1 < TJ-(n+J-1)} 2^{2s|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \|u\|^2_{\mathcal{H}^t_{\text{mix}}}. \qquad (4.12)
\end{aligned}$$

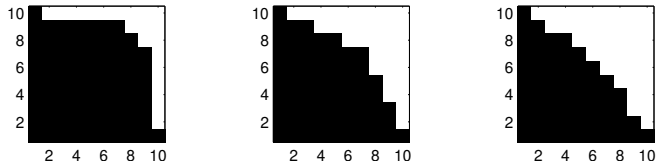

FIGURE 3. Index sets I_{10}^{-5}, I_{10}^{-1} and $I_{10}^{-1/2}$, from left to right.

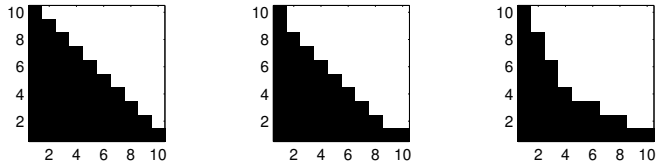

FIGURE 4. Index sets $I_{10}^{0}, I_{10}^{1/8}$ and $I_{10}^{1/2}$, from left to right.

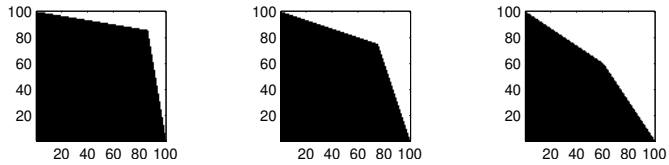

FIGURE 5. Index sets $I_{100}^{-5}, I_{100}^{-1}$ and $I_{100}^{-1/2}$, from left to right.

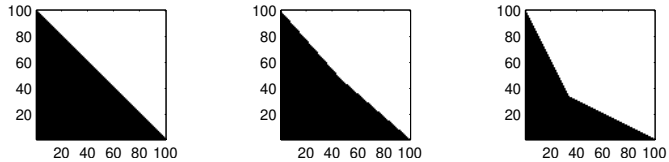

FIGURE 6. Index sets $I_{100}^{0}, I_{100}^{1/8}$ and $I_{100}^{1/2}$, from left to right.

It is straightforward to show that for $T \geq s/t$ the maximum is obtained for $|\mathbf{j}|_\infty = \lceil ((1-T)J - n + 1)/(n-T) \rceil$, and for $T \leq s/t$ the maximum is obtained in $\mathbf{j} = (J+1, 1, \cdots, 1)$.
We continue (4.12) and have for $T \geq s/t$

$$\begin{aligned}\inf_{v \in V_J^T} \|u - v\|_{\mathcal{H}^s}^2 &\leq C \cdot 2^{2(s-nt)((1-T)J-n+1)/(n-T)} \|u\|_{\mathcal{H}^t_{\text{mix}}}^2 \\ &= C \cdot 2^{2(s-nt)(1-n)/(n-T)} 2^{2(s-t+(Tt-s)(n-1)/(n-T))J} \|u\|_{\mathcal{H}^t_{\text{mix}}}^2\end{aligned}$$

and for $T \leq s/t$, we have

$$\begin{aligned}\inf_{v \in V_J^T} \|u - v\|_{\mathcal{H}^s}^2 &\leq C \cdot 2^{-2t(n-1)} 2^{2(s-t)(J+1)} \|u\|_{\mathcal{H}^t_{\text{mix}}}^2 \\ &= C \cdot 2^{2(s-t)-2t(n-1)} 2^{2(s-t)J} \|u\|_{\mathcal{H}^t_{\text{mix}}}^2 .\end{aligned}$$

Note that for $T = s/t$ both estimates give the same approximation order.
This estimate shows once more that for $u \in \mathcal{H}^t_{\text{mix}}$ there appears no loss of asymptotic approximation power if the full grid is replaced by an optimized grid induced

by the index set I_J^T with $T \le s/t$. Note that I_J^T is of lower dimension than the index set $I_J^{-\infty}$ of the full grid. However, a further reduction of the number of grid points by using an index set I_J^T with $T > s/t$ results in a deterioration of the order of approximation. In this case the order of approximation is reduced by $(Tt-s)(n-1)/(n-T)$.
Note that smoothness assumptions on the right hand side f in the variational problem (2.2) imply smoothness properties of the solution. Consider for example the case of a differential operator. Then for example $f \in \mathcal{H}^t_{\text{mix}}$ implies $u \in \mathcal{H}^{t,s}_{\text{mix}}$. Therefore we now deal also with the more general case $u \in \mathcal{H}^{t,l}_{\text{mix}}$. We summarize the discussion in a Theorem.

Theorem 4.1. *Let $0 \le t < r$ and $-\tilde{r} < s < t+l < r$. Then for $u \in \mathcal{H}^{t,l}_{\text{mix}}$ we have*
(4.13)
$$\inf_{v\in V_J^T} \|u-v\|^2_{\mathcal{H}^s} \le \begin{cases} C\cdot 2^{2(s-l-t+(Tt-s+l)(n-1)/(n-T))J}\|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}} & \text{for } T \ge (s-l)/t \\ C\cdot 2^{2(s-l-t)J}\|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}} & \text{for } T \le (s-l)/t. \end{cases}$$

Specifically for $u \in \mathcal{H}^t = \mathcal{H}^{0,t}_{\text{mix}}$ we have

$$\inf_{v\in V_J^T} \|u-v\|^2_{\mathcal{H}^s} \le C\cdot 2^{2(s-t)((1-T)/(n-T))J}\|u\|^2_{\mathcal{H}^t} \tag{4.14}$$

and for $u \in \mathcal{H}^t_{\text{mix}} = \mathcal{H}^{t,0}_{\text{mix}}$ we have

$$\inf_{v\in V_J^T} \|u-v\|^2_{\mathcal{H}^s} \le \begin{cases} C\cdot 2^{2(s-t+(Tt-s)(n-1)/(n-T))J}\|u\|^2_{\mathcal{H}^t_{\text{mix}}} & \text{for } T \ge s/t \\ C\cdot 2^{2(s-t)J}\|u\|^2_{\mathcal{H}^t_{\text{mix}}} & \text{for } T \le s/t. \end{cases} \tag{4.15}$$

Proof. Let $u \in \mathcal{H}^{t,l}_{\text{mix}}$. To show (4.13) we use the upper estimate from the norm equivalence (3.2) and the lower estimate from (3.5). Then

$$\begin{aligned} \inf_{v\in V_J^T} \|u-v\|^2_{\mathcal{H}^s} &\le \|u - \sum_{\mathbf{j}\in I_J^T} w_{\mathbf{j}}\|^2_{\mathcal{H}^s} \overset{(3.2)}{\approx} \sum_{\mathbf{j}\notin I_J^T} 2^{2s|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \\ &\le \max_{\mathbf{j}\notin I_J^T} 2^{2(s-l)|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \cdot \sum_{\mathbf{j}\notin I_J^T} 2^{2l|\mathbf{j}|_\infty + 2t|\mathbf{j}|_1}\|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \\ &\overset{(3.5)}{\le} C\cdot \max_{\mathbf{j}\notin I_J^T} 2^{2(s-l)|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1}\|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}}. \end{aligned}$$

Evaluating the maximum with respect to I_J^T shows (4.13). The inequalities (4.14) and (4.15) are special cases of the inequality (4.13) with $t = 0$ and $l = 0$, respectively. □

Theorem 4.1 shows that the optimal order of approximation of a function in $\mathcal{H}^{t,l}_{\text{mix}}$ is kept when changing from the full grid approximation space $V_J^{-\infty}$ to an approximation space V_J^T with $T \le (s-l)/t$. The use of approximation spaces V_J^T with $T > (s-l)/t$ leads to a deterioration of the optimal order of convergence. Hence, for purposes of discretization of large scale problems with solution in the space $\mathcal{H}^{t,l}_{\text{mix}}$, the spaces $V_J^{(s-l)/t}$ with $T \le (s-l)/t$ are well-suited. From the nestedness of the spaces V_J^T we conclude that the choice $T = (s-l)/t$ will lead to the most economical algorithms. This holds true especially in higher dimensions where the benefits of the spaces V_J^T become most obvious, as we will see in Section 4.2.

4.2. **Dimension of the approximation spaces V_J^T.** The following Lemma discusses the dimension of the spaces V_J^T. In general, we may split the basis functions into two sets, one with those functions that correspond to the interior of the unit cube and the other with those functions that correspond to the boundary. For ease of exposition we again restrict ourselves to homogeneous boundary conditions, that is, we count only those basis functions/indices that correspond to the interior of the unit cube. Hence $\mathbf{j} \in \mathbb{N}^n$ and the index $\mathbf{j}$ with minimal $|\cdot|_\infty$ and $|\cdot|_1$-norm in an index set I_J^T is $\mathbf{j} = (1, \dots, 1)$. Note that other boundary conditions could be dealt with analogous reasoning, but we would then have to count also indices $\mathbf{j}$ with $j_i = 0$ for some $1 \le i \le n$.

Lemma 4.2. *It holds*

$$\dim\left(V_J^T\right) \quad \le \quad \begin{cases} \frac{n}{2}\left(1/(1-2^{-1/(1/T-1)})\right)^n \cdot 2^J = 0(2^J) & \textit{for } 0 < T < 1, \\ O(2^{((T-1)/(T/n-1))J}) & \textit{for } T < 0. \end{cases} \tag{4.16}$$

The case $T = 0$ is covered by the estimate

$$\dim\left(V_J^T\right) \le \left(\frac{J^{n-1}}{(n-1)!} + O(J^{n-2})\right) \cdot 2^J = O(2^J J^{n-1}) \ \textit{for } 0 \le T \le 1/J. \tag{4.17}$$

Proof. The case $T \ge 0$:
Let $|\mathbf{j}|_1 = n + J - 1 - i$ and $0 < T \le 1$. Then

$$W_{\mathbf{j}} \subset V_J^T \Leftrightarrow -|\mathbf{j}|_1 + T|\mathbf{j}|_\infty \ge -(n+J-1) + TJ \Leftrightarrow |\mathbf{j}|_\infty \ge J - \frac{1}{T}i.$$

Since

$$\sum_{|\mathbf{j}|_1 = n+J-1-i} 1 = \binom{|\mathbf{j}|_1 - 1}{n-1}$$

and

$$\sum_{\substack{|\mathbf{j}|_1 = n+J-1-i, \\ |\mathbf{j}|_\infty \ge J - i/T}} 1 \le \binom{\lceil |\mathbf{j}|_1 - (J - i/T) \rceil}{n-1} = \binom{\lceil n - 1 + (1/T - 1)i \rceil}{n-1}, \tag{4.18}$$

the number of subspaces $W_{\mathbf{j}}$ with $|\mathbf{j}|_1 = n + J - 1 - i$ belonging to V_J^T is bounded by

$$n\binom{\lceil n - 1 + (1/T - 1)i \rceil}{n-1}.$$

Hence, with the definition of V_J^T, we have

$$\begin{aligned} |V_J^T| \quad &= \quad \sum_{i=0}^{J-1} \sum_{\substack{|\mathbf{j}|_1 = n+J-1-i, \\ |\mathbf{j}|_\infty \ge J - i/T}} |W_j| \le \sum_{i=0}^{J-1} 2^{J-1-i} n \binom{\lceil n - 1 + (1/T - 1)i \rceil}{n-1} \\ &= \quad 2^{J-1} n \sum_{i=0}^{J-1} 2^{-i} \binom{\lceil n - 1 + (1/T - 1)i \rceil}{n-1}. \end{aligned} \tag{4.19}$$

For $T < 1$ the substitution $i \to i/(1/T - 1)$ leads to

$$|V_J^T| \le 2^{J-1} n \sum_{i=0}^{\lceil (1/T-1)(J-1) \rceil} 2^{-i/(1/T-1)} \binom{n-1+i}{i}.$$

Since $(x^{n-1+i})^{(n-1)} = ((n-1+i)!)x^i/i!\ \ \forall x \in \mathbb{R}$, we get

$$
\begin{aligned}
|V_J^T| &\le 2^{J-1} n \ \frac{1}{(n-1)!} \sum_{i=0}^{\lceil (1/T-1)(J-1)\rceil} (x^{n-1+i})^{(n-1)} \Big|_{x=2^{-1/(1/T-1)}} \\
&= 2^{J-1} n \frac{1}{(n-1)!} \left(x^{n-1} \sum_{i=0}^{\lceil (1/T-1)(J-1)\rceil} x^i \right)^{(n-1)} \Big|_{x=2^{-1/(1/T-1)}} \\
&= 2^{J-1} n \ \frac{1}{(n-1)!} \left(x^{n-1} \ \frac{1-x^{\lceil (1/T-1)(J-1)\rceil+1}}{1-x} \right)^{(n-1)} \Big|_{x=2^{-1/(1/T-1)}} \\
&= 2^{J-1} n \frac{1}{(n-1)!} \left[\left(\frac{x^{n-1}}{1-x} \right)^{(n-1)} - \left(\frac{x^{\lceil (1/T-1)(J-1)\rceil+n}}{1-x} \right)^{(n-1)} \right] \Big|_{x=2^{-1/(1/T-1)}}.
\end{aligned}
$$

Since

$$
\begin{aligned}
\frac{1}{(n-1)!} \left(x^k \ \frac{1}{1-x} \right)^{(n-1)} &= \frac{1}{(n-1)!} \sum_{i=0}^{n-1} \binom{n-1}{i} (x^k)^{(i)} \left(\frac{1}{1-x} \right)^{(n-i-1)} \\
&= \frac{1}{(n-1)!} \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{k!}{(k-i)!} x^{k-i} (n-1-i)! \ \left(\frac{1}{1-x} \right)^{n-i} \\
&= x^{k-n} \sum_{i=0}^{n-1} \binom{k}{i} \left(\frac{x}{1-x} \right)^{n-i}, \qquad (4.20)
\end{aligned}
$$

we get

$$
\begin{aligned}
|V_J^T| \ &\le \ 2^{J-1} n \ \left[\left(\frac{1}{1-x} \right)^n - x^{\lceil (1/T-1)(J-1)\rceil+1} \times \right. \\
&\qquad \left. \sum_{i=0}^{n-1} \binom{\lceil (1/T-1)(J-1)\rceil + n}{i} \left(\frac{x}{1-x} \right)^{n-i} \right] \Big|_{x=2^{-1/(1/T-1)}} \\
&\le \ 2^{J-1} n \ \left(\frac{1}{1-2^{-1/(1/T-1)}} \right)^n.
\end{aligned}
$$

Hence we obtain (4.16).

To prove (4.17) we again let $|\mathbf{j}|_1 = n+J-1-i$ and $T \le 1/J$. Then

$$
W_{\mathbf{j}} \subset V_J^T \Leftrightarrow -|\mathbf{j}|_1 + T|\mathbf{j}|_\infty \ge -(n+J-1) + TJ \Leftrightarrow |\mathbf{j}|_\infty \ge 0.
$$

That is, every $W_{\mathbf{j}}$ with $|\mathbf{j}|_1 \le n+J-1-i$ is in V_J^T. Hence

$$
\begin{aligned}
\left|V_J^T\right| \ &= \ \sum_{|\mathbf{j}|_1 \le n+J-1} |W_j| = \sum_{i=0}^{J-1} 2^{J-1-i} \sum_{|\mathbf{j}|_1 = n+J-1-i} 1 \\
&= \ \sum_{i=0}^{J-1} 2^{J-1-i} \binom{n-1+J-1-i}{n-1} = \sum_{i=0}^{J-1} 2^i \binom{n-1+i}{n-1}.
\end{aligned}
$$

This results in (see [5], proof of Lemma 7 for details)

$$
\left|V_J^T\right| = (-1)^n + 2^J \sum_{i=0}^{n-1} \binom{n+J-1}{i} (-2)^{n-1-i} = \left(\frac{J^{n-1}}{(n-1)!} + O(J^{n-2}) \right) \cdot 2^J.
$$

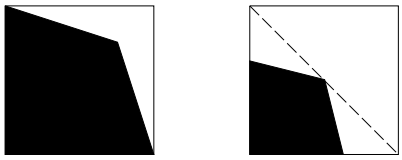

FIGURE 7. Schematic representation of I_J^T (left) and $I_{\widehat{J}}^T$ (right) for $T < 0$, $n = 2$.

This completes the proof for the case $T \geq 0$.

The case $T < 0$:
Now we deal with the approximation spaces V_J^T, $T < 0$. We introduce an auxiliary index set $I_{\widehat{J}}^T$ with $I_{\widehat{J}}^T \subset I_J^0$ given by

$$I_{\widehat{J}}^T = \left\{ \mathbf{j} : -|\mathbf{j}|_1 + T|\mathbf{j}|_\infty \geq -(n + J - 1) + TJ/n \right\}$$

and the related approximation spaces $V_{\widehat{J}}^T := \bigoplus_{\mathbf{j} \in I_{\widehat{J}}^T} W_{\mathbf{j}}$. Note that $I_{\widehat{J}}^T$ is just a shifted version of I_J^T. See Figure 7 for a schematic comparison of the index sets I_J^T and $I_{\widehat{J}}^T$ in the case $n = 2$.
Obviously $\dim(I_{\widehat{J}}^T) = O(2^J)$. Equation (4.7) shows that if $T \leq s/t$, then the order of approximation of the space $V_{\widehat{J}}^T$ for functions from $\mathcal{H}_{\text{mix}}^t$ is the same as for the space V_J^0, which is $O(2^{(s/n-t)J})$. On the other hand (4.15) implies that the order of approximation of the space $V_{\widehat{J}}^{s/t}$ is $O(2^{(s-t)\widehat{J}})$. This shows that $O(2^{(s/n-t)J}) = O(2^{(s-t)\widehat{J}})$ must hold. Hence we have that $J = ((s-t)/(s/n-t))\widehat{J} + C$ and therefore $\dim(V_{\widehat{J}}^{s/t}) = O(2^{((s-t)/(s/n-t))\widehat{J}})$ and $\dim(V_J^{s/t}) = O(2^{((s-t)/(s/n-t))J})$.
This completes the proof. □

Note that the coefficient in the asymptotic estimate of the second inequality in (4.16) is unbounded for $T \to 0$ whereas the coefficient in the estimate (4.17) remains bounded. Asymptotically, for $T > 0$, the estimate (4.16) is sharper than (4.17). However, for computationally relevant sizes of J, estimate (4.17) might be sharper than (4.16) for T near 0. Similar results have been obtained in [5] for $s = 1, t = 2, l = 0$, and approximation spaces spanned by piecewise linear functions.
The estimates (4.16) and (4.17) should be compared to the results for the full grid spaces $V_J^{-\infty}$ with dimension $\dim(V_J^{-\infty}) = (2^J - 1)^n$. The first two estimates in (4.16) show that for $T > 0$ the dependence of the dimension of the approximation space on the dimension n of the problem has been reduced from 2^{nJ} to $nC^n \cdot 2^J$, with some constant C independent of n and J. Note that C is explicitly given by Lemma 4.2 for this case. For the case $T < 0$ we see that using the spaces V_J^T in the Galerkin method leads to a significant reduction of the number of unknowns, and hence the number of entries in the stiffness matrices. Note that $\dim(V_J^T) \ll \dim(V_J^{-\infty})$ for large n or large T. Hence, using the spaces given above in the Galerkin method leads to a significant reduction of the numbers of unknowns and hence the number of entries in the stiffness matrices.
In summary, Theorem 4.1 and Lemma 4.2 show that for approximation problems with $u \in \mathcal{H}_{\text{mix}}^{t,l}$ the use of the approximation spaces V_J^T with $T \leq (s-l)/t$ leads to a

significant reduction of the number of degrees of freedom compared to the full grid, while the order of approximation remains the same as for the full grid. This will become even more clear in Section 5 where we consider the overall cost of solving the operator equations up to a prescribed tolerance.

4.3. **Optimization procedures and subspace selection.** In this section we present another way of obtaining the approximation spaces V_J^T. The idea is to explicitly use an optimization procedure to select subspaces. We describe this briefly in the following. See [5] for a longer discussion in the case of $s = 1$ where we use basis functions of piecewise linear splines. Further details can be found in [6, 21].

As we already noticed several times, the norm equivalence (3.2) and the ellipticity condition (2.3) together with the local stability (2.10) the two-sided estimate for $s \in (-\tilde{r}, r)$ yields that

$$a(u,u) \overset{(2.3)}{\approx} \|u\|_{\mathcal{H}^s}^2 \overset{(3.2)}{\approx} \sum_{\mathbf{j}} 2^{2s|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \overset{(2.10)}{\approx} \sum_{\mathbf{j}} 2^{2s|\mathbf{j}|_\infty} \Big(\sum_{m \in \tau_{\mathbf{j}}} \langle u, \tilde{\psi}_{\mathbf{j},\mathbf{m}} \rangle^2 \Big).$$

From this we see that the contribution of the subspace $W_{\mathbf{j}}$ to $a(u,u)$ is bounded by $\mathrm{Profit}_{\mathbf{j}} \cdot C$, where

$$\mathrm{Profit}_{\mathbf{j}} := 2^{2s|\mathbf{j}|_\infty} \|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2 \approx 2^{2s|\mathbf{j}|_\infty} \sum_{\mathbf{m} \in \tau_{\mathbf{j}}} \langle u, \tilde{\psi}_{\mathbf{j},\mathbf{m}} \rangle^2. \tag{4.21}$$

Together with an upper estimate of $\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2$ or of the coefficients $\langle u, \tilde{\psi}_{\mathbf{j},\mathbf{m}} \rangle$, the resulting upper estimate of $\mathrm{Profit}_{\mathbf{j}}$ can be considered a measure of how the approximation power improves when $W_{\mathbf{j}}$ is included into the approximation space. Note that such an estimate of $\|w_{\mathbf{j}}\|_{\mathcal{L}^2}^2$ or an upper estimate of $\langle u, \tilde{\psi}_{\mathbf{j},\mathbf{m}} \rangle$ by $\|u\|_{\mathcal{H}_{\mathrm{mix}}^{t,l}}^2$ can be obtained easily for elements of the considered smoothness classes by exploiting the vanishing moment condition on the dual wavelets $\tilde{\psi}_{\mathbf{j},\mathbf{m}}$ (compare the Jackson inequality). Implicitly we used this several times in the last sections.

On the other hand, the inclusion of $W_{\mathbf{j}}$ into the approximation space causes some cost in the discretization and hence in the solution procedure. The easiest measure for this cost is the dimension of the subspace $|W_{\mathbf{j}}|$. The task is now to find a grid (i.e., to select subspaces $W_{\mathbf{j}}$) such that a given error bound is minimal for some fixed cost, that is, the dimension of the approximation space is bounded by some given value b. This problem of deciding which subspaces should be included into the approximation space given some prescribed overall cost can be reformulated as a classical binary knapsack problem. Restricting the range of possible subspaces $W_{\mathbf{j}}$ to $|\mathbf{j}|_\infty \le J$ for some integer J, and arranging the possible indices $\mathbf{j}$ in some linear order, the optimization problem reads as follows:

Find a binary vector $y \in \{0,1\}^{n \times J}$ such that

$$\sum_{|\mathbf{j}|_\infty \le J} \mathrm{Profit}_{\mathbf{j}} \cdot y_{\mathbf{j}} \text{ constrained to } \sum_{|\mathbf{j}|_\infty \le J} |W_{\mathbf{j}}| \cdot y_{\mathbf{j}} \le b \tag{4.22}$$

is maximal.

Here the binary array y indicates which subspaces are to be included into the approximation space. Unfortunately such a binary knapsack problem is NP-hard. However, the situation changes, when we allow the array y to be a rational array in

$([0,1]\cap\mathbb{Q})^{n\times J}$. Then we know that the solution can be obtained by the following algorithm [36]:

(1) Arrange the possible indices in some linear order such that $\{\mathrm{Profit}_j/|W_j|\}_j$ is decreasing in size, that is
$$\frac{\mathrm{Profit}_{i_1}}{|W_{i_1}|}\geq\frac{\mathrm{Profit}_{i_2}}{|W_{i_2}|}\geq\cdots.$$
(2) Let $M:=\max\{i:\sum_{k=1}^{i}|W_{j_k}|\leq b\}$.
(3) The solution of the rational knapsack problem is given by
$$\begin{aligned}&y_1=\cdots=y_M=1;\\&y_{M+1}=\frac{b-\sum_{i=1}^{M}|W_{j_i}|}{|W_{j_{M+1}}|};\\&y_{M+2}=y_{M+3}=\cdots=0.\end{aligned}$$

Note that y_{M+1} may be rational in $[0,1]$. Therefore the solution of the rational knapsack problem is in general no solution of the binary knapsack problem. However, we still have the freedom of slightly changing the size of the cost b. We can do this in such a way that y_{M+1} is in $\{0,1\}$. Then, y is also a solution of the corresponding binary knapsack problem. We refer to [5, 6, 21] for more details of this optimization procedure. The optimization process can thus be reduced to the discussion (of the upper bounds) of the profit/cost quotients of the subspaces

$$\gamma_{\mathbf{j}}:=\frac{\mathrm{Profit}_{\mathbf{j}}}{|W_{\mathbf{j}}|},\tag{4.23}$$

that is, for an optimal grid in this sense one has to include those $W_{\mathbf{j}}$ into the approximation space with $\gamma_{\mathbf{j}}$ bigger than some threshold. Note that the optimization has to be performed with the use of upper bounds for $\mathrm{Profit}_{\mathbf{j}}$ and not with the exact (but unknown) values. Hence, the optimization procedure is optimal only in this sense. Combining (4.21), (4.23) and using the moment condition on the dual wavelets together with the smoothness assumptions on the solution, we end up with spaces similar to V_J^T as in Section 4.1.2 with basically the same properties.

5. Cost estimates

In this section we deal with the cost of solving the elliptic variational problem (2.2) up to some prescribed error when using the approximation spaces V_J^T and preconditioners arising from the norm equivalences from Section 3, compare Remark 3.7. We consider the worst case setting, that is the error of an approximation u_{FE} from a finite element approximation space V_{FE} compared to the exact solution u is measured in the $\mathcal{H}^s$-norm. The cost of computing an approximation to the solution of the variational problem (2.2) can be divided into two parts, namely the cost of obtaining the discrete system (2.6) and the cost of computing an approximate solution to this discrete system. The prices for these two parts are often called the informational and combinatorial cost, respectively.

Note that due to the larger supports of the wavelets from coarser scales, the resulting stiffness matrices $\mathcal{A}_J$ are rather densely populated. Here we have to distinguish two cases, namely integral and differential operators. In the case of integral operators, $\mathcal{A}_J$ is dense and thus has $O(\dim(V_{\mathrm{FE}})^2)$ entries. In the case of differential operators

$\mathcal{A}_J$ has $O\left(\dim(V_{\mathrm{FE}})\left(\log\left(\dim(V_{\mathrm{FE}})\right)\right)^n\right)$ entries and is therefore much sparser than in the case of integral operators.

Let us take a closer look at the case of integral operators first. There are techniques to estimate the size of the entries in the stiffness matrix a priori and to avoid the computation of entries below a prescribed threshold [13, 43]. See [28] for numerical experiments regarding compression with respect to the single layer potential equation and approximation spaces built with the index sets I_J^0. Here we refrain from incorporating the effect of additional compression into the overall cost complexity, as not to mix the effects of the use of the approximation spaces V_J^T and of compression. Moreover, note that additional compression provides us with purely asymptotic estimates only, whereas the choice of optimized approximation spaces pays for computationally relevant problem sizes especially in higher dimensions.

For differential operators it is important to note that one needs not to assemble the stiffness matrix, because all that is required in an iterative scheme is the application of the preconditioned stiffness matrix to a vector. Exploiting the pyramid structure of the multiscale transformations and the tensor-product structure of our wavelet basis functions, the matrix-vector product can be performed with $O(\dim(V_{\mathrm{FE}}))$ operations for differential operators with constant or separable coefficients. The same holds true in the case of general coefficient functions on uniform grids, i.e., for the approximation space $V_J^{-\infty}$. However, note that computing a matrix-vector product with linear cost is a very involved and delicate task.

In the following, we assume that the matrix-vector product can be performed with $O(\dim(V_{\mathrm{FE}}))$ operations for differential operators and with $O(\dim(V_{\mathrm{FE}})^2)$ operations for integral operators. We furthermore assume that arbitrary continuous linear information [48, 50] is permissible, i.e., that the stiffness matrix as well as the load vector have been computed exactly (or at least with sufficient accuracy). Once the stiffness matrix and the load vector have been computed, we are left with the issue of proposing an algorithm for the approximate solution of the discrete problem.

We discuss an algorithm whose cost is $O(\dim(V_{\mathrm{FE}}))$ for differential operators and $O(\dim(V_{\mathrm{FE}})^2)$ for integral operators. Concerning the computational cost, we mentioned already in Remark 3.7 that a simple diagonal scaling of the stiffness matrix is enough to obtain optimal preconditioning if the related norm equivalences hold. This allows us to construct solvers whose cost is of the order of the number of entries in the stiffness matrix. To be a bit more precise, let us estimate the cost to solve (4.1) up to discretization error ϵ, which is of order $O(2^{-cJ})$, with some $c > 0$ depending on the order of approximation of the wavelet basis. From (3.6) in Section 3 we have that the preconditioned (diagonally scaled) Galerkin stiffness matrix $\{2^{-s|\mathbf{l}+\mathbf{l}'|_\infty} a(\psi_{\mathbf{l},\mathbf{k}}, \psi_{\mathbf{l}',\mathbf{k}'})\}_{\mathbf{l},\mathbf{l}',\mathbf{k},\mathbf{k}'}$ has a condition number that is bounded independently of the number of levels involved. Hence, the convergence rate ξ of a gradient method is independent of the dimension of the finite element approximation space V_{FE} if the stiffness matrix is symmetric. Applied to the preconditioned system, the initial error is reduced by at least the factor ξ in every iteration step and the number of iterations necessary to obtain an approximation within the prescribed accuracy is then $|\log_\xi(\epsilon)| = cJ$. Hence, the overall cost of computing an approximation to the solution of the variational problem (2.2) within discretization accuracy ϵ is $O(J \cdot \dim(V_{\mathrm{FE}})^2)$ if the stiffness matrix is dense and $O(J \cdot \dim(V_{\mathrm{FE}}))$ if the matrix-vector product can be performed with $O(\dim(V_{\mathrm{FE}}))$ operations. Note

that it is possible to get rid of the J-term in the cost estimate by embedding the solver in a nested iteration scheme [35]. The idea is to compute a suitable starting value by first applying some iteration steps to the problem on a coarser level and to apply this procedure recursively starting from the coarsest level. This makes the optimized spaces defined in Section 4 good candidates for the approximation space V_{FE} provided that the required regularity assumptions on the solution of the variational problem hold.

To obtain an approximation of the exact solution that has an error of $O(\epsilon)$ in the energy norm, we must choose the number J of levels such that the approximation error is smaller than $O(\epsilon)$. Combining the results about the approximation error from Theorem 4.1 with the estimate of the dimension of the space $V_J^{(s-l)/t}$ in Section 4.2 gives us the order of the cost. Tables 1 and 2 summarize the discussion above. There the cost of solving the problem (2.2) in the space $\mathcal{H}^{t,l}_{\mathrm{mix}}$ up to an error of the order of ϵ is given for positive and for negative smoothness parameters s.

Table 1. Cost of solving an $\mathcal{H}^s$-elliptic variational problem with a differential operator up to an error of $O(\epsilon)$ measured in the $\mathcal{H}^s$-norm under the constraint that the solution is in $\mathcal{H}^{t,l}_{\mathrm{mix}}$.

	$V_J^{(s-l)/t}$	$V_J^{-\infty}$
$s > l$	$O\left(\epsilon^{1/(s-l-t)}\right)$	$O\left(\epsilon^{n/(s-l-t)}\right)$
$s = l$	$O\left(\epsilon^{-1/t}\left(\ln(\epsilon^{-1/t})\right)^{n-1}\right)$	$O\left(\epsilon^{-n/t}\right)$
$s < l$	$O\left(\epsilon^{1/((s-l)/n-t)}\right)$	$O\left(\epsilon^{n/(s-l-t)}\right)$

Table 2. Cost of solving an $\mathcal{H}^s$-elliptic variational problem with an integral operator up to an error of $O(\epsilon)$ measured in the $\mathcal{H}^s$-norm under the constraint that the solution is in $\mathcal{H}^{t,l}_{\mathrm{mix}}$.

	$V_J^{(s-l)/t}$	$V_J^{-\infty}$
$s > l$	$O\left(\epsilon^{2/(s-l-t)}\right)$	$O\left(\epsilon^{2n/(s-l-t)}\right)$
$s = l$	$O\left(\epsilon^{-2/t}\left(\ln(\epsilon^{-1/t})\right)^{2(n-1)}\right)$	$O\left(\epsilon^{-2n/t}\right)$
$s < l$	$O\left(\epsilon^{2/((s-l)/n-t)}\right)$	$O\left(\epsilon^{2n/(s-l-t)}\right)$

Tables 1 and 2 show that for problems with solution $u \in \mathcal{H}^{t,l}_{\mathrm{mix}}$ and optimized approximation spaces, the asymptotic cost is independent of the dimension n if $s - l > 0$. For fixed dimension n and $1 - (s-l)/t < n$, the cost is in favor of the approximation space $V_J^{(s-l)/t}$ also for the case $s - l < 0$.

Note that the cost for integral operators in Table 2 are not yet optimal, as we made no use of the potential of further compression of the stiffness matrix [11, 13, 15, 32, 33, 34, 43, 47].

6. Examples

In this section we give some applications of the ideas described above. We deal with the Poisson problem with homogeneous Dirichlet boundary conditions and with the screen problem. These are two prominent elliptic problems that show the conceptual approach and may therefore serve as a guideline for dealing with other elliptic variational problems. First of all, we are looking for candidates of univariate wavelet bases that fulfill our requirements. Note that because of our tensor-product approach we can reduce the questions to the one-dimensional case. Specifically, those basis functions whose support intersects with the boundary have to fulfill special boundary conditions. To this end, we refer the interested reader to the literature and state that these problems can be settled. See [7, 8] for appropriate constructions of localized functions and their boundary adaptation.

Sobolev spaces of interest for the study of integral and differential equations on the n-dimensional unit square $I^n = [0,1]^n$ are defined by

$$\mathcal{H}^s(I^n) = \{f \in \mathcal{D}'(I^n) : \exists g \in \mathcal{H}^s(\mathbb{R}^n) : g|_{I^n} = f \text{ and } \|f\|_{\mathcal{H}^s(I^n)} = \inf_{f=g|_{I^n}} \|g\|_{\mathcal{H}^s(\mathbb{R}^n)}\}$$

and

$$\tilde{\mathcal{H}}^s(I^n) = \{f = g|_{I^n} : g \in \mathcal{H}^s(\mathbb{R}^n) \text{ and supp } g \subset I^n\}$$

equipped with the norm

$$\|f\|_{\tilde{\mathcal{H}}^s(I^n)} = \|g\|_{\mathcal{H}^s(\mathbb{R}^n)}.$$

The interpolation spaces $\mathcal{H}^s(I^n)$ and $\tilde{\mathcal{H}}^s(I^n)$ are dual to each other, i.e.,

$$(\mathcal{H}^s(I^n))' = \tilde{\mathcal{H}}^{-s}(I^n),\ (\tilde{\mathcal{H}}^s(I^n))' = \mathcal{H}^{-s}(I^n),\ -\infty < s < \infty.$$

Furthermore

$$\tilde{\mathcal{H}}^s(I^n) = \mathcal{H}^s_0(I^n) \equiv \mathrm{clos}_{\mathcal{H}^s(I^n)}\mathcal{C}^\infty_0(I^n) \text{ for } s > \tfrac{1}{2}, s \neq k + \tfrac{1}{2}, k \in \mathbb{N},$$

i.e., $\tilde{\mathcal{H}}^s(I^n)$ is the appropriate space for problems with homogeneous essential boundary conditions and $\tilde{\mathcal{H}}^s(I^n) = \mathcal{H}^s$ for $-\frac{1}{2} < s < \frac{1}{2}$. Which of these spaces is appropriate depends on the application. For example $\tilde{\mathcal{H}}^1(I^n) = \mathcal{H}^1_0(I^n)$ is the appropriate space for the Poisson problem with homogeneous essential boundary conditions. For the screen problem the space $\tilde{\mathcal{H}}^{1/2}$ is appropriate.

Sobolev spaces of functions in other spaces of interest, like those with dominating mixed derivative on I^n are defined analogously. For example we have

$$\mathcal{H}^s_{\mathrm{mix}}(I^n) := \mathcal{H}^s(I) \otimes \cdots \otimes \mathcal{H}^s(I) \text{ and } \tilde{\mathcal{H}}^s_{\mathrm{mix}}(I^n) := \tilde{\mathcal{H}}^s(I) \otimes \cdots \otimes \tilde{\mathcal{H}}^s(I).$$

To be able to repeat the reasoning above, function spaces fulfilling the required boundary conditions and a Jackson and a Bernstein inequality have to be constructed. Then the argument of Section 4 can be repeated with obvious modifications.

Here, we concentrate on semi-orthogonal linear spline wavelets (prewavelets) on uniform dyadic grids as introduced in [7]. Figure 8 (left) shows a prewavelet in the interior of the domain. Concerning our cases of interest, suitable boundary constructions have been given for example in [2, 8, 27] and [28], respectively.

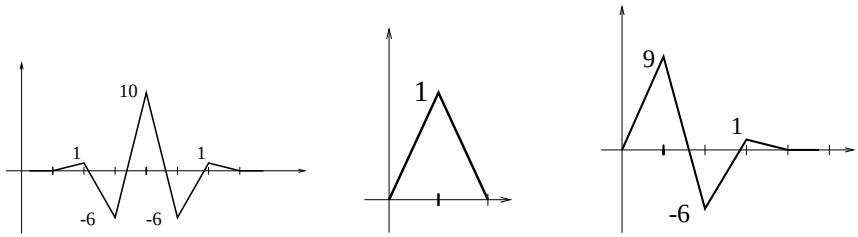

FIGURE 8. Semi-orthogonal linear spline prewavelet (left), nodal basis function $\psi_{1,0}$ corresponding to the coarsest level W_1 (middle) and boundary wavelet for the left boundary for homogeneous Dirichlet boundary conditions (right).

6.1. **Example: The Poisson equation.** We consider the problem

$$-\Delta u = f \tag{6.1}$$

with homogeneous Dirichlet boundary conditions in its variational form on $\mathcal{H}_0^1(I^n)$. In this case we have $s = 1$. Estimates of the cost[10] of solving (6.1) for $u \in \mathcal{H}^2_{\text{mix}}$ and continuous linear information up to an accuracy ϵ have been given in [5]. The authors constructed a finite element method using tensor-products of piecewise linear splines and index sets that are asymptotically equal to $I_J^{2/5}$. They proposed the use of a multilevel method to solve the resulting discrete problems. The resulting overall cost is then $O(\epsilon^{-1})$ because of the optimality of the proposed multilevel method and $\dim(I_J^{2/5}) = O(2^J)$.

Let us discuss our method in more detail. The basis function assigned to the coarsest level is the usual nodal basis function, see Figure 8 (middle). The orthogonal complement spaces W_j, for $j \geq 2$, are spanned by scaled and dilated versions of the functions shown in Figure 8 (left) for the interior grid points and Figure 8 (right) for the left boundary and an analogous construction for the right boundary. The resulting multilevel system incorporating homogeneous Dirichlet boundary conditions is a semi-orthogonal Riesz basis in $\mathcal{H}_0^s(I)$ for $0 \leq s < \frac{3}{2}$.

We assume that the solution of the variational problem is in the space $\mathcal{H}^{t,l}_{\text{mix}}$ for some parameters t, l with $t + l \geq 1$. From Table 1 we take the following orders of the cost:

$$\text{cost}(\epsilon) = \begin{cases} O\left(\epsilon^{1/(1-l-t)}\right) & \text{for } l < 1, \\ O\left(\epsilon^{-1/t}\left(\ln(\epsilon^{-1/t})\right)^{n-1}\right) & \text{for } l = 1, \\ O\left(\epsilon^{1/((1-l)/n-t)}\right) & \text{for } l > 1. \end{cases}$$

Specifically for the cases $u \in \mathcal{H}^2 = \mathcal{H}^{0,2}_{\text{mix}}, \mathcal{H}^2_{\text{mix}} = \mathcal{H}^{2,0}_{\text{mix}}$ and $\mathcal{H}^{1,1}_{\text{mix}}$ we obtain

$$\text{cost}(\epsilon) = \begin{cases} O\left(\epsilon^{-1}\right) & \text{for } u \in \mathcal{H}^2_{\text{mix}}, \\ O\left(\epsilon^{-1}\left(\ln(\epsilon^{-1})\right)^{n-1}\right) & \text{for } u \in \mathcal{H}^{1,1}_{\text{mix}}, \\ O\left(\epsilon^{-n}\right) & \text{for } u \in \mathcal{H}^2. \end{cases}$$

Hence we regain the result $O(\epsilon^{-1})$ of [5] as a special case. Note that for $u \in \mathcal{H}^{t,1}_{\text{mix}}$ the resulting optimized approximation space is V_J^0. Hence the cost is of the order $O\left(\epsilon^{-1/t}(\ln(\epsilon^{-1/t}))^{n-1}\right)$.

[10]Estimates of the ϵ-complexities of solving (6.1) with $f \in \mathcal{H}^t_{\text{mix}}$ and standard information can be found in [51].

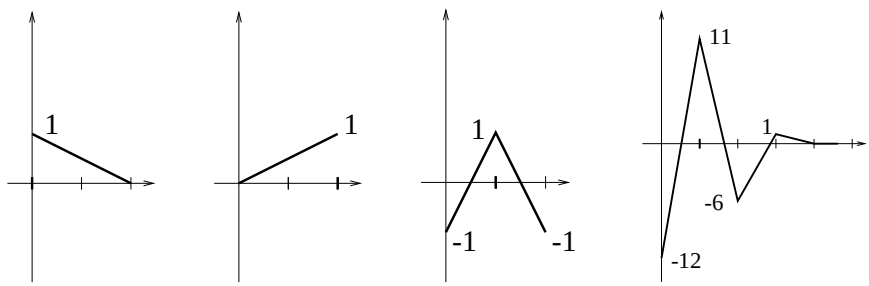

FIGURE 9. Basis functions $\psi_{0,0}, \psi_{0,1}$ and $\psi_{1,0}$ corresponding to the spaces W_0 and W_1 (first, second and third from left), boundary wavelet for the left boundary (right).

6.2. **Example: Single layer potential equation.** The second example we consider is the single layer potential equation

$$\frac{1}{c} \int\limits_{I^n} \frac{u(y)}{|x-y|} \, dy = f(x)$$

in its variational form on $\tilde{\mathcal{H}}^{-1/2}(I^n)$. Here we have $s = -\frac{1}{2}$. The corresponding bilinear form

$$a(u,v) = \left(\frac{1}{c} \int\limits_{I^n} \frac{u(y)}{|\cdot - y|} \, dy, v \right)_{\mathcal{H}^{1/2} \times \tilde{\mathcal{H}}^{-1/2}}, \qquad \forall u, v \in \tilde{\mathcal{H}}^{-1/2}(I^n),$$

is symmetric and $\tilde{\mathcal{H}}^{-1/2}$-elliptic. For problems in $\tilde{\mathcal{H}}^{-1/2}$ the basis does not have to fulfill special boundary conditions. The bases for W_0 and W_1 are shown in Figure 9. The orthogonal complement spaces W_j for $j \geq 2$ are spanned by scaled and dilated versions of the functions shown in Figure 8 (left) for the interior grid points and Figure 9 (right) for the left boundary and an analogous construction for the right boundary. The resulting multilevel system is a semi-orthogonal Riesz basis in $\mathcal{H}^s(I)$ for $0 \leq s < \frac{3}{2}$, and for $\tilde{\mathcal{H}}^s(I)$ for $-\frac{3}{2} < s < 0$. Hence this example is fully covered by the theory of Sections 3 and 4. In particular, the preconditioning and approximation results and the estimates of Section 5 can be applied. Regularity theory for the screen problem shows that if the right hand side vector f is smooth enough, then the solution u can be decomposed into a regular part u^{reg} and a singular part due to corner and edge singularities, compare also [28]. Here we restict ourselves to an approximation of the regular part of the solution. For a treatment of the singular parts see [28, 39].

Hence we assume that the solution of the variational problem is in the space $\tilde{\mathcal{H}}^{t,l}_{\text{mix}}$ for some parameters t, l with $t + l \geq -\frac{1}{2}$. From Table 2 we take the following orders of the cost.

$$\text{cost}(\epsilon) = \begin{cases} O\left(\epsilon^{-2/(1/2+l+t)}\right) & \text{for } l < -\frac{1}{2}, \\ O\left(\epsilon^{-2/t} \left(\ln(\epsilon^{-1/t})\right)^{2(n-1)}\right) & \text{for } l = -\frac{1}{2}, \\ O\left(\epsilon^{-2/((1/2+l)/n+t)}\right) & \text{for } l > -\frac{1}{2}. \end{cases}$$

Note that a further reduction of the cost complexity can be achieved by compression strategies as described in [11, 13, 15, 32, 33, 34, 43, 47].

7. Further aspects

We now give some hints on extensions to problems with large ellipticity constants, non-stable splittings and expansion systems other than wavelet-type multilevel bases. Specifically we derive modifications of the optimized spaces by incorporating additional information from the operators considered. This leads to the definition of anisotropic sparse grids [41]. Furthermore we discuss the potential possibilities of incorporating a priori known information about singularities of the solution into the construction process of optimized grids. Finally, we show how non-stable multilevel splittings, which only allow for an upper estimate instead of a full norm equivalence, still may be used to construct optimal sparse grid approximation spaces.

7.1. **Anisotropic sparse grids.** For problems with large ellipticity constants, the constants in the estimates of the approximation error become large and badly influence the behavior of the approximation in actual implementations, as the constants may dominate the error approximation for practical problem sizes. In these cases, the asymptotic estimates do not provide full insight into the behavior of the approximants. It is advisable to spare the detour via the $\mathcal{H}^s$-norm and to use norm estimates applied directly to $a(\cdot,\cdot)$. Then a further adaptation of the approximation space to the operator at hand can be obtained. This is important for preconditioning purposes also.
As a simple example consider the anisotropic elliptic problem

$$-\sum_{i=1}^{n} d_i \frac{\partial^2}{\partial^2 x_i} u = f, \quad d_i > 0, \tag{7.1}$$

in its variational form on $\mathcal{H}^1(I^n)$. Tensor-product approximation spaces are well-suited for such problems as they allow easily for anisotropic refinement. Let $a(\cdot,\cdot)$ denote the corresponding $\mathcal{H}^1$-elliptic variational form. The problem with the numerical solution of (7.1) is that the condition number of the Galerkin stiffness matrix on a isotropic full grid is linearly dependent on $\max_{1\le i\le n} d_i / \min_{1\le i\le n} d_i$. The same is true for the coefficient in the asymptotic estimate of the approximation error. Hence, for a fixed refinement level J and varying coefficients $d \equiv (d_1,\dots,d_n)$, the convergence rate of iterative methods, as well as the approximation error, depend on d. For problems with large anisotropies this leads to a slowdown of convergence and a deterioration of approximation. It is well-known that some kind of semi-coarsening in the subspace splittings or in the construction of the approximation spaces can remedy these problems. These ideas can also be used for the approximation spaces defined here. It amounts to the use of a norm estimate on $a(\cdot,\cdot)^{\frac{1}{2}}$ instead of $\|\cdot\|_{\mathcal{H}^1}$.
Again using Propositions 3.1 and 3.2, an argument analogous to that in the proof of Theorem 3.3 shows that

$$a(u,u) \approx \sum_{\mathbf{j}} \left(\sum_{i=1}^{n} d_i\, 2^{2j_i} \right) \|w_j\|^2_{\mathcal{L}^2} \approx \sum_{\mathbf{j}} \max_{1\le i\le n} (d_i 2^{2j_i}) \|w_{\mathbf{j}}\|^2_{\mathcal{L}^2} \tag{7.2}$$

for $u \in \mathcal{H}^1, u = \sum_{\mathbf{j}} w_{\mathbf{j}}$.[11] Compared to the norm equivalence (3.2) (set $s = 1$) the weight $2^{2|\mathbf{j}|_\infty}$ is substituted by the weight $\max_{1\le i\le n}(d_i 2^{2j_i})$ including information

[11] See [27] for a proof in the case of prewavelets where this norm estimate was used to obtain robust preconditioners for anisotropic problems.

from the anisotropy. Let $u \in \mathcal{H}^{t,l}_{\text{mix}}$ where $0 \le t < r$ and $1 \le t + l < r$. Let I_J be a subset of $I_J^{-\infty}$ and let V_{FE} be the corresponding approximation space. Then (7.2) together with (3.5) shows

$$\inf_{v \in V_{\text{FE}}} a(u-v, u-v) \le C \cdot \max_{\mathbf{j} \notin I_J} \left(\max_{1 \le i \le n} (d_i 2^{2j_i}) \cdot 2^{-2l|\mathbf{j}|_\infty - 2t|\mathbf{j}|_1} \right) \|u\|^2_{\mathcal{H}^{t,l}_{\text{mix}}} . \tag{7.3}$$

Without loss of generality we may assume that

$$d_1 = \text{argmax}_{1 \le i \le n}\{d_i\} \quad \text{and} \quad d_n = \text{argmin}_{1 \le i \le n}\{d_i\}.$$

Fixing $(J, 1, \cdots, 1)$ to be the index with maximal $|\cdot|_\infty$-norm to be included into the index sets leads to $c = \frac{1}{2t}\log_2(d_1 2^{2J}) - n + 1 - \frac{1}{t}J$ and the index sets

$$I_J^{l,t,d} := \{\mathbf{j} \in \mathbb{N}^n : -|\mathbf{j}|_1 - \frac{l}{t}|\mathbf{j}|_\infty + \frac{1}{2t}\log_2\big(\max_{1 \le i \le n}\big(\frac{d_i}{d_1} 2^{2(j_i - 2J)}\big)\big) \ge -(n+J-1) - \frac{l}{t}J\}$$

where the index d indicates the dependence on the parameters $d_i, 1 \le i \le n$. Fixing $(1, \cdots, 1, J)$ to be the index with maximal n-th component to be included into the index sets leads to $c = \frac{1}{2t}\log_2(d_n 2^{2J}) - n + 1 - \frac{1}{t}J$ and the index sets

$$\widehat{I}_J^{l,t,d} := \{\mathbf{j} \in \mathbb{N}^n : -|\mathbf{j}|_1 - \frac{l}{t}|\mathbf{j}|_\infty + \frac{1}{2t}\log_2\big(\max_{1 \le i \le n}\big(\frac{d_i}{d_n} 2^{2(j_i - 2J)}\big)\big) \ge -(n+J-1) - \frac{l}{t}J\}.$$

Then the corresponding approximation spaces keep the order of approximation of the full grid approximation space. Estimates on the dimension and the order of approximation can be derived in the spirit of the preceding sections. We obtain the same orders of approximation as for the spaces $V_J^{(1-l)/t}$ but with different coefficients. Note that in the case of the index set $I_J^{l,t,d}$ the coefficient is dependent on $d_1 = \text{argmax}_{1 \le i \le n}\{d_i\}$ and the number of unknowns is further reduced, as $I_J^{l,t,d} \subset I_J^{(1-l)/t}$. For the case $\widehat{I}_J^{l,t,d}$ the number of unknowns is increased compared to $I_J^{(1-l)/t}$, but the coefficient is only depending on $d_n = \text{argmin}_{1 \le i \le n}\{d_i\}$. The norm equivalence (7.2) leads also to robust preconditioners, compare [27]. In the case of extreme anisotropy, the resulting grid consists of extremely stretched grids in the direction of the anisotropy, corresponding to semicoarsening. Figures 10 and 11 show some examples in two dimensions.

At this point note that there is a close relation to the so-called weighted spaces from [44, 45] where weights are introduced into Sobolev norms. There, the curse of dimensionality no longer shows up if certain conditions on the weights are fulfilled. Thus strong tractability [48] of integration can be achieved. These weights resemble to some extent our diffusion coefficients d_i.

7.2. **Adaptive sparse grids.** So far, our theory involves an a priori approach, i.e., we beforehand assume the solution to be from a specific function class and we then determine the best approximation space with respect to cost and accuracy. For practical purposes however such an a priori approach is not always feasible. This may be because the class of data (and thus the regularity of the solution of the problem) is not known beforehand or due to the fact that the smoothness needed is not present. Then, our previous algorithms need to be complemented with some special treatment of singular parts of the solution of the variational problem. The idea is that a few wavelets of high level clustered around the singularity will suffice, while the optimized grids of the previous sections are enough to treat the smooth parts of the solution. To this end, it is helpful to refine the selection criterion to

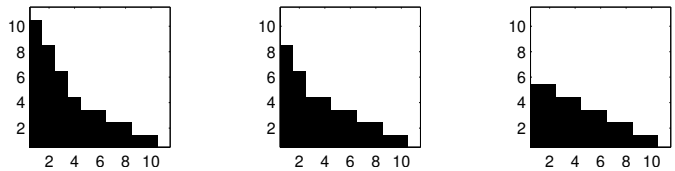

FIGURE 10. Index sets $I_{10}^{2,0,d}$ for $d_1/d_2 = 1$, $d_1/d_2 = 10$ and $d_1/d_2 = 1000$ from left to right, two dimensional case.

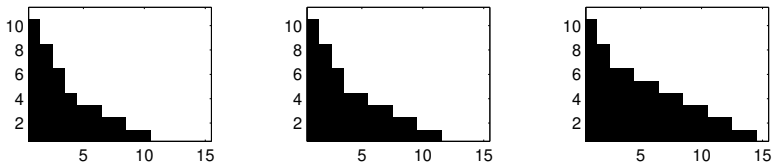

FIGURE 11. Index sets $\widehat{I}_{10}^{2,0,d}$ for $d_1/d_2 = 1$, $d_1/d_2 = 10$ and $d_1/d_2 = 1000$ from left to right, two dimensional case.

the atomic level, i.e., to allow for single basis functions/grid points to be selected. From (4.23) together with (4.21) we obtain the profit/cost quotient of a single basis function

$$\gamma_{\mathbf{j},\mathbf{l}} := 2^{2s|\mathbf{j}|_\infty} \langle u, \tilde{\psi}_{\mathbf{j},\mathbf{l}} \rangle^2. \tag{7.4}$$

Suppose that for example the leading singularity component χ of the solution u is known. Decomposing χ with respect to the given basis, we can use the weights $|\langle \chi, \tilde{\psi}_{\mathbf{j},\mathbf{l}} \rangle|$ in (7.4) instead of the weights $|\langle u, \tilde{\psi}_{\mathbf{j},\mathbf{l}} \rangle|$. This leads to the definition of grids adapted to χ by choosing those indices that have

$$\gamma_{\mathbf{j},\mathbf{l}} := 2^{2s|\mathbf{j}|_\infty} \langle \chi, \tilde{\psi}_{\mathbf{j},\mathbf{l}} \rangle^2 \tag{7.5}$$

bigger than some threshold. This a priori adaptivity leads to a relatively high degree of adaptivity without complicated mesh refinement strategies especially for problems in higher dimensions. Nevertheless, for singularly perturbed problems with large ellipticity constants and problems that exhibit boundary singularities where χ is not known beforehand, a posteriori adaptivity is still necessary. Locally adaptive sparse grid methods can be found in [3, 4, 6, 17, 20]. We further refer to [10, 14] for results on nonlinear approximation and adaptivity and to [37] for results on nonlinear approximation with sparse grids.

7.3. **Other multilevel systems.** The constructions of the approximation spaces presented in this paper are not restricted to biorthogonal wavelets as basis functions, but can be carried over to other multiscale basis functions as well. Specifically, the construction of optimized grids given in Subsection 4.3 is not limited to stable multilevel splittings, that is, to multilevel finite element spaces that possess norm equivalences like those described in Section 3. Instead only an upper estimate is needed. Consider for example the case of an $\mathcal{H}^1$-elliptic operator and multiscale basis functions of tensor-products of piecewise linear splines $\phi_{\mathbf{j},\mathbf{k}}$. Let $W_{\mathbf{j}} = \operatorname{span}\{\phi_{\mathbf{j},\mathbf{k}}, \mathbf{k} \in \tau_{\mathbf{j}}\}$ denote the hierarchical difference space between two successive spaces spanned by piecewise n-linear functions. It is easy to see that in this

case there holds a Bernstein inequality

$$\|w_{\mathbf{j}}\|_{\mathcal{H}^1} \leq C2^{|\mathbf{j}|_\infty}\|w_{\mathbf{j}}\|_{\mathcal{L}^2} \quad \forall w_{\mathbf{j}} \in W_{\mathbf{j}}, \mathbf{j} \in \mathbb{N}_0^n, \tag{7.6}$$

and that the estimate

$$\|w_{\mathbf{j}}\|_{\mathcal{L}^2} \leq C2^{-2|\mathbf{j}|_1}\|u\|_{\mathcal{H}^2_{\text{mix}}} \quad \forall u = \sum_{\mathbf{j}} w_{\mathbf{j}} \in \mathcal{H}^2_{\text{mix}}, \tag{7.7}$$

holds, see [5]. Inequality (7.7) can again be inferred from decay properties of the coefficients in the representation of u in the bases of piecewise linear splines. Then, applying the triangle inequality together with (7.6) and (7.7) yields

$$\inf_{v \in V_J^T} \|u - v\|_{\mathcal{H}^1} \leq \|\sum_{\mathbf{j} \notin I_J^T} w_{\mathbf{j}}\|_{\mathcal{H}^1} \leq \sum_{\mathbf{j} \notin I_J^T} \|w_{\mathbf{j}}\|_{\mathcal{H}^1} \overset{(7.6),(7.7)}{\leq} C \sum_{\mathbf{j} \notin I_J^T} 2^{|\mathbf{j}|_\infty - 2|\mathbf{j}|_1}\|u\|_{\mathcal{H}^2_{\text{mix}}}$$

for $u \in \mathcal{H}^2_{\text{mix}}$. Summing up, a longer calculation gives a generalized Jackson inequality

$$\inf_{v \in V_J^T} \|u - v\|_{\mathcal{H}^1} \leq C2^{-J}\|u\|_{\mathcal{H}^2_{\text{mix}}} \tag{7.8}$$

for $T < \frac{1}{2}$ where $C = C(T)$. Hence the optimal order of approximation is kept as long as $T < \frac{1}{2}$. That is, we obtain a similar result for a multilevel approximation space without the direct use of norm equivalences. This can also be used as the starting point for enlarging the range of the validity of the estimates presented in this paper. Especially the upper range of the parameters t and l which were restricted from above by $t + l < r$ and $t < r$ could be enlarged to the whole range $t + l \leq m$ and $t \leq m$, see (2.17). Apart from eventual logarithmic terms in the extremal cases, the results remain the same.

8. Concluding remarks

In this paper we constructed approximation spaces for elliptic variational problems with solution in $\mathcal{H}^{t,l}_{\text{mix}}$. We gave cost estimates for the case of continuous linear information. We showed these results in a constructive manner by proposing a Galerkin method together with optimal preconditioning. Specifically, we identified smoothness assumptions that make it possible to choose the approximation space in such a way that the number of degrees of freedom is $O(2^J)$ compared to $O(2^{nJ})$ for the full grid space, while keeping the optimal order of approximation.

A disadvantage of the approaches described in this paper is that generalizations to more general geometries are not easy to handle. Research in this direction is mainly based either on domain transformation techniques or on some kind of domain decomposition approach where the computational domain is decomposed locally and transformed to unit cubes. On these local domains the wavelet techniques can be applied. Note however that $[0, 1]^n$ is a natural computational domain for many higher-dimensional physical applications.

References

1. R. Adams, *Sobolev spaces*, Academic Press, New York, 1975.
2. P. Auscher, *Wavelets with boundary conditions on the interval*, in C. Chui, (editor), *Wavelets: A tutorial in theory and applications*, Acad. Press, Boston, 217–236, 1992.
3. H.-J. Bungartz, *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*, Dissertation, TU München, Institut für Informatik, 1992.

4. H.-J. Bungartz, *Finite elements of higher order on sparse grids*, Habilitation, TU München, Institut für Informatik, 1998.
5. H.-J. Bungartz, M. Griebel, *A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives*, J. Complexity, 15:167–199, 1999.
6. H.-J. Bungartz, M. Griebel, *Sparse grids*, Acta Numerica, 13:1–123, 2004.
7. C. Chui, J. Wang, *On compactly supported spline wavelets*, Trans. Amer. Math. Soc., 330:903–915, 1992.
8. A. Cohen, I. Daubechies, P. Vial, *Wavelets on the interval and fast wavelet transforms*, Appl. Comput. Harm. Anal., 1:54–81, 1993.
9. W. Dahmen, *Stability of multiscale transformations*, Journal of Fourier Analysis and Applications, 2:341–361, 1996.
10. W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, 6:55–228, 1997.
11. W. Dahmen, H. Harbrecht, R. Schneider, *Adaptive methods for boundary integral equations: Complexity and convergence estimates*, Math. Comp., 76:1243–1274, 2007.
12. W. Dahmen, A. Kunoth, *Multilevel preconditioning*, Numer. Math., 63:315–344, 1992.
13. W. Dahmen, S. Prößdorf, R. Schneider, *Wavelet approximation methods for pseudodifferential equations II: Matrix compression and fast solution*, Advances in computational Mathematics, 1:259–335, 1993.
14. R. DeVore, *Nonlinear approximation*, Acta Numerica, 7:51–150, 1998.
15. R. DeVore, B. Jawerth, V. Popov, *Compression of wavelet decompositions*, Amer. J. Math., 114:737–785, 1992.
16. K. Frank, S. Heinrich, S. Pereverzev, *Information complexity of multivariate Fredholm equations in Sobolev classes*, J. Complexity, 12:17–34, 1996.
17. C. Feuersänger, Dünngitterverfahren für hochdimensionale elliptische partielle Differentialgleichungen, Diplomarbeit, Institut für Numerische Simulation, Universität Bonn, 2005.
18. M. Griebel, *A parallelizable and vectorizable multi-level algorithm on sparse grids*, Parallel Algorithms for Partial Differential Equations, Notes on Num. Fluid Mech. 31, W. Hackbusch, (editor), Vieweg, Braunschweig, 1991.
19. M. Griebel, *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen*, Teubner Skripten zur Numerik, Teubner, Stuttgart, 1994.
20. M. Griebel, *Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences*, Computing, 61(2):151–179, 1998.
21. M. Griebel, *Sparse grids and related approximation schemes for higher dimensional problems* In Foundations of Computational Mathematics, L. Pardo, A. Pinkus, E. Suli, M. Todd, (editors), LMS 331, Cambridge University Press, Cambridge, 2006.
22. M. Griebel, J. Hamaekers, *Sparse grids for the Schrödinger equation*, Mathematical Modeling and Numerical Analysis, 41(2):215–247, 2007.
23. M. Griebel, J. Hamaekers, *A wavelet based sparse grid method for the electronic Schrdinger equation*, in M. Sanz-Sol, J. Soria, J. Varona, J. Verdera, (editors), Proceedings of the International Congress of Mathematicians, volume III, 1473–1506, Madrid, Spain, 22.-30. August 2006, European Mathematical Society.
24. M. Griebel, S. Knapek, *Optimized approximation spaces for operator equations*, Technical Report 568, SFB 256, Univ. Bonn, 1999.
25. M. Griebel, S. Knapek, *Optimized tensor-product approximation spaces*, Constructive Approximation, 16(4):525–540, 2000.
26. M. Griebel, D. Oeltz, *A sparse grid space-time discretization scheme for parabolic problems*, Computing, 81(1):1–34, 2007.
27. M. Griebel, P. Oswald, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Advances in Computational Mathematics, 4:171–206, 1995.
28. M. Griebel, P. Oswald, T. Schiekofer, *Sparse grids for boundary integral equations*, Numer. Mathematik, 83(2):279–312, 1999.
29. M. Griebel, M. Schneider, C. Zenger, *A combination technique for the solution of sparse grid problems*, in Iterative Methods in Linear Algebra, P. de Groen, R. Beauwens, (editors), Elsevier, Amsterdam, 263–281, 1992.
30. R. Hochmuth, S. Knapek, G. Zumbusch, *Tensor products of Sobolev spaces and applications*, Technical Report 685, SFB 256, Univ. Bonn, 2000.

31. S. Jaffard, *Wavelet methods for fast resolution of elliptic equations*, SIAM J. Numer. Anal., 29:965–986, 1992.
32. S. Knapek, *Hyperbolic cross approximation of integral operators with smooth kernel*, Technical Report 665, SFB 256, Univ. Bonn, 2000.
33. S. Knapek, *Compression of anisotropic tensor-product discretizations*, Technical Report 0200, Institut für Numerische Simulation, Universität Bonn, 2002.
34. S. Knapek, F. Koster, *Integral operators on sparse grids*, SIAM J. Num. Anal., 39(5):1794–1809, 2002.
35. L. Kronsjö, G. Dahlquist, *On the design of nested iteration for elliptic differential equations*, BIT, 11:63–71, 1971.
36. S. Martello, P. Toth, *Knapsack problems: Algorithms and computer implementations*, John Wiley & Sons, Chichester, 1990.
37. P. Nitsche, *Sparse approximation of singularity functions*, Constructive Approximation, 21:63–81, 2004.
38. P. Oswald, *On discrete norm estimates related to multilevel preconditioners in the finite element method*, in: Constructive Theory of Functions, K.G. Ivanov, P. Petrushev, B. Sendov, (editors), Proc. Int. Conf. Varna, 1991, Bulg. Acad. Sci., Sofia, 203–214, 1992.
39. P. Oswald, *Best N-term capacitance approximation on sparse grids*, Proc. 12th Intern. Conf. on Domain Decomposition Methods in Science and Engineering, T. Chan et al., (editors), Chiba, 1999, 437–444, 2001.
40. S. Pereverzev, *On the complexity of finding solutions of Fredholm equations of the second kind with differentiable kernels II*, Ukrain. Mat. Zh., 41:169–173, 1989.
41. D. Röschke, *Über eine Kombinationstechnik zur Lösung partieller Differentialgleichungen*, Diplomarbeit, Institut für Informatik, TU München, 1991.
42. H.-J. Schmeisser, H. Triebel, *Fourier analysis and functions spaces*, John Wiley, Chichester, 1987.
43. R. Schneider, *Multiskalen- und Wavelet-Kompression: Analysisbasierte Methoden zur effizienten Lösung großer vollbesetzter Gleichungssysteme*, Advances in Numerical Analysis, Teubner, Stuttgart, 1998.
44. I. Sloan, H. Wozniakowski, *When are Quasi-Monte Carlo algorithms efficient for high dimensional problems*, Journal of Complexity, 14:1–33, 1998.
45. I. Sloan, H. Wozniakowski, *Tractability of integration in non-periodic and periodic weighted tensor product Hilbert spaces*, Journal of Complexity, 18:479–499, 2002.
46. S. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 4:240–243, 1963.
47. R. Stevenson, *On the compressibility of operators in wavelet coordinates*, SIAM J. Math. Anal., 35(5):1110–1132, 2004.
48. J. Traub, G. Wasilkowski, H. Wozniakowski, *Information-based complexity*, Academic Press, New York, 1988.
49. J. Weidmann, *Linear operators in Hilbert spaces*, Springer, New York, 1980.
50. A. Werschulz, *The computational complexity of differential and integral equations, An information based approach*, Oxford University Press, New York, 1991.
51. A. Werschulz, *The complexity of the Poisson problem for spaces of bounded mixed derivatives*, Technical Report CUCS-016-95, 1995.
52. C. Zenger, *Sparse grids*, In W. Hackbusch, (editor), Parallel Algorithms for Partial Differential Equations, *Notes on Num. Fluid Mech. 31*, Vieweg, Braunschweig, 1991.

Institut für Numerische Simulation, Universität Bonn, D-53115 Bonn
E-mail address: `griebel@ins.uni-bonn.de`