



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Space-Time Approximation with Sparse Grids

M. Griebel, D. Oeltz, P. S. Vassilevski

April 26, 2005

SIAM Journal on Scientific Computing

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

SPACE-TIME APPROXIMATION WITH SPARSE GRIDS

MICHAEL GRIEBEL[†] DANIEL OELTZ[‡] PANAYOT VASSILEVSKI[‡]

Abstract. In this article we introduce approximation spaces for parabolic problems which are based on the tensor product construction of a multiscale basis in space and a multiscale basis in time. Proper truncation then leads to so-called *space-time sparse grid spaces*. For a uniform discretization of the spatial space of dimension d with $O(N^d)$ degrees of freedom, these spaces involve for $d > 1$ also only $O(N^d)$ degrees of freedom for the discretization of the whole space-time problem. But they provide the same approximation rate as classical space-time Finite Element spaces which need $O(N^{d+1})$ degrees of freedoms. This makes these approximation spaces well suited for conventional parabolic and for time-dependent optimization problems.

We analyze the approximation properties and the dimension of these sparse grid space-time spaces for general stable multiscale bases. We then restrict ourselves to an interpolatory multiscale basis, i.e. a hierarchical basis. Here, to be able to handle also complicated spatial domains Ω , we construct the hierarchical basis from a given spatial Finite Element basis as follows: First we determine coarse grid points recursively over the levels by the coarsening step of the algebraic multigrid method. Then, we derive interpolatory prolongation operators between the respective coarse and fine grid points by a least squares approach. This way we obtain an algebraic hierarchical basis for the spatial domain which we then use in our space-time sparse grid approach.

We give numerical results on the convergence rate of the interpolation error of these spaces for various space-time problems with two spatial dimensions. Also implementational issues, data structures and questions of adaptivity are addressed to some extent.

Key words. parabolic differential equations, multilevel basis, sparse grids, space-time discretization

AMS subject classifications. 35K20, 65M99, 65Y20

1. Introduction. The modeling of various phenomena in physics, chemistry, biology and financial engineering leads to time dependent partial differential equations. Mostly there is no analytical solution for these equations available and hence they have to be solved approximately by an appropriate discretization scheme. If we employ, depending on the smoothness of the solution, a discretization scheme on a uniform grid in time and a uniform grid in d -dimensional space with an error of the order p in time and order q in space we need $O(N^d)$ degrees of freedom for each time step and $O(N^{d+q/p})$ degrees of freedom for the whole approximate solution in space-time to obtain an *overall* order of q . In particular, for a method which is second order in space and in time this would result in $O(N^{d+1})$ degrees of freedom. An approach which is only first order in time, e.g. the Euler methods, but second order in space needs $O(N^{d+2})$ degrees of freedom. The associated additional storage requirements then prohibit to work for parabolic PDEs directly in space-time if $d = 3$ or larger, but enforces a time-slice approach, where only a few spatial grids at successive time points need to be stored.

In some applications it is necessary to solve a system of parabolic equations, where some of the equations are forward and the other are backward in time. Such systems typically arise in parabolic optimal control problems where the equations for

[†]Sonderforschungsbereich 611, Singuläre Phänomene und Skalierung in mathematischen Modellen, Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, D-53115 Bonn, Germany, (griebel, oeltz)@ins.uni-bonn.de

[‡]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, Mail Stop L-560, Livermore, CA 94551, U.S.A., panayot@llnl.gov

This work was performed under the auspices of the U. S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

the adjoint variables are backward in time but also involve the state variables for which forward equations have to be solved, see e.g. [34]. Here, all values of the state variable in space *and* time must be stored to be used for the solution of the adjoint. This, however, is not computationally feasible for the case $d \geq 3$ due to the complexity of $O(N^{d+1})$ or $O(N^{d+2})$, even if N is moderately small.

One approach to overcome this storage problem is the so-called reduced order method [35, 38]. Here, one tries to compute the most important spatial additional storage of the instationary solution from snapshots of the solution at different times. This results in (hopefully) only few spatial basis functions which approximate the given solution with sufficient accuracy over time. To this end, the instationary problem has however to be approximately solved at the beginning to obtain the snapshots at certain time instances at all. Also, for any update of the reduced order model during an optimization it is necessary to compute an approximate solution to the actual full instationary problem again. Due to the above mentioned complexity orders this is still very expensive. Furthermore, the choice of the time points at which snapshots are taken strongly influence the reduced order model. Up to now, the question how to choose the amount and the locations of the time points for the snapshots is not finally answered and these topics are still a subject of current research.

Another approach is the so-called checkpointing [33]. Here, the solution of the state equation is stored at only some properly chosen time points, called checkpoints. Now, if the value of the state equation at a certain time is needed, the forward state equation is solved with start time of the nearest checkpoint and the state variable at this point as initial value. Clearly, checkpointing leads to a further increase of computational time on the already expensive solution of the instationary problem.

In the present article we introduce new space-time sparse grid spaces which allow for complexity reduction of the space-time problem from $O(N^{d+q/p})$ to $O(N^d)$. If a slightly more involved regularity assumption is fulfilled, namely if the mixed second derivative in space and time is bounded, the approximation rate for the space-time sparse grid space remains basically the same as for classical Finite Element spaces in space and time. Thus, the additional complexity stemming from the time part of a parabolic PDE is eliminated and we obtain the same complexity order as for a stationary problem. Consequently, the memory and storage requirements are substantially reduced. These space-time sparse grid spaces are constructed from a d -dimensional multilevel basis in space and a one-dimensional multilevel basis in time via a tensor product approach. In this sense they are similar to classical sparse grid spaces, see for example [13], which are constructed via a tensor product approach from a one-dimensional multilevel basis for each coordinate direction. We, however only use a d -dimensional isotropic multilevel basis in spatial space. For this purpose, any available stable multilevel basis in space can be employed. For the ease of presentation, we restrict ourselves to an interpolatory multilevel basis, namely a hierarchical basis. To construct the necessary spatial multilevel basis from a given general Finite Element basis we proceed as follows: First we determine coarse grid points recursively over the levels by the coarsening step of an algebraic multigrid method. Then, to assign hierarchical basis functions to these points we use interpolatory prolongation operators between the respective coarse and fine grids which are derived by a least squares approach. We need higher accuracy of the prolongation operators than in the classical AMG methods in order to guarantee the desired approximation property of the hierarchical basis. In particular, by the least-squares approach we employed the linear functions are well represented on coarse levels. Note that the classical AMG methods

have only “weak approximation property” guaranteed, namely only the constant function is well represented on coarse levels. Alternatives to the least squares approach could be the smoothed aggregation AMG [49] or the multiple vector preserving AMG interpolation proposed in [50]. These are not considered in the present paper but worth pursuing in the future. The AMG approach in general, allows us to employ any given Finite Element package on the finest level of the spatial discretization. Thus, we are able to deal also with complicated spatial domains Ω where classical sparse grid spaces run into difficulties.

In the context of instationary control problems, these spaces can be used for an interpolant of the state variable which was solved with a classical discretization scheme. Then this cheap space-time sparse grid interpolant is employed in the adjoint equation. Alternatively, these spaces can be directly used for the discretization of both, the state problem and the adjoint problem. Altogether, our approach overcomes the storage complexity in time-dependent two- and three-dimensional optimization problems and allows to deal with quite complicated spatial geometries.

The remainder of this article is organized as follows:

In section 2 we describe the general construction of the proposed space-time sparse grid spaces from a multilevel basis in space and a multilevel basis in time. To this end, we simply use a tensor product and properly truncate the resulting multivariate series expansion. We then present results on the approximation rates for these spaces. Here, provided that a slightly stronger regularity assumption is fulfilled, it turns out that the approximation order of classical full grid spaces in space-time is preserved while the overall complexity is significantly reduced.

In section 3 we discuss how a spatial hierarchical basis can be derived from a given nodal basis and a given sequence of prolongation operators. Here we use the classical coarsening scheme from algebraic multigrid methods to obtain coarser grids and a least squares approach to compute the necessary interpolation operators.

In section 4 we address some implementational issues arising for the proposed space-time sparse grid approach. We first describe the necessary operations which must be performed on the data to compute the space-time sparse grid interpolant for a given function. These insights are then used as guideline for the development of efficient data structures for space-time sparse grids. Furthermore, we discuss additional requirements to take adaptivity in space and time into account.

In section 5 we consider numerical experiments with the proposed space-time sparse grid space. Here, we focus on the interpolation error. The results show that the approximation rates with respect to the L^2 -norm deteriorate only slightly in comparison to that of classical Finite Element spaces in space and time. For the H^1 -norm even the same rates are achieved. In addition, the results for an adaptive space-time sparse grid applied to a non-smooth function indicate that also singular functions can be dealt with in an efficient manner.

Finally, in section 6 we draw some conclusions.

2. Space-Time Sparse Grids. In this section we describe the general construction of space-time sparse grid spaces following the presentation in [30]. We discuss their approximation properties as well as their cost complexity. It will turn out that these spaces provide the same approximation rate as classical Finite Element spaces in space-time under just slightly stronger regularity assumptions. The number of the involved degrees of freedom, however, is significantly reduced, i.e. by a factor of $O(N^{q/p})$ where p denotes the order of the associated time discretization and q the order of the space discretization, respectively.

For a d -dimensional spatial domain $\Omega \subset \mathbb{R}^d$ and $T > 0$ we define $\Omega_T := \Omega \times (0, T)$. In the following, let $H^m(\Omega)$, $m \in \mathbb{N}$, denote the usual Sobolev space on Ω with weak derivatives of order m . Now, for $m \in \mathbb{N}$, we define

$$H^{2m,m}(\Omega_T) := (H^{2m}(\Omega) \otimes L^2((0, T))) \cap (L^2(\Omega) \otimes H^m((0, T))) \quad (2.1)$$

with associated norm

$$\|u\|_{H^{2m,m}(\Omega_T)}^2 := \sum_{2 \cdot p + \|\mathbf{q}\|_1 \leq 2m} \|\partial_t^p \partial_x^{\mathbf{q}} u\|_{L^2(\Omega_T)}^2. \quad (2.2)$$

Here and in the following we use boldface letters to denote multi-indices, i.e. $\mathbf{q} \in \mathbb{N}^d$. Function spaces with bounded mixed derivatives play an important role in the theoretical analysis of sparse grids, see [13, 46, 47]. For our space-time setting we define

$$H_{mix}^{2m,m}(\Omega_T) := H^{2m}(\Omega) \otimes H^m((0, T)) \quad (2.3)$$

with associated norm

$$\|u\|_{H_{mix}^{2m,m}(\Omega_T)}^2 := \sum_{\substack{p \leq m, \\ \|\mathbf{q}\|_1 \leq 2m}} \|\partial_t^p \partial_x^{\mathbf{q}} u\|_{L^2(\Omega_T)}^2, \quad (2.4)$$

and

$$H_{mix}^{m,m}(\Omega_T) := H^m(\Omega) \otimes H^m((0, T)), \quad (2.5)$$

with associated norm

$$\|u\|_{H_{mix}^{m,m}(\Omega_T)}^2 := \sum_{\substack{p \leq m, \\ \|\mathbf{q}\|_1 \leq m}} \|\partial_t^p \partial_x^{\mathbf{q}} u\|_{L^2(\Omega_T)}^2. \quad (2.6)$$

Note that the space $H_{mix}^{2m,m}(\Omega_T)$ is of similar structure as the space $H^{2m,m}$ which is usually employed in regularity estimates of parabolic problems, c.f. [39]. Under only slightly stronger regularity assumptions for the initial condition and right hand side than in classical regularity theory, the solution of the Cauchy problem can actually be shown to be in $H_{mix}^{2m,m}(\Omega_T)$, see [30] for details.

In the following we want to construct approximation spaces for the approximation of functions $u \in H_{mix}^{2m,m}(\Omega_T)$ and $u \in H_{mix}^{m,m}(\Omega_T)$, respectively. They will possess a substantially reduced number of overall degrees of freedom compared to a conventional Finite Element space on Ω_T without a change in the approximation order. Let us assume that there are finite dimensional spaces V_j^Ω and V_j^T , $j \in \mathbb{N}$, such that

$$\begin{aligned} L^2(\Omega) &= \bigoplus_{j \geq 0} V_j^\Omega, \\ L^2((0, T)) &= \bigoplus_{j \geq 0} V_j^T, \end{aligned}$$

and $V_j^\Omega \subset V_{j+1}^\Omega$, $V_j^T \subset V_{j+1}^T$. Furthermore we define increment spaces W_j^Ω and W_j^T , such that

$$\begin{aligned} V_j^\Omega &= V_{j-1}^\Omega \oplus W_j^\Omega, \quad j \geq 1, \\ V_j^T &= V_{j-1}^T \oplus W_j^T, \quad j \geq 1, \end{aligned}$$

and $W_0^\Omega := V_0^\Omega$, $W_0^T := V_0^T$. Since $L^2(\Omega_T) = L^2(\Omega) \otimes L^2((0, T))$, c.f. [36], we obtain

$$L^2(\Omega_T) = \bigoplus_{\mathbf{j} \in \mathbb{N}^2} V_{\mathbf{j}}, \quad V_{\mathbf{j}} = V_{j_1}^\Omega \otimes V_{j_2}^T.$$

We define the hierarchical increment spaces $W_{\mathbf{j}}$, $\mathbf{j} \in \mathbb{N}^2$, as

$$W_{\mathbf{j}} = W_{(j_1, j_2)} := W_{j_1}^\Omega \otimes W_{j_2}^T. \quad (2.7)$$

To simplify notation we will omit the domain of integration in the norms, e.g. we write $\|\cdot\|_{L^2}$ instead of $\|\cdot\|_{L^2(\Omega)}$. If not stated otherwise we assume that the hierarchical increments W_j^T and W_j^Ω are L^2 -stable, i.e.¹

$$\begin{aligned} \left\| \sum_j w_j \right\|_{L^2(\Omega)}^2 &\simeq \sum_j \|w_j\|_{L^2(\Omega)}^2 \text{ for all } \sum_j w_j \in L^2(\Omega), w_j \in W_j^\Omega, \\ \left\| \sum_j w_j \right\|_{L^2((0, T))}^2 &\simeq \sum_j \|w_j\|_{L^2((0, T))}^2 \text{ for all } \sum_j w_j \in L^2((0, T)), w_j \in W_j^T. \end{aligned}$$

With this notation, we define the full grid spaces

$$V_l^\infty := \bigoplus_{\substack{2j_1 \leq 2l, \\ j_2 \leq 2l}} W_{(j_1, j_2)} \quad (2.8)$$

and

$$\tilde{V}_l^\infty := \bigoplus_{\|\mathbf{j}\|_\infty \leq l} W_{(j_1, j_2)} \quad (2.9)$$

as well as the sparse grid spaces

$$V_l^0 := \bigoplus_{2j_1 + j_2 \leq 2l} W_{(j_1, j_2)} \quad (2.10)$$

and

$$\tilde{V}_l^0 := \bigoplus_{\|\mathbf{j}\|_1 \leq l} W_{\mathbf{j}}. \quad (2.11)$$

Note that the space-time sparse grid spaces V_l^0 and \tilde{V}_l^0 are similar to classical sparse grid spaces, see [11, 13, 29]. Now, however, the supports of the basis functions are anisotropic with respect to space versus time but they are isotropic in space. This results from our construction (2.7) which involves a tensor product between a (isotropic) multilevel basis in space and a one-dimensional multilevel basis in time which is in contrast to the conventional sparse grid approach where the basis in space is formed by a tensor product of a one-dimensional multiscale basis for each coordinate direction. Furthermore, the space V_l^0 is twice as much refined in time direction than in space direction. This can also be seen in Figure 2.1 (right) where sparse grids of level $l = 4$ on the domain $\Omega_T = (0, 1)^2$ are shown. To this end, the one-dimensional piecewise linear hierarchical basis [23, 51, 52] was used for both, space and time, within the tensor product construction. The sparse grid points are the center points of the supports of the associated basis functions used for the respective spaces. Here we

¹Here, $A \simeq B \Leftrightarrow c_1 B \leq A \leq c_2 B$ with positive constants c_1, c_2 .

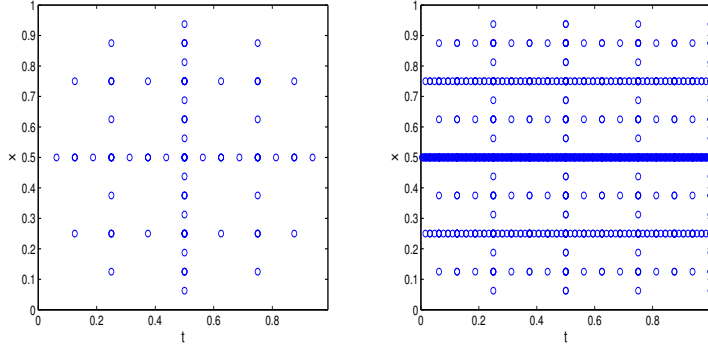


FIGURE 2.1. Grid points for the space \tilde{V}_4^0 (left) and V_4^0 (right) for the one dimensional domain $\Omega = (0, 1)$ and the time interval $(0, 1)$.

see that V_l^0 involves more grid points in time than in space direction which results in an higher overall number of degrees of freedom than for \tilde{V}_l^0 . As we will see later in Lemma 2.3 the sparse grid V_l^0 is more suited to problems with lower regularity in time than in space, whereas \tilde{V}_l^0 is more efficient for problems with the same regularity in space and time.

To derive estimates of the approximation rates of the different spaces, we need the following Lemma.

LEMMA 2.1. For $t \geq 0$ fixed we assume that for every $v^\Omega \in H^{2t}(\Omega)$, $v^\Omega = \sum_j w_j^\Omega$, $w_j^\Omega \in W_j^\Omega$,

$$\|v^\Omega\|_{H^{2t}}^2 \simeq \sum_j 2^{2(2t)j} \|w_j^\Omega\|_{L^2}^2 \quad (2.12)$$

and for every $v^T \in H^t((0, T))$, $v^T = \sum_j w_j^T$, $w_j^T \in W_j^T$

$$\|v^T\|_{H^t}^2 \simeq \sum_j 2^{2tj} \|w_j^T\|_{L^2}^2. \quad (2.13)$$

Then, for $u \in H_{mix}^{2t,t}(\Omega_T)$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, we have

$$\|u\|_{H_{mix}^{2t,t}(\Omega_T)}^2 \simeq \sum_{\mathbf{j}} 2^{2t\|(2j_1, j_2)\|_1} \|w_{\mathbf{j}}\|_{L^2(\Omega_T)}^2. \quad (2.14)$$

If we replace (2.12) by

$$\|v^\Omega\|_{H^t}^2 \simeq \sum_j 2^{2tj} \|w_j^\Omega\|_{L^2}^2 \quad (2.15)$$

we get, together with (2.13), for $u \in H_{mix}^{t,t}(\Omega_T)$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$,

$$\|u\|_{H_{mix}^{t,t}(\Omega_T)}^2 \simeq \sum_{\mathbf{j}} 2^{2t\|(j_1, j_2)\|_1} \|w_{\mathbf{j}}\|_{L^2(\Omega_T)}^2. \quad (2.16)$$

The detailed proof of these norm equivalences can be found in [30]. Here, we restrict ourselves to (2.14) to show that the estimates follow directly from a proposition about

additive subspace splittings taken from [31]. To this end we use the notation $\{V; a\}$ to denote a Hilbert space V which is equipped with the scalar-product $a(\cdot, \cdot)$. We now consider two Hilbert spaces H_1 and H_2 , and a collection of closed subspaces $V_{1i} \subset H_1$ and $V_{2i} \subset H_2$, $i \in \mathbb{N}$, such that topologically

$$H_1 = \sum_i V_{1i} \text{ and } H_2 = \sum_i V_{2i}.$$

For a sequence $b_{li}(\cdot, \cdot)$ of bilinear forms on V_{li} , $l = 1, 2$, we call $\{H_l; a_l\} = \sum_i \{V_{li}; b_{li}\}$ an additive subspace splitting which is stable if

$$a_l(u, u) \simeq \|u\|^2 \equiv \inf_{\substack{u_i \in V_{li}, \\ u = \sum_i u_i}} \sum_i b_{li}(u_i, u_i) \quad (2.17)$$

holds true. From [31] we have:

PROPOSITION 2.2. *If the splittings $\{H_l; a_l\} = \sum_i \{V_{li}; b_{li}\}$, $l = 1, 2$, are stable, then the tensor-product splitting*

$$\{H_1 \otimes H_2; a_1 \otimes a_2\} = \sum_{i_1} \sum_{i_2} \{V_{1i_1} \otimes V_{2i_2}; b_{1i_1} \otimes b_{2i_2}\}$$

is stable.

Now we can start the proof of the above Lemma 2.1.

Proof. Here, we only consider (2.14), the other norm equivalences can be derived analogously. For $u \in H_{mix}^{2t,t}(\Omega_T)$, we obtain with Proposition 2.2 the stability of

$$\{H^{2t}(\Omega) \otimes H^t((0, T)); \|\cdot\|_{H^{2t}} \otimes \|\cdot\|_{H^t}\} = \sum_j \{W_j; 2^{2t\|(2j_1, j_2)\|_1} \|\cdot\|_{L^2} \otimes \|\cdot\|_{L^2}\}$$

□

Now, we are in the position to state the approximation properties of the spaces $V_l^\infty, V_l^0, \tilde{V}_l^\infty$ and \tilde{V}_l^0 .

LEMMA 2.3. *Let us assume that (2.12) and (2.13) hold true for $s, t, t > s$. For $u \in H^{2t,t}(\Omega_T)$ we get*

$$\inf_{v \in V_l^\infty} \|u - v\|_{H^{2s,s}(\Omega_T)}^2 \leq c \cdot 2^{4(s-t)l} \|u\|_{H^{2t,t}(\Omega_T)}^2. \quad (2.18)$$

Moreover, if $u \in H_{mix}^{2t,t}(\Omega_T)$

$$\inf_{v \in V_l^0} \|u - v\|_{H^{2s,s}(\Omega_T)}^2 \leq c \cdot 2^{4(s-t)l} \|u\|_{H_{mix}^{2t,t}(\Omega_T)}^2 \quad (2.19)$$

holds true. Analogously, if $u \in H^{t,t}(\Omega_T)$ and (2.15) and (2.13) hold true for $s, t, t > s$, we have

$$\inf_{v \in \tilde{V}_l^\infty} \|u - v\|_{H^{s,s}(\Omega_T)}^2 \leq c \cdot 2^{2(s-t)l} \|u\|_{H^{t,t}(\Omega_T)}^2 \quad (2.20)$$

and in the case $u \in H_{mix}^{t,t}(\Omega_T)$ we have

$$\inf_{v \in \tilde{V}_l^0} \|u - v\|_{H^{s,s}(\Omega_T)}^2 \leq c \cdot 2^{2(s-t)l} \|u\|_{H_{mix}^{t,t}(\Omega_T)}^2. \quad (2.21)$$

Proof. We consider only the case (2.19). The other estimates follow in an analogous way. For $u \in H_{mix}^{2t,t}(\Omega_T)$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, $w_{\mathbf{j}} \in W_{\mathbf{j}}$, using the norm equivalences (2.12) and (2.13) we obtain

$$\begin{aligned}
\inf_{v \in V_l^0} \|u - v\|_{H^{2s,s}(\Omega_T)}^2 &\leq \|u - \sum_{2j_1+j_2 \leq 2l} w_{\mathbf{j}}\|_{H^{2s,s}(\Omega_T)}^2 \\
&\leq c \cdot \sum_{2j_1+j_2 > 2l} 2^{2s\|(2j_1,j_2)\|_{\infty}} \|w_{\mathbf{j}}\|_{L^2(\Omega_T)}^2 \\
&= c \cdot \sum_{2j_1+j_2 > 2l} 2^{2s\|(2j_1,j_2)\|_{\infty} - 2t\|(2j_1,j_2)\|_1} 2^{2t\|(2j_1,j_2)\|_1} \|w_{\mathbf{j}}\|_{L^2(\Omega_T)}^2 \\
&\leq c \cdot \max_{2j_1+j_2 > 2l} 2^{2s\|(2j_1,j_2)\|_{\infty} - 2t\|(2j_1,j_2)\|_1} \|u\|_{H_{mix}^{2t,t}(\Omega_T)}^2 \\
&\leq c \cdot 2^{4(s-t)l} \|u\|_{H_{mix}^{2t,t}(\Omega_T)}^2.
\end{aligned}$$

□

The above Lemma shows that the sparse grid spaces constructed with any multilevel basis which fulfills the norm equivalency (2.14) provide the same approximation rates for functions from $H_{mix}^{2t,t}(\Omega_T)$ as the full grid spaces. For a simple domain Ω which is (e.g. after a suitable differentiable mapping) a d -cube, there exist a variety of such bases like orthogonal wavelets, spline-wavelets, prewavelets, biorthogonal wavelets, lifting wavelets and similar constructions which are constructed from a mother function by translation and dilation, see [14, 16, 21, 22, 45] and the references cited therein. Also a multilevel generating system, i.e. the collection of finite element bases for V_j^{Ω} , $j \geq 0$, results in an estimate similar to (2.12) or (2.15).² The same holds for a domain Ω which is composed from simple non-overlapping subdomains. Here, a wavelet-type basis or a multilevel basis is employed within each subdomain.³ However, for a more complicated, general spatial domain Ω , the construction of a multilevel basis which fulfills norm equivalencies like (2.12) or (2.15) can be quite difficult and demanding or even impossible. For wavelet-like approaches on polygonal domains see [17, 20, 44, 43]. In view of the approximation properties, however, it is not really necessary that such norm equivalences are fulfilled. As the next Lemma shows, we will only need upper bounds for the hierarchic increments $w_{\mathbf{j}} \in W_{\mathbf{j}}$ of a function $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$ to obtain nearly the same rates of approximation.

LEMMA 2.4. *Let us assume that for $t > s \geq 0$ fixed there is a constant $c > 0$ such that*

$$\|w_{\mathbf{j}}\|_{H^{s,s}(\Omega_T)} \leq c 2^{s\|\mathbf{j}\|_{\infty} - t\|\mathbf{j}\|_1} \|u\|_{H_{mix}^{t,t}(\Omega_T)} \quad \forall \mathbf{j} \in \mathbb{N} \quad (2.23)$$

²Here, the non-uniqueness of the representation $v^{\Omega} = \sum_j v_j^{\Omega}$, $v_j^{\Omega} \in V_j^{\Omega}$ is compensated by taking the prewavelets of all possible multilevel representations into account, i.e. instead of (2.15) we would have

$$\|v^{\Omega}\|_{H^t}^2 \simeq \inf_{\substack{v^{\Omega} = \sum_j v_j^{\Omega}, \\ v_j^{\Omega} \in V_j^{\Omega}}} \sum_j 2^{2tj} \|w_j^{\Omega}\|_{L^2}^2. \quad (2.22)$$

³In these cases sparse grids can also be applied for the spatial discretization. This results in further substantial savings in cost provided that an additional smoothness prerequisite in space like the boundedness of the second mixed derivatives is fulfilled. For details see the survey article [13] and the references cited therein.

for every $u \in H_{mix}^{t,t}(\Omega_T)$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$. Then we have

$$\inf_{v \in \tilde{V}_l^0} \|u - v\|_{H^{s,s}(\Omega_T)}^2 \leq \begin{cases} c \cdot 2^{2(s-t)l} \cdot l^2 \|u\|_{H_{mix}^{t,t}(\Omega_T)}^2 & \text{for } s = 0, t > 0, \\ c \cdot 2^{2(s-t)l} \|u\|_{H_{mix}^{t,t}(\Omega_T)}^2 & \text{for } s > 0, t > 0. \end{cases} \quad (2.24)$$

This Lemma is a simple generalization of Theorem 4 in [36].

The above Lemma shows that it is possible to replace the norm equivalences (2.12) or (2.15) by simple upper bounds for the hierarchic increments of the type (2.23) while maintaining nearly the same approximation rates. As we will see in section 3 such upper bounds can easily be derived by tensor product arguments.

We are now interested in the dimension of the spaces $V_l^\infty, V_l^0, \tilde{V}_l^\infty$ and \tilde{V}_l^0 .

LEMMA 2.5. *With $\dim(W_j^\Omega) = O(2^{d \cdot j})$ and $\dim(W_j^T) = O(2^j)$ we obtain*

$$\dim(V_l^\infty) = O(2^{(d+2) \cdot l}), \quad (2.25)$$

$$\dim(V_l^0) = \begin{cases} O(2^{2dl}) & \text{for } d = 1, \\ O(2^{dl}) & \text{for } d = 2, \\ O(2^{d \cdot l}) & \text{for } d > 2, \end{cases} \quad (2.26)$$

$$\dim(\tilde{V}_l^\infty) = O(2^{(d+1) \cdot l}), \quad (2.27)$$

$$\dim(\tilde{V}_l^0) = \begin{cases} O(2^l) & \text{for } d = 1, \\ O(2^{d \cdot l}) & \text{for } d > 1. \end{cases} \quad (2.28)$$

Proof. Since the estimates (2.25) and (2.27) for the full grid spaces are well known and the derivation of (2.26) can be found in [30] we will only discuss (2.28). We have

$$\begin{aligned} \tilde{V}_l^0 &= \bigoplus_{\|\mathbf{j}\|_1 \leq l} W_{\mathbf{j}} = \bigoplus_{j_1 \leq l} \bigoplus_{j_2 \leq l-j_1} W_{\mathbf{j}} \\ &= \bigoplus_{j_1 \leq l} W_{j_1}^T \otimes \left(\bigoplus_{j_2 \leq l-j_1} W_{j_2}^\Omega \right) \\ &= \bigoplus_{j_1 \leq l} W_{j_1}^T \otimes V_{l-j_1}^\Omega. \end{aligned}$$

Therefore, we get

$$\begin{aligned} \dim(\tilde{V}_l^0) &= \sum_{j_1=0}^l \dim(W_{j_1}^T) \cdot \dim(V_{l-j_1}^\Omega) \\ &\leq c \sum_{j_1=0}^l 2^{j_1} \cdot 2^{d \cdot (l-j_1)} \\ &= c \cdot 2^{d \cdot l} \sum_{j_1=0}^l 2^{(1-d)j_1}, \end{aligned}$$

and (2.28) follows immediately. \square

This Lemma shows that the sparse grid spaces V_l^0 and \tilde{V}_l^0 have a lower dimension than the full grid spaces V_l^∞ and \tilde{V}_l^∞ . The overall complexity is reduced by one and two orders of magnitudes, respectively, where for the cases $d = 1$ and $d \leq 2$ the additional logarithmic factors l and l^{d-1} appear. The numerical results in section

5 furthermore indicate that \tilde{V}_l^0 has just about four times the number of degrees of freedom of that of the spatial multilevel basis alone, i.e.

$$\dim(\tilde{V}_l^0) \approx 4 \cdot \dim(V_l^\Omega).$$

Taking the approximation results from Lemma 2.3 into account, we see that under some mild additional regularity assumptions we can approximate space-time functions by means of sparse grid spaces with basically the same complexity as functions stemming from stationary problems. The additional order of complexity due to time is not more present and therefore the associated problems discussed in the introduction are circumvented. Thus our sparse grid space-time spaces provide a powerful tool for a variety of instationary PDEs and parabolic optimal control problems.

3. A Spatial Multilevel Basis. In this section we discuss the construction of a multilevel splitting and a simple hierarchical basis for the spatial discretization. In view of Lemma 2.4 it is sufficient to use a multilevel basis which just fulfills upper bounds of the type (2.23) within the sparse grid concept to obtain approximation properties similar to that of full grid spaces. To this end, we first focus on the hierarchical basis [51] constructed from a nested sequence of spatial Finite Element spaces in subsection 3.1. It turns out that, under suitable conditions on the sequence of the spatial Finite Element spaces, the use of the isotropic linear hierarchical basis in the spatial domain and the linear hierarchical basis in time results in sparse grid spaces which fulfill upper bounds of the type (2.23) for the L^2 - and the energy-norm. Therefore, Lemma 2.4 can be applied and thus estimates for the approximation rates of the resulting space-time sparse grids based on the spatial hierarchical basis are obtained.

But even for the most simple multiscale basis, i.e. the hierarchical basis, there is still the problem of the resolution of the boundary on all scales. The coarsest triangulation determines the domain from which the finer scale triangulations are derived by refinement. To overcome this problem we construct in subsection 3.2 a special spatial hierarchical basis from a Finite Element space given on the finest scale. Here, we determine so-called coarse grids by purely algebraic coarsening, i.e. following the approach of algebraic multigrid [27, 41, 48] we use the AMG-coarsening scheme to create sets of coarse grid unknowns. Then, to assign hierarchical basis functions to these points, we use prolongation operators between the respective coarse and fine grids which are derived from a least squares approach.

3.1. Multilevel subspace splitting, interpolation and hierarchical basis.

Let us assume that we have a sequence of nested triangulations/grids \mathcal{T}_j on the domain Ω , $j \in \mathbb{N}$, $\mathcal{T}_j \subset \mathcal{T}_{j+1}$, i.e. the set of the nodes \mathcal{N}_j of \mathcal{T}_j is a subset of the set of nodes \mathcal{N}_{j+1} of \mathcal{T}_{j+1} ,

$$\mathcal{N}_j \subset \mathcal{N}_{j+1}.$$

As usual, we denote the diameter of $T \in \mathcal{T}_j$ with $h(T)$ and define

$$\rho(T) := \sup\{\text{diam}(B) \mid B \subset T, B \text{ is a ball}\}.$$

Throughout this section we assume that \mathcal{T}_j is a sequence of regular triangulations, i.e. that there is a constant σ such that

$$\frac{h(T)}{\rho(T)} \leq \sigma, \text{ for all } j \text{ and } T \in \mathcal{T}_j$$

and that there is a constant $c > 0$ such that

$$h_j := \max_{T \in \mathcal{T}_j} h(T) \leq c2^{-j}.$$

We denote by V_j^Ω the associated Finite Element space of piecewise linear or d -linear functions, and define a mapping $P_j : C^0(\Omega) \rightarrow V_j^\Omega$. Then each $u_l \in V_l^\Omega$ has the representation

$$u_l = P_0 u_l + \sum_{j=0}^{l-1} (P_{j+1} u_l - P_j u_l), \quad (3.1)$$

with the increment spaces

$$W_{j+1}^\Omega := \text{range}(P_{j+1} - P_j) \quad (3.2)$$

and $W_0^\Omega := V_0^\Omega$. For example, if we choose $P_j u$ as the interpolant of $u \in C^0(\Omega)$ in V_j^Ω , i.e.

$$P_j u(x) = u(x) \text{ for all } x \in \mathcal{N}_j, \quad (3.3)$$

we obtain a splitting into increment spaces which are spanned by the hierarchical basis as described in [51]. For the choice $P_j u$ as the L^2 -projection of u onto V_j we would obtain an L^2 -orthogonal wavelet basis. Other choices of projection operators are discrete L^2 -projections, approximate L^2 -projections, quasi-interpolation or Clément-projections which all lead to a stable multilevel basis.

In the following we focus on the choice of P_j as the interpolation operator (3.3) onto V_j . Since we are dealing with a regular triangulation, the classical interpolation theory of Finite Elements spaces, c.f. [3, 15], shows that

$$\|u - P_j u\|_{L^2(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^2(\Omega)} \quad (3.4)$$

for all $u \in H^2(\Omega)$ with constant $c > 0$ independent of j . Therefore, using the hierarchical decomposition (3.1) we obtain for the parts w_j of the splitting $u = P_0 u + \sum_{j \in \mathbb{N}} (P_{j+1} u_l - P_j u_l) = \sum_{j \in \mathbb{N}} w_j$, $u \in H^2(\Omega)$, the estimate

$$\|w_j\|_{L^2(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^2(\Omega)} \quad (3.5)$$

with constant $c > 0$ independent of j . An analogous estimate follows directly for all $u \in H^2((0, T))$ with the linear interpolation operator and thus the linear hierarchical basis in time. Then, arguments on the tensor product of operators [32] lead for the hierarchical increment spaces $W_{\mathbf{j}} = W_{j_1}^\Omega \otimes W_{j_2}^T$ with the splitting $u = \sum_{\mathbf{j} \in \mathbb{N}^2} w_{\mathbf{j}}$ to the estimate

$$\|w_{\mathbf{j}}\|_{L^2(\Omega_T)} \leq c \cdot 2^{-2 \cdot |\mathbf{j}|_1} \|u\|_{H_{mix}^{2,2}(\Omega)} \quad \forall \mathbf{j} \in \mathbb{N}^2, \quad (3.6)$$

with c independent of \mathbf{j} and independent of $u \in H_{mix}^{2,2}(\Omega_T)$. Here, $W_{j_2}^T$ are the one-dimensional linear increment spaces in time which result analogously to (3.2) from the differences of piecewise linear interpolation operators in time.

Now we can apply Lemma 2.4 and obtain the following result on the approximation rate of the space-time sparse grid space using a linear hierarchical basis in time and the isotropic d -dimensional hierarchical basis in space.

LEMMA 3.1. *Let the sequence of Finite Element spaces V_j^Ω fulfill (3.4) with a constant c independent of j for all $u \in H^2(\Omega)$. Then, with the hierarchical basis increment spaces (3.2) and the linear hierarchical basis in time in the construction of the space-time sparse grid space \tilde{V}_l^0 , there holds*

$$\inf_{v \in \tilde{V}_l^0} \|u - v\|_{L^2(\Omega_T)} \leq c \cdot 2^{-2l} \cdot l \cdot \|u\|_{H_{mix}^{2,2}(\Omega_T)} \quad (3.7)$$

for all $u \in H_{mix}^{2,2}(\Omega_T)$.

We define the energy norm $\|\cdot\|_E$ for $u \in H^{1,1}(\Omega_T)$ as usual by

$$\|u\|_E^2 = \int_0^T \int_\Omega (\partial_t u)^2 + \sum_{i=1}^d (\partial_{x_i} u)^2 dx dt.$$

Furthermore, analogous to the definition of $H_{mix}^{2,2}(\Omega_T)$, we now define

$$H_{mix}^{2,2,\infty}(\Omega_T) := H^{2,\infty}(\Omega) \otimes H^{2,\infty}((0, T)),$$

where $H^{2,\infty}(\Omega)$ and $H^{2,\infty}((0, T))$ denote the classical Sobolev spaces of functions $u \in L^\infty(\Omega)$ and $u \in L^\infty((0, T))$ with weak derivatives up to the order 2 which are contained in $L^\infty(\Omega)$, respectively. We then can state the following Lemma:

LEMMA 3.2. *Let the family V_k of Finite Element spaces be affine equivalent. For \tilde{V}_l^0 constructed from the hierarchical increment spaces (3.2) and the one-dimensional hierarchical basis in time there holds*

$$\inf_{v \in \tilde{V}_l^0} \|u - v\|_E \leq c 2^{-l} \|u\|_{H_{mix}^{2,2,\infty}(\Omega_T)} \quad (3.8)$$

for all $u \in H_{mix}^{2,2,\infty}(\Omega_T)$ with a constant $c > 0$ independent of l .

For the proof of this Lemma, one has to verify condition (2.23). Then the application of Lemma 2.4 leads to the estimate (3.8). The verification of (2.23) is quite technical, we therefore postpone it to the appendix.

The above discussion shows that it is sufficient to have a sequence of grids and linear or d -linear Finite Element spaces which induce the hierarchical increment spaces in the space-time sparse grid construction to maintain the approximation order of a full grid space in the energy norm. For the L^2 -norm only an additional logarithmic factor comes into play. Now, the question is how such a sequence of grids and prolongation operators can be obtained. Quite a few Finite Elements packages meanwhile use geometric multigrid methods for the solution of the linear systems which result from the discretization of a partial differential equation. These codes then provide in one way or the other also access to interpolation operators P_j which we can exploit.

However, for complicated geometries and unstructured grids it is difficult or even impossible to construct a sequence of nested subspaces and a hierarchy of grids which are needed for geometric multigrid methods. Here, a main problem is the resolution of the domain on coarser levels which is one of the reasons algebraic multigrid methods (AMG) [4, 5, 6, 7, 41] were developed. Here, given a matrix A stemming from a Finite Element discretization, AMG computes (among other things) a sequence of coarse grids, i.e. a set of unknowns, and a sequence of prolongation operators. Now, taking a closer look at the construction (3.2) of the hierarchical increment spaces W_j^Ω , we find that geometric information about the grids \mathcal{T}_j is not necessary but only the prolongation operators P_j are needed. Therefore, AMG offers a solution to the

problem of the resolution of the geometry. We discuss in the next subsection how we use AMG ingredients to construct a sequence of prolongations which we will use in our space-time sparse grid construction.

3.2. Construction of Algebraic Prolongation Operators. In this subsection we will focus on the construction of a sequence of prolongation operators P_j needed for the construction (3.2) of the hierarchic increment spaces W_j . Here, prolongation operators as well as sets C_j of coarse grid points are algebraically constructed using only the entries of the system matrix A^l on the finest level of discretization.⁴

For most AMG methods, the setup phase, i.e. the phase where the prolongation operators and the coarse grid operators are computed, can be divided into two parts. In a first step, the coarse grid, i.e. a subset $C_j \subset N_j$ of the set of all unknowns N_j on level j , is chosen. Then, given the set of coarse grid points which form the new set of unknowns of the next coarser level, the prolongation $P_j = (p_{kl}^j)$ is computed. Here, one usually requires that each unknown associated to a fine grid point which is also contained in the coarser grid is interpolated by the value of the respective coarse grid unknown, i.e. after reordering the fine and coarse grid unknowns, the prolongation operator P_j has the block-matrix representation

$$P_j = \begin{pmatrix} I_j \\ \tilde{P}_j \end{pmatrix}, \quad (3.9)$$

$n_j := |N_j| = |C_{j-1}|$, $I_j \in \mathbb{R}^{n_{j-1} \times n_{j-1}}$ is the identity and $\tilde{P}_j \in \mathbb{R}^{(n_j - n_{j-1}) \times n_{j-1}}$.

For the coarsening process, i.e. the computation of C_j , we use the coarsening algorithm due to Ruge and Stüben as described in [41]. To derive \tilde{P}_j we used in a first attempt some classical interpolation formulas, the so-called standard interpolation and the so-called direct interpolation, see [48]. These formulas involve the entries a_{kl}^j of the matrix A^j within the interpolation formulas. Numerical experiments however showed that, especially on unstructured grids, the resulting prolongations did not provide satisfying approximation rates when used in the sparse grid context. The reason is that the prolongation formulas in AMG are developed to interpolate algebraic smooth errors as exact as possible, i.e. errors which cannot efficiently be reduced by a given stationary linear iteration scheme. This requirement however does not have to lead to good approximation results when used in space-time sparse grids.

Therefore we use a least squares approach to compute interpolation formulas which are able to reproduce (global) linear or d -linear functions as exact as possible. We denote by $\bar{u}^j \in \mathbb{R}^{n_j}$ the vector of coefficients of the basis representation of $u_j \in V_j^\Omega$. We choose a basis $\{\zeta_i\}$ of the space of *globally* linear or, alternatively, d -linear functions, i.e. $\{1, x_i, i = 1, \dots, d\}$ or $\{1, x_{i_1}, x_{i_1}x_{i_2}, \dots, x_{i_1} \cdots x_{i_d}, i_j = 1, \dots, d, i_1 < i_2, i_1 < i_2 < i_3, \dots, i_1 < i_2 < \dots < i_d\}$, and a norm $\|\cdot\|_\star$ on \mathbb{R}^{n_j} and define the functional $J : \mathbb{R}^{(n_j - n_{j-1}) \times n_{j-1}} \mapsto \mathbb{R}$ as

$$J(\tilde{P}_j) := \sum_{\zeta_i} \|\bar{\zeta}_i^j - (\tilde{P}_j \bar{\zeta}_i^{j-1})\|_\star^2. \quad (3.10)$$

⁴AMG needs to be given a fine grid stiffness matrix A^l to set up the interpolation operators and coarser grid matrices A^j via the Galerkin identity. But in our context we have no unique naturally given stiffness matrix A^l at our disposal. Since we are merely interested in the construction of a multilevel basis for the space part of a space-time sparse grid discretization by means of AMG, we used the stiffness matrix which stems from the discretization of the Laplacian in the given Finite Element space in the numerical experiments presented in section 5. For this choice AMG methods provide nearly the same hierarchy as geometric multigrid methods on structured uniform grids.

Now we compute the block \tilde{P}_j in (3.9) as the solution of the constrained minimization problem

$$\min_{\tilde{P}_j \in \mathbb{R}^{(n_j - n_{j-1}) \times n_{j-1}}} J(\tilde{P}_j). \quad (3.11)$$

In general, this would lead to densely populated operators $\tilde{P}_j = (\tilde{p}_{kl}^j)$. Therefore, we impose a second constraint on the above minimization problem, namely

$$a_{kl}^j = 0 \Rightarrow \tilde{p}_{k,l}^j = 0. \quad (3.12)$$

Loosely speaking this constraint enforces that only neighboring grid points are used in the prolongation process to fine grid points. Note that due to the constraint (3.12) on the sparsity pattern of the prolongation operator, the overall constraint minimization (3.11) can be split into local small least squares problems which can be efficiently solved by simple *LU*-factorization.

For the choice of $\|\cdot\|_*$ in (3.10) there are several possibilities. The norm induced by the coarse grid operators A^j which are set up from the AMG approach anyway, i.e.

$$\|x\|_{A^j} := (x^T A^j x)^{1/2}$$

would lead to cost functionals similar to the ones used for the construction of prolongation operators in the context of algebraic multigrid methods based on element interpolation (AMGe) [8]. In the following, we will use the Euclidean norm instead. It is easy to compute and provided good results in our numerical experiments. Nevertheless, future research is needed to study the influence of the choice of the norm $\|\cdot\|_*$ on the space-time sparse grid approach in more detail.

4. Data Structures and Adaptivity. In this section we will discuss the underlying data structure to store the space-time sparse grid coefficients. The aim is to obtain a storage format for the coefficients which allows for an efficient implementation of a sparse grid interpolant for given nodal values and vice versa, see [2, 42] for more details.

For classical sparse grids, there are several approaches [2, 12, 37, 42] on this topic which use hash maps or binary trees to manage the coefficients of the sparse grid basis. Note however that we are not dealing with classical sparse grids which are constructed by a tensor product from a purely one-dimensional multilevel basis. There, each basis function has a fixed number of hierarchical sons such that tree data structures can be efficiently applied. In our application the space part of the multilevel basis is allowed to stem from general unstructured grids instead. Therefore, the simple concept of recursively used binary trees is not efficiently applicable any longer. Instead we use a slightly different approach, i.e. binary trees of hash maps, which is described in the following. The associated data structures allow for an efficient implementation of the operations that have to be performed on a space-time sparse grid, like e.g. the computation of the function values of the interpolant [42]. They are also well suited for adaptive refinement procedures on space-time sparse grids.

4.1. Algorithmic Approach and Data Structures. Let $\mathcal{B}(V_l^T) = \{\psi_{j,i}^T\}$ be the piecewise linear hierarchical basis in time for V_l^T , i.e. $\psi_{j,i}^T \in W_j^T$, where the index j indicates the level, $\|\psi_{j,i}^T\|_\infty = 1$ and

$$\text{supp}(\psi_{j,i}^T) \cap \text{supp}(\psi_{j,k}^T) = \emptyset \text{ for all } i \neq k. \quad (4.1)$$

Furthermore, let $\mathcal{B}(V_l^\Omega) = \{\psi_{j,i}^\Omega\}$ be the spatial algebraic hierarchical basis constructed from a Finite Element space as described in section 3.2. Then, we obtain a basis of the space-time sparse grid space \tilde{V}_l^0 by

$$\mathcal{B}(\tilde{V}_l^0) := \{\psi_{\mathbf{j},\mathbf{i}} := \psi_{j_1,i_1}^\Omega \cdot \psi_{j_2,i_2}^T \mid j_1 + j_2 \leq l\}. \quad (4.2)$$

For the ease of presentation, we will use the set of multilevel indices $\mathcal{I}_{\tilde{V}_l^0}$ contained in the sparse grid \tilde{V}_l^0 ,

$$\mathcal{I}_{\tilde{V}_l^0} := \{(\mathbf{j}, \mathbf{i}) \mid j_1 + j_2 \leq l, 0 \leq i_1 \leq \dim(W_{j_1}^\Omega), 0 \leq i_2 \leq \dim(W_{j_2}^T)\}.$$

Furthermore, to each multilevel index (\mathbf{j}, \mathbf{i}) we associate a grid point $(x_{j_1,i_1}, t_{j_2,i_2})$ such that

$$\psi_{(\mathbf{j},\mathbf{i})}(x_{j_1,i_1}, t_{j_2,i_2}) = 1.$$

We define the usual hierarchical relation $>$ between two hierarchical indices (j_1, i_1) and (\bar{j}_1, \bar{i}_1) of the multilevel basis in space as

$$(j_1, i_1) > (\bar{j}_1, \bar{i}_1) :\Leftrightarrow \text{supp } \psi_{j_1, \bar{i}_1}^\Omega \subset \text{supp } \psi_{j_1, i_1}^\Omega \quad (4.3)$$

and we use the analogous definition for the hierarchical indices of the spatial multilevel basis. Now, given the nodal values $u(x_{j_1,i_1}, t_{j_2,i_2})$ of a continuous function u at the sparse grid points $(x_{j_1,i_1}, t_{j_2,i_2})$ we want to compute the coefficients $u_{\mathbf{j},\mathbf{i}}$ of the sparse grid interpolant u_l^{sp} ,

$$u_l^{sp} = \sum_{(\mathbf{j},\mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}} u_{\mathbf{j},\mathbf{i}} \psi_{\mathbf{j},\mathbf{i}}(x, t),$$

i.e.

$$u_l^{sp}(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2}) = \sum_{(\mathbf{j},\mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}} u_{\mathbf{j},\mathbf{i}} \psi_{\mathbf{j},\mathbf{i}}(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2}),$$

for all sparse grid points $(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2})$. Using (4.1) we obtain

$$\begin{aligned} u(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2}) &= \sum_{(\mathbf{j},\mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}} u_{\mathbf{j},\mathbf{i}} \psi_{\mathbf{j},\mathbf{i}}(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2}) \\ &= \sum_{\substack{(j_2, i_2) \geq (\bar{j}_2, \bar{i}_2) \\ (\mathbf{j}, \mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}}} \psi_{j_2, i_2}^T(t_{\bar{j}_2, \bar{i}_2}) \left(\sum_{\substack{j_1, i_1 \\ (\mathbf{j}, \mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}}} u_{\mathbf{j},\mathbf{i}} \psi_{j_1, i_1}^\Omega(x_{\bar{j}_1, \bar{i}_1}) \right). \end{aligned}$$

Therefore, to obtain the hierarchical coefficients we proceed in two steps. First, according to the second sum, we compute spatial multilevel coefficients $\hat{u}_{\mathbf{j},\mathbf{i}}$ such that for each time step $t_{\bar{j}_2, \bar{i}_2}$

$$u(x_{\bar{j}_1, \bar{i}_1}, t_{\bar{j}_2, \bar{i}_2}) = \sum_{\substack{j_1, i_1 \\ ((\bar{j}_1, \bar{j}_2), (\bar{i}_1, \bar{i}_2)) \in \mathcal{I}_{\tilde{V}_l^0}}} \hat{u}_{\mathbf{j},\mathbf{i}} \psi_{j_1, i_1}^\Omega(x_{\bar{j}_1, \bar{i}_1}), \quad \forall x_{\bar{j}_1, \bar{i}_1}, (\mathbf{j}, \mathbf{i}) \in \mathcal{I}_{\tilde{V}_l^0}. \quad (4.4)$$

Then we apply a basis transformation from the nodal basis to the multilevel basis in time to transform $\hat{u}_{j,i}$ to the sparse grid coefficients $u_{j,i}$. Here, according to the first sum which involves all hierarchical ancestors, this can be done by a simple top-down tree traversal if the coefficients $\hat{u}_{j,i}$ are stored in a binary tree.

Thus, this algorithm suggests the use of a binary tree for the multilevel basis in time, where each node contains an array for the coefficients which belong to the respective time step. However, we also want to be able to adaptively refine our space-time sparse grids, e.g. by starting with a regular sparse grid \tilde{V}_l^0 and locally inserting new grid points and associated basis functions according to certain error estimators or indicators. Therefore we use hash maps in each node instead of simple arrays. For a detailed discussion on the use of hash maps in the context of sparse grids and hierarchical bases see [42, 28]. For a general introduction we refer to [40]. Here, a hash table M is used to store the data and each data entry has to be uniquely identified by a key, for which we simply use the spatial hierarchical index (j_1, i_1) . Now, we need a hash function h to compute the position $h((j, i))$ where the associated coefficient is stored in the hash table M . In each node of the binary tree which represents the time level j_2 we use a hash table with $\dim(V_{l-j_2}^\Omega)$ entries and we define the hash map h by

$$h((j, i)) = (\dim(V_{j-1}^\Omega) + i) \mod \dim(V_{l-j_2}^\Omega),$$

where $V_0^\Omega := \emptyset$. Obviously, for a regular sparse grid \tilde{V}_l^0 , there are exactly $\dim(V_{l-j_2}^\Omega)$ data entries in each node which belong to the time level j_2 and the above hash function h provides a mapping without collision. However, for adaptively refined sparse grids, different coefficients can result in the same hash values. Therefore, we use so called *direct chaining hashing* [26] where all coefficients with the same hash value are stored in a separate list. Then, to obtain the coefficient value for a given spatial index, we have to search the list which belongs to the hash value of the key. Numerical experiments showed that these lists are usually very small, so that the computational cost of this search is bounded by a small constant.

4.2. Adaptivity. Sparse grids can easily be refined adaptively, c.f. [2, 9, 37]. This is necessary if the function to be approximated does not fulfill the smoothness requirements which are a prerequisite for regular sparse grids. This way singularities or strong variations in the function can be taken care of. It can be shown that the favorable approximation rates and complexities of regular sparse grids can be obtained also for non-smooth functions provided that the right adaptivity scheme with a proper error estimator is used. Here one can derive classical error estimators involving residual- or duality-based error indicators in a way similar to classical Finite Elements, see [10, 13]. An alternative is the use of the size of the coefficients in a wavelet-like representation as indicators for local refinement along the lines of [18, 19, 21]. It can be shown that, for any stable multiscale basis, such an approach provides true error estimators, i.e. error indicators which are reliable and efficient. Note that there is an easy way to switch from our hierarchical basis representation to such a wavelet-type representation by means of the lifting scheme [37, 45].

In the following we simply use the weighted size of the coefficients in the hierarchical basis representation as an error indicator for local refinement. This approach was successfully used in several different application areas, e.g. in visualization [24], in numerical integration [25] and for the solution of partial differential equation [2, 9, 13, 37]. It results in a reliable but in general not perfectly efficient error estimator.⁵ Never-

⁵Note that the upper bound (which implies reliability) in the norm equivalency for the hierarchical

theless, numerical results show that the proposed adaptivity criterion provides quite good results.

We proceed as follows: For a basis function $\psi_{\mathbf{j},\mathbf{i}} = \psi_{j_1,i_1}^T \cdot \psi_{j_2,i_2}^\Omega$ we define the set $\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}})$ of hierarchical sons, i.e.

$$\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}}) := \{\psi_{\tilde{\mathbf{j}},\tilde{\mathbf{i}}} \mid (\tilde{j}_1, \tilde{j}_2) = (j_1 + 1, j_2) \text{ or } (\tilde{j}_1, \tilde{j}_2) = (j_1, j_2 + 1), \quad (4.5)$$

$$\psi_{j_1,i_1}^T > \psi_{\tilde{j}_1,\tilde{i}_1}^T \text{ and } \psi_{j_2,i_2}^\Omega > \psi_{\tilde{j}_2,\tilde{i}_2}^\Omega\} \quad (4.6)$$

using the $>$ -relation from (4.3). Now, given the sparse grid approximation $u^{sp} = \sum_{(\mathbf{j},\mathbf{i}) \in \Lambda} u_{\mathbf{j},\mathbf{i}} \cdot \psi_{\mathbf{j},\mathbf{i}}$ where Λ denotes the set of involved indices, we add $\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}})$ to the actual sparse grid space⁶ and thus the indices $(\tilde{\mathbf{j}}, \tilde{\mathbf{i}})$ of the hierarchical sons to the actual set Λ whenever

$$|u_{\mathbf{j},\mathbf{i}}| > 2^{\frac{1}{2}(d \cdot j_1 + j_2)} \cdot \varepsilon, \quad (4.7)$$

for a given threshold ε . Here, the factor $\frac{1}{2}$ in the exponent resembles a measurement of the error in the L^2 -norm, for other norms, appropriate values have to be substituted. The term $d \cdot j_1 + j_2$ corresponds to an approximation scheme with the same order in space and time. Such a refinement strategy leads asymptotically to an equilibration of the error indicators in the sense of [1].

5. Numerical Results. In this section we present some numerical results concerning the convergence rates of the space-time sparse grid interpolant in \tilde{V}_l^0 for the regular case and the adaptive refinement case. To this end, we use a bilinear Finite Element space V_l^Ω , where $h := 2^{-l} > 0$ indicates the mesh width of the underlying grid, on the spatial domain $\Omega \subset \mathbb{R}^2$ to construct the spatial multilevel basis as described in section 3. Here we use the stiffness matrix of the Laplacian in the AMG coarsening algorithm.

For $u \in C^0(\Omega_T)$ given, we set $h := 2^{-l}$ and denote by u_h the full grid interpolant of u and by u_l^{sp} the sparse grid interpolant. We define the interpolation errors $e_h := u - u_h$ and $e_l^{sp} := u - u_l^{sp}$. Then obviously

$$\|e_l^{sp}\| \leq \|e_h\| + \|u_h - u_l^{sp}\|, \quad (5.1)$$

for any given norm $\|\cdot\|$. In the following, we will use the norms

$$\begin{aligned} \|u\|_{L^2 \times L^\infty} &:= \text{ess sup}_{t \in (0,T)} \|u(\cdot, t)\|_{L^2(\Omega)}, \\ \|u\|_{L^2 \times L^2} &:= \left(\int_{0 < t < T} \|u(\cdot, t)\|_{L^2(\Omega)}^2 dt \right)^{1/2}, \\ \|u\|_{H^1 \times L^\infty} &:= \text{ess sup}_{t \in (0,T)} \|u(\cdot, t)\|_{H^1(\Omega)}, \\ \|u\|_{H^1 \times L^2} &:= \left(\int_{0 < t < T} \|u(\cdot, t)\|_{H^1(\Omega)}^2 dt \right)^{1/2}. \end{aligned}$$

For these norms, classical interpolation theory shows that the error e_h decreases with at least the same order as the error of the sparse grid interpolant e_l^{sp} for decreasing

basis is (after proper weighting) independent of the number of levels whereas the lower bound (which relates to efficiency) slightly depends on the number of levels involved.

⁶Of course, this is only done for the indices which are 'leaves' of the structure of Λ , i.e. for the indices $(\mathbf{j}, \mathbf{i}) \in \Lambda$ with $\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}}) \not\subset \Lambda$.

TABLE 5.1

Degrees of freedom of the space-time sparse grid \tilde{V}_l^0 and the interpolation error $\tilde{e}_h := u_l^{sp} - u_h$ in different norms for the continuous function $u(x, t) = t^2 \cdot (x_1^2 + x_2^2)$.

level	spatial DOF	\tilde{V}_l^0 DOF	$\ \tilde{e}_h\ _{L^2 \times L^\infty}$	$\ \tilde{e}_h\ _{L^2 \times L^2}$	$\ \tilde{e}_h\ _{H^1 \times L^\infty}$	$\ \tilde{e}_h\ _{H^1 \times L^2}$
5	1089	3424	2.21e-4	1.64e-4	2.87e-3	2.32e-3
6	4225	13120	6.90e-05	5.10e-05	1.44e-3	1.16e-3
7	16641	51072	2.06e-05	1.53e-05	7.21e-4	5.82e-4
8	66049	200960	6.01e-06	4.44e-06	3.61e-4	2.91e-4
9	263169	796160	1.71e-06	1.27e-06	1.80e-4	1.46e-4

mesh width. We are interested if the orders predicted by our theory in section 2 for the geometrically constructed hierarchical basis can also be achieved for the algebraically constructed increment spaces described in section 3.2. With (5.1) we see that it is sufficient to use $\tilde{e}_h := \|u_h - u_l^{sp}\|$ which is easier to compute. Therefore, we will present results in terms of \tilde{e}_h only.

In all the following experiments we consider $T = 1.0$.

Example 1. In our first example we examine the interpolation error for the space-time sparse grid space \tilde{V}_l^0 for the function

$$u(x, t) = t^2 \cdot (x_1^2 + x_2^2)$$

on the unit square $\Omega = [0, 1]^2$. This function is just a product of a quadratic polynomial in time and in space. Here it serves as the most simple model problem to discuss the basic properties of our space-time sparse grid approach. We use a uniform rectangular grid with mesh width $h = 2^{-l}$ for the underlying Finite Element discretization from which the spatial multilevel basis is derived along the lines of subsection 3.2. In this case, the AMG coarsening algorithm of Ruge and Stüben leads to a sequence of coarse grids which is nearly identical to conventional geometric coarsening. Therefore, the constructed spatial multilevel basis is basically the usual isotropic hierarchical basis [51]. Note that the function is zero at the time point $t = 0$ and therefore all degrees of freedom at this time point can be eliminated.

Table 5.1 shows the error as well as the dimension of the space-time sparse grid space. First of all note that the number of degrees of freedom of the space-time sparse grid space \tilde{V}_l^0 increases with the same order, i.e. $O(h^{-2})$, as the degrees of freedom of the spatial grid as expected from (2.28) of Lemma 2.5. Moreover the additional constant amount for the space-time sparse grid in comparison to the spatial grid is just a factor of three where we eliminated the degrees of freedom (dof) for $t = 0$ (it would be four including those dofs). Figure 5.1 shows the resulting space-time sparse grid of level $l = 5$. Here, one can clearly see the finest grids with $h = 1/32$ at the time points $t = 0.5$ and $t = 1.0$ which dominate the overall number of degrees of freedom.

Furthermore, from Lemma 3.1 in section 3 we expect a behavior of $O(2^{-2l})$ ($O(h^2 \log h)$) for the error in the $L^2 \times L^2$ -norm on level l ($h = 2^{-l}$). In fact, looking at the results in Table 5.1 we see that the quotient of the errors \tilde{e}_h of two successive levels of discretization measured in $\|\cdot\|_{L^2 \times L^2}$ and $\|\cdot\|_{L^\infty \times L^2}$ slowly approaches 0.25 from above for decreasing mesh width. This value would be the rate achieved with a full grid discretization. In more detail, from level $l = 5$ to level $l = 6$ the error is reduced by a factor of about 0.31 and decreases for growing l . We finally obtain a factor of about 0.28 for the reduction from $l = 8$ to $l = 9$. This difference to the value 0.25 accounts for the additional $\log h$ -term in the convergence rate which becomes more and more unimportant with growing number of levels in comparison to the

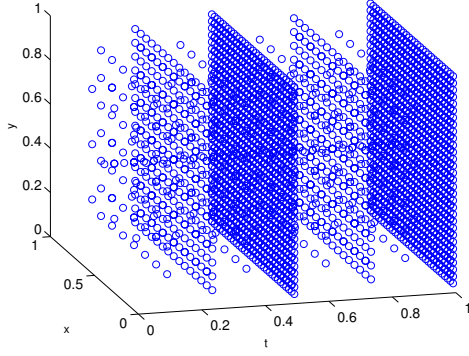


FIGURE 5.1. Space-time sparse grid associated to \tilde{V}_l^0 with two spatial dimensions for level $l = 5$ ($h = 1/32$).

exponential term. For the quotient of the error estimate of two consecutive levels we obtain $0.25 \cdot (l+1)/l$ which is also an upper bound for the reduction factor. Moreover, using this formula, we get a theoretical estimate of about 0.28 for the reduction of the error from level $l = 8$ to $l = 9$ which is exactly the rate we observe in our experiment.

From Lemma 3.2 we would expect the same error reduction rate for the sparse grid interpolant as for the full grid interpolant if the H^1 -norm is involved. Indeed, looking at the results in the $H^1 \times L^2$ - and $H^1 \times L^\infty$ -norms we observe a factor of 0.5 independent of the number of levels which is just the same as that of a full grid interpolant.

Example 2 . We now present results for the more complicated function

$$u(x, t) = \sin(4\pi t x_1) + \sin(4\pi t x_2)$$

on the unit square, $\Omega = (0, 1)^2$. Here, $u(x, t)$ is zero for $t = 0$ but develops more and more oscillations for $t \rightarrow 1.0$. A plot of the function at different time points is shown in Figure 5.2

Although predicted by the theoretical results given in Lemma 3.1 and 3.2, it is not directly obvious that the sparse grid approach still works well in practice for this case due to the oscillations in the function $u(x, t)$. Nevertheless, looking at the results in Table 5.2, we see that the behavior of the error measured in the $\|\cdot\|_{L^2 \times L^2}$ and $\|\cdot\|_{L^2 \times L^\infty}$ norms for increasing level number is the same as in the previous example. Here, the initial reduction rate from level $l = 5$ to $l = 6$ for the $\|\cdot\|_{L^2 \times L^2}$ norm amounts to about 0.35 which is a little bit larger than in the previous example. On level $l = 9$ we obtain a reduction rate of about 0.30.

For the two norms involving the H^1 -seminorm we expect from Lemma 3.2 an error reduction rate of about 0.5. In fact, the results of Table 5.2 clearly show that we achieve this predicted approximation rate.

Example 3 . In this example we show results for the proposed space-time sparse grid method for a slightly more complex geometry in space. We again consider the function

$$u(x, t) = t^2 \cdot (x_1^2 + x_2^2) \tag{5.2}$$

but choose now the unit circle

$$\Omega = \{x \mid \|x\| \leq 1.0\}$$

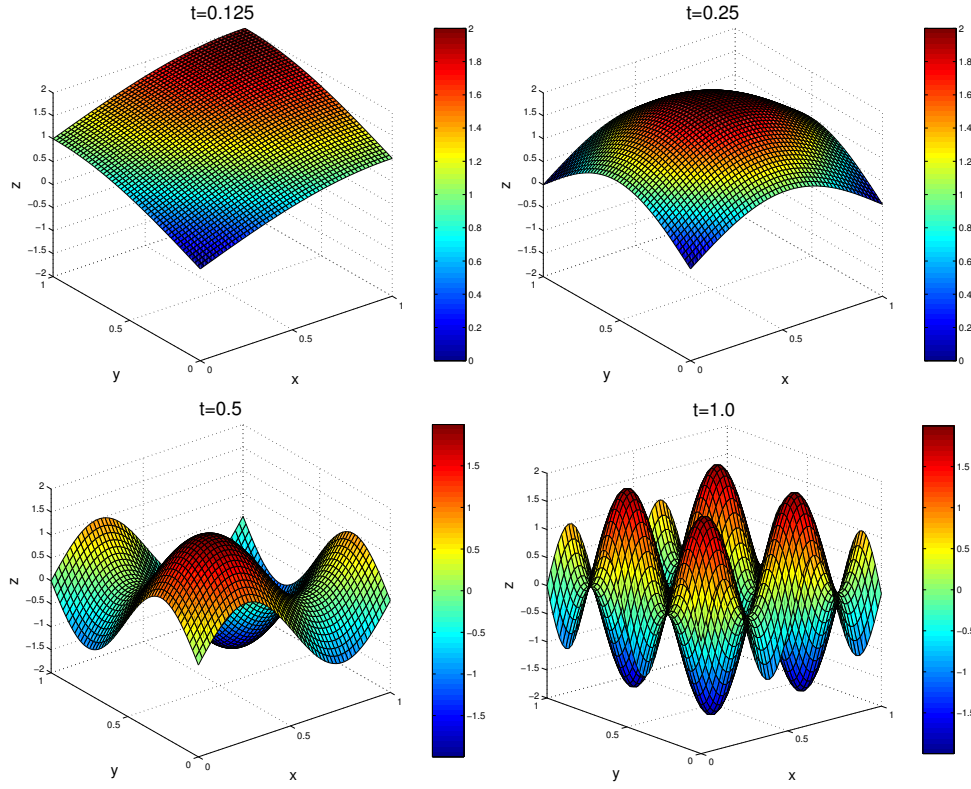


FIGURE 5.2. Plots of the function $u(x, t) = \sin(4\pi tx_1) + \sin(4\pi tx_2)$ of example 2 at time points $t = 0.125, 0.25, 0.5, 1.0$.

TABLE 5.2

Degrees of freedom of the space-time sparse grid \tilde{V}_l^0 and the interpolation error $\tilde{e}_h := u_l^{sp} - u_h$ in different norms for the continuous function $u(x, t) = \sin(4 \cdot \pi \cdot t \cdot x_1) + \sin(4 \cdot \pi \cdot t \cdot x_2)$.

level	spatial DOF	\tilde{V}_l^0 DOF	$\ \tilde{e}_h\ _{L^2 \times L^\infty}$	$\ \tilde{e}_h\ _{L^2 \times L^2}$	$\ \tilde{e}_h\ _{H^1 \times L^\infty}$	$\ \tilde{e}_h\ _{H^1 \times L^2}$
5	1089	3424	2.67e-2	1.34e-2	4.05e-1	2.25e-1
6	4225	13120	9.83e-3	4.66e-3	2.09e-1	1.16e-1
7	16641	51072	3.31e-3	1.51e-3	1.04e-1	5.87e-2
8	66049	200960	1.03e-3	4.67e-4	5.24e-2	2.95e-2
9	263169	796160	3.10e-4	1.39e-4	2.63e-3	1.48e-2

as spatial domain. Note that for such a domain a sequence of geometric discretizations on different scales with piecewise linear finite elements no longer results in nested Finite Element spaces. This is due to the circular boundary which only is linearly approximated on each level. Here, the algebraic coarsening by AMG sets in. Together with the prolongation operator constructed by (3.11) and the constraint (3.12), we obtain an algebraic hierarchical basis with good approximation properties also near the boundary. An example of a grid which is used for the finite element discretization of the Laplacian to obtain the initial system matrix for the AMG approach is given in Figure 5.3 (left), an associated space-time sparse grid is depicted in Figure 5.3 (right).

The results for three different space-time sparse grids are shown in Table 5.3. For

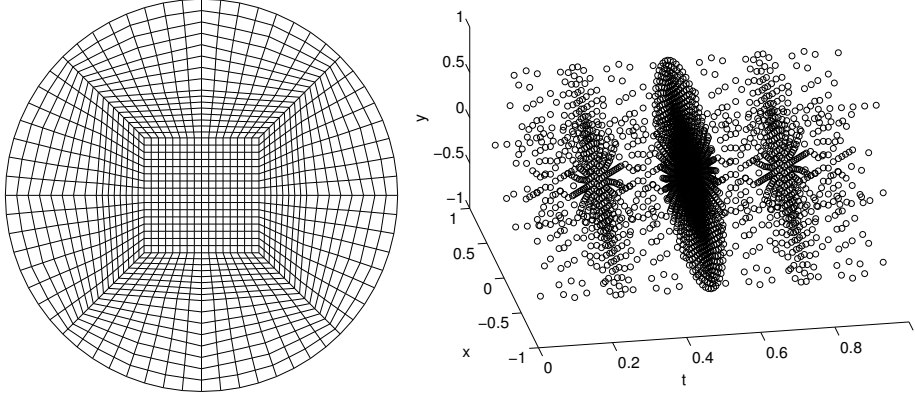


FIGURE 5.3. A spatial grid for the unit circle (left) and the space-time sparse grid (right).

TABLE 5.3

Degrees of freedom of the spatial grid, the space-time sparse grid and the interpolation error $\tilde{e}_h := u_{sp} - u_h$ in different norms for $u(x, t) = t^2 \cdot (x_1^2 + x_2^2)$ on the domain $\Omega = \{x | \|x\| \leq 1.0\}$.

spatial DOF	overall DOF	$\ \tilde{e}_h\ _{L^2 \times L^\infty}$	$\ \tilde{e}_h\ _{L^2 \times L^2}$	$ \tilde{e}_h _{H^1 \times L^\infty}$	$ \tilde{e}_h _{H^1 \times L^2}$
5185	10203	5.91e-4	4.67e-4	1.14e-2	9.22e-4
20609	40747	1.61e-4	1.23e-4	5.61e-3	4.53e-4
82177	162907	6.19e-05	4.81e-05	2.85e-3	2.32e-4

the $H^1 \times L^2$ - and the $H^1 \times L^\infty$ -norms we obtain nearly the same approximation rate of about 0.5 as in the previous examples. For the $L^2 \times L^2$ - and the $L^2 \times L^\infty$ -norms the error is reduced by the factors 0.27 and 0.38, respectively. Here, the grids used for the construction of the spatial Finite Element spaces vary in their quality depending on their mesh width, i.e. the maximal and minimal interior angles of the rectangles vary between the different grids. This might influence the approximation property of the space-time sparse grid and, consequently, the error reduction rate to some extent.

Example 4. In this example we present numerical results for the function

$$u(x, t) = \sqrt{t \cdot x_1} \cdot x_2$$

which has a singularity on the edge $x_1 = 0$ for all times $t > 0$ and a singularity at $t = 0$. Note that this function is no longer in $H_{mix}^{2,2}(\Omega_T)$ as the functions of the previous examples and we therefore expect a decrease of the approximation rates. Note furthermore that we would also obtain worse approximation results for a full grid approximation in space and time, since u is no longer in $H^{2,2}(\Omega_T)$ either. Therefore, using the error \tilde{e}_h instead of e_l^{sp} does no longer provide reliable bounds for the space time sparse grid approximation rates. Instead we use a very fine full grid ($l = 11$) as reference solution, interpolate the adaptive sparse grid solution to this grid and measure the error between the full grid and the sparse grid solution there.

To cope with the singularities in $u(x, t)$ we adaptively refine our sparse grids in space and time as described in section 4.2. We use the error indicator (4.7) with the thresholds $\varepsilon = 0.001$ and $\varepsilon = 0.005$. Figure 5.4 shows the error history in the $L^2 \times L^2$ -norm of the adaptive cycle when we apply the adaptive refinement starting on level $l = 4$ with the two different thresholds. For comparison the convergence result for regular sparse grids is given as well. We clearly see that the uniform refinement provides only a poor cost-benefit ratio. Here, starting with an error of about 0.02 for

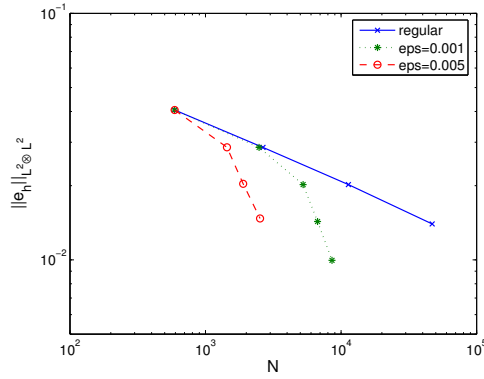


FIGURE 5.4. Interpolation error in the $\|\cdot\|_{L^2 \times L^2}$ -norm for the function $\sqrt{(t \cdot x_1)} \cdot x_2$, regular sparse grids for different levels and adaptive refinement starting from level $l = 4$ with thresholds $\varepsilon = 0.001$ and $\varepsilon = 0.005$ versus overall degrees of freedom (N).

level $l = 4$ (i.e. 592 dof) the space-time sparse grid results in an error of 0.0024 on level $l = 8$ (i.e. 46976 dof). This leads to an error reduction rate with respect to the number of degrees of freedom N of 0.48, i.e. the error behaves as $O(N^{-0.48})$ which can be seen from the slope of the respective curve in Figure 5.4. Remember that for a smooth solution⁷ we would obtain a behavior of nearly $O(N^{-1})$. Now, looking at the results for the adaptive cycles we see that we obtain a much higher accuracy with substantially less degrees of freedom. In the case of $\varepsilon = 0.001$ we obtain a final error of 0.0015 with only 8578 dof and the use of $\varepsilon = 0.005$ results in a final error of 0.0038 with 3286 dof. Now, if we look at the cost-benefit ratio, i.e. if we compare the error and the dof of the initial grid on level $l = 4$ and the grids obtained at the end of the adaptive cycles, we see that we get a rate of about 1.0 for both thresholds which is exactly the rate we would expect for a regular sparse grid and a smooth function. In Figure 5.5, we show the adaptive grid which results from the refinement process for $\varepsilon = 0.001$. Here we see that the non-smoothness in space and time direction is detected by the error indicator and gets resolved by the insertion of additional points in *time-space*. This indicates that the proposed error indicator works well.

6. Conclusion. We presented a space-time sparse grid method for the approximation of functions which depend on space and time. It involves the tensor product of a multilevel basis in space and a multilevel basis in time. A proper truncation of the resulting series expansion then gives a space-time sparse grid. This approach reduces the number of degrees of freedom significantly in comparison to a conventional space-time Finite Element method.

Furthermore we proposed a method to algebraically compute the spatial multilevel basis from a given Finite Element space. It involves algebraic multigrid and special prolongation operators which are determined by a local minimization procedure. This allows to handle also complex spatial geometries which pose a problem for classical sparse grid spaces. In the present paper we utilized a simple least-squares approach

⁷In the previous examples the degrees of freedom grew roughly with a factor of 4 from level to level and we gave the quotient of the error on two successive levels as reduction rate. Now we directly give the exponent of the error as a function of degrees of freedom which is more appropriate in the adaptive case.

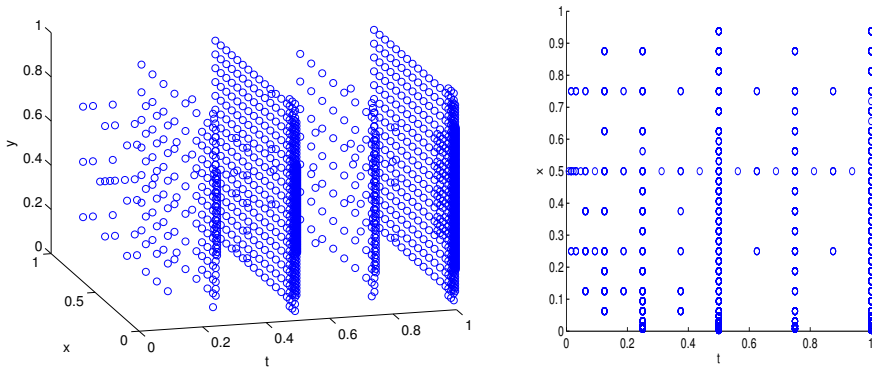


FIGURE 5.5. *Adaptively refined space-time grid ($\varepsilon = 0.001$) after 3 refinement steps for the function $u(x, t) = \sqrt{(t \cdot x_1)} \cdot x_2$.*

to guarantee approximation property of the coarse spaces. Other AMG interpolation procedures (i.e., [49] and [50]) are also feasible and worth pursuing in the future.

We showed that these space-time sparse grid spaces possess nearly the same approximation rates as classical Finite Element spaces if just slightly stronger regularity assumptions are fulfilled. Furthermore it turned out that their number of degrees of freedoms is of the same order as that of the spatial Finite Element space alone, i.e. the dimension of the space-time sparse grid spaces is by one order of magnitude smaller than the dimension of conventional space-time Finite Element spaces. In addition numerical experiments showed that the approximation rates predicted by the theory can really be achieved in practice, even for complicated spatial geometries which do not allow for a sequence of nested finite element spaces. We also demonstrated for a problem with singularities in space and time that space-time adaptivity can easily be implemented and used. This results in adaptively refined space-time sparse grids which resolve singularities, and thus allow to regain the superior convergence rates and complexities of space-time sparse grids also in the case of non-smooth functions.

In this article we restricted ourselves to the hierarchical basis in space and in time as a special case of a multiscale basis. Note however that our approach is by no means confined to this type of basis. It works in the same (or even better) way for any stable multiscale basis, like wavelets or frames. However, to get rid of the relatively tight restriction on the spatial domain imposed by wavelet-like schemes due to translation and dilation we opted for an algebraic multigrid construction with associated algebraic hierarchical basis instead.

The proposed space-time sparse grid spaces can be efficiently used for the discretization of parabolic problems and related time-dependent control and optimization problems. To this end it is necessary to discretize and to apply the respective differential operator, e.g. in weak formulation, in space and time. The associated computational work should be proportional to the number of degrees of freedom. This is easily possible whenever the coefficient functions of the operator which belong to the spatial derivatives are constant in time. Then an algorithm which is based on the so-called unidirectional principle [9, 13] can be applied. Also the efficient solution of the arising linear system in optimal complexity is necessary by e.g. a multilevel method which must be able to tackle the space-time discretized sparse grid problem directly. This however is future work.

7. Appendix. In this section, we give the proof of Lemma 3.2. Since the application of Lemma 2.4 shows the estimate, we just have to verify that inequality (2.23) holds true.

In the following, we denote a nodal basis of the spatial Finite Element space V_j^Ω by $\varphi_{j,i}^\Omega$, $1 \leq i \leq \dim(V_j^\Omega)$. Analogously, the hierarchical increment space W_j^Ω is spanned by a subset of the nodal basis of V_j^Ω . For the ease of presentation we denote the basis elements of W_j^Ω by $\psi_{j,i}^\Omega$, $1 \leq i \leq \dim(W_j^\Omega)$. Furthermore, (after reordering of the nodal set \mathcal{N}_j) we assume that the index i of the function $\psi_{j,i}^\Omega$ is associated to the nodal point where the basis function is equal to one, i.e. for every $n_j \in \mathcal{N}_j$ we have

$$\psi_{j,i}^\Omega(n_k) = \begin{cases} 1 & \text{if } k = i, \\ 0 & \text{otherwise} \end{cases}.$$

Using the definition (3.2) of W_j^Ω and the interpolation property of the prolongation operators P_j , we obtain the coefficients $w_{j,i}^\Omega$ of the basis representation $w_j^\Omega = \sum_{i=1}^{\dim(W_j)} w_{j,i}^\Omega \psi_{j,i}^\Omega$ of an element w_j^Ω , $w_j^\Omega \in W_j^\Omega$, by the formula

$$w_{j,i}^\Omega = (P_j u - P_{j-1} u)(n_i), \quad (7.1)$$

where $n_i \in \mathcal{N}_j$ is the nodal point with $\psi_{j,i}^\Omega(n_i) = 1.0$. Classical interpolation properties of Finite Element spaces, c.f. [15], show that

$$\|P_j u - P_{j-1} u\|_{L^\infty(\Omega)} \leq \|P_j u - u\|_{L^\infty(\Omega)} + \|u - P_{j-1} u\|_{L^\infty(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^{2,\infty}},$$

and relation (7.1) leads to

$$|w_{j,i}^\Omega| \leq \|P_j u - P_{j-1} u\|_{L^\infty(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^{2,\infty}(\Omega)}, \quad (7.2)$$

with $c > 0$ independent of u , j and i . We denote by $\psi_{j,i}^T$ an element of the one-dimensional hierarchical basis on $(0, T)$ and derive the analogous estimate

$$|w_{j,i}^T| \leq c \cdot 2^{-2j} \|u\|_{H^{2,\infty}((0,T))} \quad (7.3)$$

for $u \in H^{2,\infty}((0, T))$, with $u = \sum_j \sum_i w_{j,i}^T \psi_{j,i}^T$ and $c > 0$ independent of u , j or i .

A basis of the increment space $W_{\mathbf{j}}$, $\mathbf{j} \in \mathbb{N}^2$, is given by $\{\psi_{\mathbf{j},\mathbf{i}}\}$ where $\psi_{\mathbf{j},\mathbf{i}} := \psi_{j_1,i_1}^\Omega \cdot \psi_{j_2,i_2}^T$ and we obtain the basis representation

$$w_{\mathbf{j}} = \sum_{\mathbf{i}} w_{\mathbf{j},\mathbf{i}} \psi_{\mathbf{j},\mathbf{i}} \quad (7.4)$$

for a function $w_{\mathbf{j}} \in W_{\mathbf{j}}$. Using (7.1) and (7.3) and a tensor product argument, c.f. [32], we obtain for $u \in H_{mix}^{2,2,\infty}(\Omega_T)$, $u = \sum_{\mathbf{j},\mathbf{i}} w_{\mathbf{j},\mathbf{i}} \psi_{\mathbf{j},\mathbf{i}}$, the estimate

$$|w_{\mathbf{j},\mathbf{i}}| \leq c \cdot 2^{-2\|\mathbf{j}\|_1} |u|_{H_{mix}^{2,2,\infty}(\Omega)}$$

with $c > 0$ independent of u , \mathbf{j} and \mathbf{i} .

Now we want to derive an estimate for the term $\|\psi_{\mathbf{j},\mathbf{i}}\|_E$. The tensor product structure of $\psi_{\mathbf{j},\mathbf{i}}$ leads to

$$\|\psi_{\mathbf{j},\mathbf{i}}\|_E = \|\psi_{j_1,i_1}^\Omega\|_{L^2(\Omega)} \cdot \|\psi_{j_2,i_2}^T\|_{E,(0,T)} + \|\psi_{j_1,i_1}^\Omega\|_{E,\Omega} \cdot \|\psi_{j_2,i_2}^T\|_{L^2((0,T))} \quad (7.5)$$

Since we assume that the Finite Element spaces V_j^Ω are affine equivalent and we are dealing with a regular triangulation, classical Finite Element theory, c.f. [15], leads to

$$\|\psi_{j_1, i_1}^\Omega\|_{E, \Omega} \leq c 2^{-(d-2) \cdot (j_1/2)}, \quad (7.6)$$

$$\|\psi_{j_1, i_1}^\Omega\|_{L^2(\Omega)} \leq c 2^{-dj_1/2}. \quad (7.7)$$

For the one-dimensional hierarchical basis a straightforward calculation (see [13]) gives

$$\|\psi_{j_2, i_2}^T\|_{E, (0, T)} \leq c 2^{j_2/2}, \quad (7.8)$$

$$\|\psi_{j_2, i_2}^T\|_{L^2((0, T))} \leq c 2^{-j_2/2}. \quad (7.9)$$

This, together with (7.5), leads to

$$\|\psi_{\mathbf{j}, \mathbf{i}}\|_E \leq c \left(2^{-(d-2) \cdot (j_1/2)} \cdot 2^{-j_2/2} + 2^{-dj_1/2} \cdot 2^{j_2/2} \right) \quad (7.10)$$

$$\leq c 2^{-dj_1/2} 2^{-j_2/2} (2^{j_1} + 2^{j_2}) \quad (7.11)$$

$$\leq c 2^{-dj_1/2} 2^{-j_2/2} 2^{\|\mathbf{j}\|_\infty}. \quad (7.12)$$

Now, for $u \in H_{mix}^{2,2,\infty}(\Omega_T)$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, we obtain

$$\begin{aligned} \|w_{\mathbf{j}}\|_{E, \Omega_T}^2 &\leq \sum_{\mathbf{i}} w_{\mathbf{j}, \mathbf{i}}^2 \|\psi_{\mathbf{j}, \mathbf{i}}\|_{E, \Omega_T}^2 \\ &\leq \sum_{\mathbf{i}} 2^{-4\|\mathbf{j}\|_1} \cdot \|u\|_{H_{mix}^{2,2,\infty}(\Omega_T)}^2 \cdot \|\psi_{\mathbf{j}, \mathbf{i}}\|_{E, \Omega_T}^2 \\ &\leq c 2^{-4\|\mathbf{j}\|_1 + 2\|\mathbf{j}\|_\infty} \cdot 2^{-dj_1 - j_2} \sum_{\mathbf{i}} \|u\|_{H_{mix}^{2,2,\infty}(\Omega_T)}^2 \\ &\leq c 2^{-4\|\mathbf{j}\|_1 + 2\|\mathbf{j}\|_\infty} \|u\|_{H_{mix}^{2,2,\infty}(\Omega_T)}^2. \end{aligned}$$

Therefore, we can apply Lemma 2.4 which then shows Lemma 3.2.

REFERENCES

- [1] I. BABUSKA AND W. RHEINBOLDT, *Estimates for adaptive finite element computations*, SIAM J. Numer. Anal., (1978), pp. 736–754.
- [2] R. BALDER, *Adaptive Verfahren für elliptische und parabolische Differentialgleichungen*, PhD thesis, Technische Universität München, 1994.
- [3] D. BRAESS, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, 2001.
- [4] A. BRANDT, *Algebraic multigrid theory: The symmetric case*, in Preliminary Proceedings for the International Multigrid Conference, Copper Mountain, Colorado, April 1983.
- [5] ———, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.
- [6] A. BRANDT, S. F. MCCORMICK, AND J. W. RUGE, *Algebraic multigrid for automatic multigrid solutions with application to geodetic computations*, Technical Report, Institute for Computational Studies, Fort Collins, Colorado, October 1982.
- [7] ———, *Algebraic multigrid for sparse matrix equations*, in Sparsity and Its Applications, D. J. Evans, ed., Cambridge University Press, 1984.
- [8] M. BREZINA, A. J. CLEARY, R. D. FALGOUT, V. E. HENSON, J. E. JONES, T. A. MANTEUFFEL, S. F. MCCORMICK, AND J. W. RUGE, *Algebraic multigrid based on element interpolation (AMGe)*, SIAM J. Sci. Comp., 22 (2000), pp. 1570–1592.
- [9] H.-J. BUNGARTZ, *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*, PhD thesis, Technische Universität München, 1992.
- [10] H.-J. BUNGARTZ, *Finite Elements of Higher Order on Sparse Grids*, Shaker Verlag, 1998.
- [11] H.-J. BUNGARTZ, T. DORNSEIFER, AND C. ZENGER, *Tensor product approximation spaces for the efficient numerical solution of partial differential equations*, in Proc. Int. Workshop on Scientific Computations, Nova Science Publishers, Konya, 2000.
- [12] H.-J. BUNGARTZ AND M. GRIEBEL, *A note on the complexity of solving Poisson’s equation for spaces of bounded mixed derivatives*, J. Complexity, 15 (1999), pp. 167–199.
- [13] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 1–123.
- [14] C. CHUI AND Y. WANG, *A general framework for compactly supported splines and wavelets*, J. Approx. Theory, 71 (1992), pp. 263–304.
- [15] P. CIARLET, *The Finite Element Method for Elliptic Problems*, Classics in Applied Mathematics, SIAM, 2002.
- [16] A. COHEN, *Numerical Analysis of Wavelet Methods*, Studies in Mathematics and its Applications, Vol. 32, North Holland, 2003.
- [17] A. COHEN, L. ECHEVERRY, AND Q. SUN, *Finite element wavelets*, tech. rep., Laboratoire d’Analyse Numérique, Université Pierre et Marie Curie, 2000.
- [18] S. DAHLKE, W. DAHMEN, AND R. DEVORE, *Nonlinear approximation and adaptive techniques for solving elliptic operator equations*, in Multiscale Wavelet Methods for PDEs, W. Dahmen, A. Kurdila, and P. Oswald, eds., Academic Press, 1997, pp. 237–284.
- [19] S. DAHLKE, W. DAHMEN, R. HOCHMUTH, AND R. SCHNEIDER, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Math., 23 (1997), pp. 21–47.
- [20] W. DAHMEN AND R. STEVENSON, *Element-by-element construction of wavelets satisfying stability and moment conditions*, SIAM J. Numer. Anal., 37 (1999), pp. 319–352.
- [21] W. DAMEN, *Wavelet and multiscale methods for operator equations*, Acta Numerica, (1997), pp. 55–228.
- [22] I. DAUBECHIES, *Orthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.
- [23] G. FABER, *Über stetige Funktionen*, Mathematische Annalen, (1909), pp. 81–94.
- [24] T. GERSTNER, *Adaptive hierarchical methods for landscape representation and analysis*, in Process Modelling and Landform Evolution, vol. 78 of Lecture Notes in Earth Sciences, Springer Verlag, 1999.
- [25] T. GERSTNER AND M. GRIEBEL, *Numerical integration using sparse grids*, Numerical Algorithms, 18 (1998), pp. 209–232.
- [26] G. GONNET AND R. BAEZA-YATES, *Handbook of Algorithms and Data Structures*, Addison Wesley, 1991.
- [27] T. GRAUSCHOPF, M. GRIEBEL, AND H. REGLER, *Additive multilevel-preconditioners based on bilinear interpolation, matrix dependent geometric coarsening and algebraic multigrid coarsening for second order elliptic PDEs*, Applied Numerical Mathematics, 23(1) (1997), pp. 63–96.
- [28] M. GRIEBEL, *Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences*, Computing, 61 (1998), pp. 151–179.

- [29] M. GRIEBEL AND S. KNAPEK, *Optimized tensor-product approximation spaces*, Constructive Approximation, 16 (2000), pp. 525–540.
- [30] M. GRIEBEL AND D. OELTZ, *A sparse grid space-time discretization scheme for parabolic problem*, to be submitted.
- [31] M. GRIEBEL AND P. OSWALD, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Advances in Computational Mathematics, 4 (1995), pp. 171–206.
- [32] M. GRIEBEL, P. OSWALD, AND T. SCHIEKOFER, *Sparse grids for boundary integral equations*, Numer. Mathematik, 83 (1999), pp. 279–312.
- [33] A. GRIEWANK, *Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation*, Optim. Methods Softw., 1 (1992), pp. 35–54.
- [34] M. GUNZBURGER, *Perspectives in Flow Control and Optimization*, SIAM, 2003.
- [35] M. HINZE AND K. KUNISCH, *Three control methods for time-dependent fluid flows*, Journal of Flow, Turbulence and Combustion, 65 (2000), pp. 1386–6184.
- [36] S. KNAPEK, *Approximation und Kompression mit Tensorprodukt-Multiskalen-Approximationsräumen*, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2000.
- [37] F. KOSTER, *Multiskalen-basierte Finite-Differenzen-Verfahren auf adaptiven dünnen Gittern*, PhD thesis, Rheinische Friedrich-Hilhelms-Universität Bonn, 2002.
- [38] K. KUNISCH AND S. VOLKWEIN, *Galerkin proper orthogonal decomposition methods for parabolic systems*, Numer. Math., (2001), pp. 117–148.
- [39] O. LADYZENSKAJA, V. SOLONNIKOV, AND N. URAL'CEVA, *Linear and Quasilinear Equations of Parabolic Type*, vol. 23 of Translations of Mathematical Monographs, American Mathematical Society, 1988.
- [40] K. MEHLHORN, *Datenstrukturen und Effiziente Algorithmen, Band 1: Sortieren und Suchen*, Teubner Verlag, 1988.
- [41] J. W. RUGE AND K. STÜBEN, *Efficient solution of finite difference and finite element equations by algebraic multigrid*, in Multigrid Methods for Integral and Differential Equations, D. J. Paddon and H. Holstein, eds., The Institute of Mathematics and its Applications Conference Series, Clarendon Press, 1985.
- [42] T. SCHIEKOFER, *Die Methode der Finiten Differenzen auf Dünne Gittern*, PhD thesis, Rheinische Friedrich-Hilhelms-Universität Bonn, 1998.
- [43] P. SCHRÖDER AND W. SWELDENS, *Spherical wavelets: Efficiently representing functions on the sphere*, Computer Graphics Proceedings (SIGGRAPH 95), (1995), pp. 161–172.
- [44] R. STEVENSON, *Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes*, Constr. Approx., 19 (2003), pp. 477–508.
- [45] W. SWELDENS, *The lifting scheme: a construction of second generation wavelets*, SIAM J. Math. Anal., 29 (1997), pp. 511–546.
- [46] V. N. TEMLYAKOV, *Approximation of functions with bounded mixed derivative*, Proc. Steklov Inst. Math., 1 (1989).
- [47] V. N. TEMLYAKOV, *On approximate recovery of functions with bounded mixed derivatives*, J. Complexity, 9 (1993), pp. 41–59.
- [48] U. TROTTEMBERG, C. W. OSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, San Diego, 2001, Appendix A: An Introduction to Algebraic Multigrid by K. Stüben, pp. 413–532.
- [49] P. VANĚK, J. MANDEL, AND M. BREZINA, *Algebraic Multigrid based on Smoothed Aggregation for Second and Fourth Order Problems*, Computing, 56 (1996), pp. 179–196.
- [50] P. S. VASSILEVSKI AND L. T. ZIKATANOV, *Multiple vector preserving interpolation mappings in algebraic multigrid*, LLNL Report UCRL-JRNL-208036, Lawrence Livermore National Laboratory, 2004.
- [51] H. YSERENTANT, *On the multi-level splitting of finite element spaces*, Numer. Math., 58 (1986), pp. 379–412.
- [52] ———, *Hierarchical bases in the numerical solution of parabolic problems*, in Large Scale Scientific Computing, P. Deufhard and et. al., eds., Birkhäuser, 1987, pp. 22–36.