A Sparse Grid Space-Time Discretization Scheme for Parabolic Problems

Michael Griebel, Daniel Oeltz

Abstract

In this paper we consider the discretization in space *and* time of parabolic differential equations where we use the so-called space-time sparse grid technique. It employs the tensor product of a one-dimensional multilevel basis in time and a proper multilevel basis in space. This way, the additional order of complexity of a direct space-time discretization can be avoided, provided that the solution fulfills a certain smoothness assumption in space-time, namely that its mixed space-time derivatives are bounded. This holds in many applications due to the smoothing properties of the propagator of the parabolic PDE (heat kernel). In the more general case, the space-time sparse grid approach can be employed together with adaptive refinement in space *and* time and then leads to similar approximation rates as the non-adaptive method for smooth functions. We analyze the properties of different space-time sparse grid discretizations for parabolic differential equations from both, the theoretical and practical point of view, discuss their implementational aspects and report on the results of numerical experiments.

AMS Subject Classification: 35K20, 65M60, 65M99, 65Y20

Key words: space-time discretization, parabolic differential equations, Crank-Nicolson, discontinuous Galerkin, sparse grids

1 Introduction

The modeling of various phenomena in physics, chemistry, biology and financial engineering leads to parabolic partial differential equations. For their numerical treatment, there are mainly three discretization techniques which differ in the order of the discretization of the space and time operators. In the method of lines (MOL), space is discretized first. This reduces the time-dependent PDE to a system of ordinary differential equations which can then be treated by ODE-solvers with adequate time integration schemes [59]. Here, the accuracy in spatial finite element space can be locally estimated by standard a posteriori error estimation techniques for stationary problems [7], but it is nevertheless difficult to handle local mesh refinement in space within this approach, see [1, 2]. In Rothe's method, time is discretized first [37]. This leads to a sequence of stationary problems which can then be solved by standard techniques available for elliptic problems. Here, adaptivity can be applied to the space part of the problem [62] using standard error estimators for elliptic problems, see [41] and the references cited therein. However, to intertwine this space adaptivity with a proper adaptive refinement in time by e.g. a local time stepping technique is not an easy task. Finally, space and time can both be discretized at once. One important example for this approach is the so–called discontinuous Galerkin method (DGM) which was originally developed for the discretization of hyperbolic problems.[1] In the recent years it has come into focus for the solution of parabolic problems [25, 26, 27, 28, 29].

Let us assume that we use a uniform grid with N grid points in each space dimension and a uniform grid with M time points. Then the MOL, Rothe's method and the DGM have a computational complexity of $O(N^d \cdot M)$ where d denotes the dimension of the spatial finite element space. If we employ, depending on the smoothness of the solution, a discretization scheme with an error of the order p in time and an error of the order q in space and if we furthermore assume the relation $\delta t^p \approx \delta x^q$, $p, q \geq 1$, for the grid sizes $\delta x \simeq 1/N$ in space and $\delta t \simeq 1/M$ in time, we end up by virtue of $\delta t^{-1} = M \approx N^{q/p}$ with $O(N^{d+q/p})$ degrees of freedom for the whole approximate solution in space-time. 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. Therefore, even for a relatively coarse discretization, it is extremely demanding to solve parabolic problems with just three space dimensions on modern workstations. Moreover, to store the whole evolution of a solution over time, easily terabytes of memory are needed, which may be completely impossible for sufficiently fine mesh sizes. This so-called *curse of dimension* becomes even more dramatic for higher dimensional problems which arise from the theory of queuing networks [42, 55], from reaction mechanisms in molecular biology [23, 56], for the viscoelasticity in polymer fluids [50, 48, 47], or in various models for the pricing of financial derivatives [49].

To some extent, the curse of dimension can be circumvented for special approximation spaces. An example are the so-called sparse grid spaces which were originally developed for d-dimensional elliptic problems of second order. They are based on tensor products of one–dimensional multiscale functions. The coefficients of a sufficiently smooth solution in the resulting multivariate series representation then exhibit a specific decay with the number of levels involved. For certain function classes, i.e. for functions with dominating r-th mixed derivatives, truncation of the associated series expansion results in sparse grid spaces which need only $O(N \log(N)^{d-1})$ degrees of freedom

[1] Here, the functions of the approximation space are typically discontinuous in the time direction.

instead of $O(N^d)$ degrees of freedom for the case of uniform full grids, see [12] and the references cited therein. The achieved accuracy, however, is only slightly reduced from $O(h^r)$ to $O(h^r(\log h^{-1})^{d-1})$ in the L^2-norm if piecewise polynomials of degree r are used in the basic one-dimensional multilevel basis. With respect to the energy norm even the same order of accuracy can be obtained for both cases. Furthermore there are so-called energy-norm based sparse grids which only need $O(N)$ degrees of freedom but result in $O(h^{r-1})$ accuracy with respect to the energy norm. This approach completely eliminates the dependence of the dimension d in the complexities at least for the N-asymptotics, the order constants however still depend exponentially on d. The sparse grid method has been successfully applied to problems from quantum mechanics [32, 63, 38], to stochastic differential equations [54, 53], to high-dimensional integration problems from physics and finance [5, 44, 8, 34] and to the solution of moderately higher-dimensional partial differential equations, mainly of elliptic type [10, 3, 4]. For a survey, see [12].

These properties make the sparse grid technique a good candidate for a direct space-time discretization of parabolic PDE problems. Already in [3] a Galerkin-type sparse grid discretization scheme was proposed for the solution of parabolic problems and in [51] a method based on the sparse grid combination technique was investigated. Due to the tensor product type construction of sparse grid spaces it is difficult to handle complicated domains. In [21, 22, 46] this problem was treated by splitting the spatial domain into subdomains which then could be transformed to the unit cube. These transformations, however, enter the resulting differential operator which may cause further problems.

In [35] we proposed a space-time sparse grid construction which makes use of the tensor product structure of the space-time cylinder. There, a general isotropic spatial multilevel basis (not necessarily constructed via a tensor product construction) and a one–dimensional multilevel basis in time were used. This approach can be employed whenever a hierarchy of nested grids in space is available and thus allows to deal with complicated geometries in space. Then, for a d–dimensional spatial finite element space, the resulting space-time sparse grid space has only $O(N^d)$ degrees of freedom instead of $O(N^{d+1})$ for a full grid space in space and time.[2] Thus, the additional complexity due to the instationarity is avoided. This makes the method especially interesting for various parabolic problems and instationary optimization problems with three spatial dimensions and general spatial geometries. In [35] only results concerning the *interpolation error* of these space-time sparse grid spaces were given. They showed that (up to a logarithmic factor) the same convergence rates were indeed obtained as for the space-time full grid spaces.

[2] Note that one can also get almost $O(N^d)$ complexity if one uses an exponentially convergent time stepping scheme of hp-dG type.

In this paper we now employ such a space-time sparse grid discretization scheme for the solution of parabolic PDEs. Here, we use the classical hierarchical basis for the spatial part and either a continuous linear, a piecewise constant or a piecewise linear hierarchical basis for the time part of the problem. We thus propose and analyze space-time sparse grid discretization schemes which are the sparse counterparts of the Crank-Nicolson scheme and the discontinuous Galerkin scheme with piecewise constant or linear functions in time. We give an estimate of the discretization error of the discontinuous Galerkin schemes in terms of the interpolation error of the space-time sparse grids and we derive a stability estimate which shows that the discrete problems possess a unique solution for the Crank-Nicolson scheme. From the analysis of the discontinuous Galerkin scheme as well as from the presented numerical experiments it turns out that these schemes on space-time sparse grids lead to (nearly) the same convergence rates as the corresponding schemes on space-time full grids. Furthermore, for problems where the solution does not fulfill the necessary smoothness assumption we present an adaptive version in space *and* time, including local time stepping, which allows for an efficient discretization also in the non-smooth case.

The remainder of this paper is organized as follows. In section 2 we discuss the basic properties of space-time sparse grid spaces including the dimension and the approximation rate. Section 3 introduces the special space-time sparse grids used in the rest of the paper. Here we give different one–dimensional temporal multilevel bases and a spatial hierarchical basis derived from a nested sequence of finite element spaces, which we employ for our space-time sparse grid spaces. Furthermore we present the associated interpolation properties based on the results from section 2. In section 4 we shortly recall some regularity results for linear parabolic problems and show that these estimates provide solutions with bounded mixed derivatives, as needed for the theory in sections 2 and 3. In section 5 we discuss the discretization of parabolic problems. Here, we analyze the discontinuous Galerkin method in our space-time sparse grid setting with piecewise constant or linear functions in time and show that the discretization error can be bounded from above by the interpolation error of the continuous solution. For the Crank-Nicolson discretization we present a stability estimate which shows that the discrete system is uniquely solvable. Furthermore, we give numerical results for the different discretization schemes. Section 6 deals with the adaptive discretization for problems where the solution does not possess the necessary smoothness. Here, analogously to classical adaptive sparse grid discretizations for elliptic problems, we apply a simple error indicator based on the value of the hierarchical coefficients (now in space and time). Numerical examples show that the proposed method works well and also leads for singular solutions to good results. Finally, we give some conclusions in section 7.

2 Space-Time Sparse Grids

In this section we briefly sketch the general concept of space-time sparse grid spaces. We recall results from [35] on their dimension and on estimates of the interpolation errors. It turns out that they provide nearly the same (up to a logarithmic factor) rates in the energy- and L^2-norm as classical full grid spaces, whereas their dimension is reduced to that of the underlying *spatial* problem.

We denote multi-indices in boldface, e.g. $\mathbf{i}$ is a vector with components $i_1, \dots, i_d$ and define $\partial^{\mathbf{i}} := \partial_{x_1}^{i_1} \dots \partial_{x_d}^{i_d}$. For a given spatial domain $\Omega \subseteq \mathbb{R}^d$ and a final time point $T \in (0, \infty)$, Ω_T denotes the associated space-time cylinder, i.e. $\Omega_T := \Omega \times (0, T)$. Note that we restrict ourselves to the case where the domain does not change over time. It is this tensor product structure of the space-time cylinder which we exploit for our space-time sparse grid construction in the following. First, we consider Sobolev spaces in space-time on Ω_T.

Definition 2.1 *Let $\Omega \subseteq \mathbb{R}^d$ and $m \in \mathbb{N}$. We define, for fixed $m, k \in \mathbb{N}$, the spaces*

$$\begin{aligned} H^{m,k}_{mix}(\Omega_T) &:= H^m(\Omega) \otimes H^k((0,T)), \\ H^{m,k}(\Omega_T) &:= \left(H^m(\Omega) \otimes L^2((0,T))\right) \cap \Big(L^2(\Omega) \otimes H^k((0,T))\Big) \end{aligned}$$

and the associated norms

$$\begin{aligned} \|u\|_{H^{m,k}_{mix}(\Omega_T)} &:= \sum_{j \le k, \|\mathbf{i}\|_1 \le m} \|\partial_t^j \partial^{\mathbf{i}} u\|_{L^2(\Omega_T)}, \\ \|u\|_{H^{m,k}(\Omega_T)} &:= \sum_{j + \|\mathbf{i}\|_1 \le \max\{m,k\}, j \le k, \|\mathbf{i}\|_1 \le m} \|\partial_t^j \partial^{\mathbf{i}} u\|_{L^2(\Omega_T)}. \end{aligned}$$

We will focus on the special cases $m = 2k$ and $m = k$.

Note that usually the smoothness m in space and the smoothness k in time are used separately, i.e. it is assumed that the solution of a parabolic problem is contained in the spaces $H^{m,k}(\Omega_T)$. For example in [45], the smoothness in space is used for a spatial sparse grid discretization, whereas the smoothness in time is used for an adaptive *hp*-DGM. This approach is based on the smoothing effect of the parabolic operator (time analycity of the solution) and thus even problems with incompatible initial and boundary conditions where the solution is not very regular in space for small times t can be solved with nearly optimal convergence rates, provided that space singularities do not appear. Thus, the smoothness in space *and* time, i.e. the existence of bounded higher mixed space–time derivatives in the sense of the above $H^{m,k}$-spaces, which ensure this regularity only for highly compatible initial and boundary data (via the assumption of the existence of a sufficiently regular g in the later Theorem 4.2), is not needed for their approach while we rely on it.

In the following we give a short introduction into the space-time sparse grid concept. Let us assume that there are multiscale decompositions

$$L^2(\Omega) = \bigcup_{j\geq 0} V_j^\Omega \quad \text{and} \quad L^2((0,T)) = \bigcup_{j\geq 0} V_j^T,$$

with increment spaces W_j^Ω and W_j^T, i.e.

$$V_j^\Omega = V_{j-1}^\Omega \oplus W_j^\Omega, \quad V_j^T = V_{j-1}^T \oplus W_j^T,$$

$W_0^T := V_0^T$ and $W_0^\Omega := V_0^\Omega$. Then

$$L^2(\Omega) = \bigoplus_{j\geq 0} W_j^\Omega \quad \text{and} \quad L^2((0,T)) = \bigoplus_{j\geq 0} W_j^T. \tag{1}$$

For $L^2(\Omega_T) = L^2(\Omega) \otimes L^2((0,T))$ we obtain the decompositions

$$L^2(\Omega_T) = \bigcup_{\mathbf{j}\in\mathbb{N}^2} V_{\mathbf{j}} \quad \text{and} \quad L^2(\Omega_T) = \bigoplus_{\mathbf{j}\in\mathbb{N}^2} W_{\mathbf{j}}$$

with the subspaces $V_{\mathbf{j}} = V_{j_1}^\Omega \otimes V_{j_2}^T$ and the increment spaces $W_{\mathbf{j}} := W_{j_1}^\Omega \otimes W_{j_2}^T$, respectively. Here, for a basis $\{\psi_{j_1,i_1}^\Omega\}$ of $W_{j_1}^\Omega$ and a basis $\{\psi_{j_2,i_2}^T\}$ of $W_{j_2}^T$ we immediately obtain a basis $\{\psi_{\mathbf{j},\mathbf{i}} := \psi_{j_1,i_1}^\Omega \cdot \psi_{j_2,i_2}^T\}$ of $W_{\mathbf{j}}$. With this notation, we define the full grid spaces

$$V_l^\infty := \bigoplus_{\substack{j_1\leq l,\\ j_2\leq 2l}} W_{(j_1,j_2)} \quad \text{and} \quad \tilde{V}_l^\infty := \bigoplus_{\|\mathbf{j}\|_\infty\leq l} W_{(j_1,j_2)} \tag{2}$$

and the sparse grid spaces

$$V_l^0 := \bigoplus_{2j_1+j_2\leq 2l} W_{(j_1,j_2)} \quad \text{and} \quad \tilde{V}_l^0 := \bigoplus_{\|\mathbf{j}\|_1\leq l} W_{\mathbf{j}}, \tag{3}$$

respectively. Note that these spaces are in general anisotropic in space-time but isotropic in space. This is an important difference compared to classical sparse grid approximation spaces, c.f. [12], where the basis functions are in general anisotropic with respect to every coordinate direction due to the tensor product construction. In [35], the following result on the dimension of these spaces is given.

Lemma 2.2 *Let us assume that* $\dim(W_j^\Omega) = O(2^{d\cdot j})$ *and* $\dim(W_j^T) = O(2^j)$*. Then, we obtain*

$$\begin{aligned}
\dim(V_l^\infty) &= O(2^{(d+2)\cdot l}),\\
\dim(V_l^0) &= \begin{cases} O(2^{2dl}) \textit{ for } d=1,\\ O(2^{dl}l) \textit{ for } d=2,\\ O(2^{d\cdot l}) \textit{ for } d>2,\end{cases}\\
\dim(\tilde V_l^\infty) &= O(2^{(d+1)\cdot l}),\\
\dim(\tilde V_l^0) &= \begin{cases} O(2^l l) \textit{ for } d=1,\\ O(2^{d\cdot l}) \textit{ for } d>1.\end{cases}
\end{aligned}$$

This lemma shows that the space-time sparse grid spaces V_l^0 and $\tilde V_l^0$ have significantly less degrees of freedom than the full grid spaces V_l^∞ and $\tilde V_l^\infty$. Moreover, for $d>2$ and $d>1$ their dimension is even of the same order as the dimension of the underlying spatial finite element space V_l^Ω.

Furthermore, in [35] it was shown that the space-time sparse grid spaces lead to the *same* interpolation rates as the full grid spaces provided that the multilevel splittings (1) are L^2-stable and fulfill certain norm equivalencies. For example in the case $u \in H^{2,2}_{mix}$ we may employ the norm equivalencies[3]

$$\|v^\Omega\|^2_{H^2} \simeq \sum_j 2^{4j}\|w_j^\Omega\|^2_{L^2} \quad \text{for all} \quad v^\Omega = \sum_j w_j^\Omega,\ w_j^\Omega \in W_j^\Omega, \tag{4}$$

$$\|v^T\|^2_{H^2} \simeq \sum_j 2^{4j}\|w_j^T\|^2_{L^2} \quad \text{for all} \quad v^T = \sum_j w_j^T,\ w_j^T \in W_j^T. \tag{5}$$

Here, to simplify notation we omit the domain of integration in the norms, i.e. we write $\|\cdot\|_{L^2}$ instead of $\|\cdot\|_{L^2(\Omega)}$, etc.

For a simple domain Ω which is (e.g. after a suitable differentiable mapping) a d-cube, there exist a variety of bases which provide norm equivalencies of the necessary type like orthogonal wavelets, spline-wavelets, prewavelets, biorthogonal wavelets, lifting wavelets and similar constructions which are derived from a mother function by translation and dilation, see [13, 15, 18, 20, 58] and the references cited therein. 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.[4] For a more complicated, general spatial domain Ω, the construction of a multilevel basis which fulfills the necessary norm equivalencies can be quite difficult and demanding. For wavelet-like approaches on polygonal domains, see [16, 19, 57].

[3] Here, $A \simeq B \Leftrightarrow c_1 B \le A \le c_2 B$ with positive constants c_1, c_2.

[4] In these cases sparse grids can also be applied for the spatial discretization. This results in substantial further savings in cost provided that an additional smoothness prerequisite in space like the boundedness of the second mixed spatial derivative is fulfilled. For details see the survey article [12] and the references cited therein.

Note that the postulated norm equivalencies (4) and (5) do not hold for the piecewise linear functions which will later be used in our numerical experiments. But in view of the approximation properties it is not really necessary that such norm equivalencies are fulfilled. Here, as the next Lemma shows, we obtain error bounds for arbitrary norms if only some *upper* bounds for the increments $w_{\mathbf{j}} \in W_{\mathbf{j}}$ of a function $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$ are valid.

Lemma 2.3 *Let $\|\cdot\|_\star$ and $\|\cdot\|_\Delta$ be two arbitrary norms. Let us assume that there is for fixed $t > s \geq 0$ a constant $c > 0$ such that, for every u with $\|u\|_\star < \infty$ and $\|u\|_\Delta < \infty$, $u = \sum_{\mathbf{j}} w_{\mathbf{j}}$, $w_{\mathbf{j}} \in W_{\mathbf{j}}$, we have*

$$\|w_{\mathbf{j}}\|_\star \leq c 2^{s\|\mathbf{j}\|_\infty - t\|\mathbf{j}\|_1} \|u\|_\Delta \ \forall \mathbf{j} \in \mathbb{N}^2. \tag{6}$$

Then we obtain

$$\inf_{v \in \tilde{V}_l^0} \|u - v\|_\star^2 \leq \begin{cases} c \cdot 2^{2(s-t)l} \cdot l^2 \|u\|_\Delta^2 & \text{for } s = 0, t > 0, \\ c \cdot 2^{2(s-t)l} \|u\|_\Delta^2 & \text{for } s > 0, t > 0. \end{cases} \tag{7}$$

If the upper estimate

$$\|w_{\mathbf{j}}\|_\star \leq c 2^{s\|(2j_1, j_2)\|_\infty - t\|(2j_1, j_2)\|_1} \|u\|_\Delta \ \forall \mathbf{j} \in \mathbb{N}^2 \tag{8}$$

is valid, we obtain

$$\inf_{v \in V_l^0} \|u - v\|_\star^2 \leq \begin{cases} c \cdot 2^{4(s-t)l} \cdot l^2 \|u\|_\Delta^2 & \text{for } s = 0, t > 0, \\ c \cdot 2^{4(s-t)l} \|u\|_\Delta^2 & \text{for } s > 0, t > 0. \end{cases} \tag{9}$$

This Lemma is a simple generalization of Theorem 4 in [39]. We will use this result in the following to prove the interpolation properties of space-time sparse grids which are constructed from hierarchical bases which do not fulfill stable norm equivalencies like (4) and (5).

3 Multilevel Bases

For the construction of the space-time sparse grid, we need, besides a multilevel splitting of the function space on the spatial domain, also a multilevel splitting of the function space on the time interval. To this end, we introduce one–dimensional splittings constructed from a hierarchical basis of piecewise constant or piecewise linear, discontinuous functions or continuous linear functions in subsection 3.1. Then, in subsection 3.2, we investigate error bounds for the resulting space-time sparse grid spaces constructed from these one–dimensional splittings in time and the classical linear (d-linear) hierarchical basis in space.

3.1 Multilevel Splittings in Time

For the ease of presentation, let us assume that $T = 1.0$ and let us restrict ourselves to zero boundary conditions for $t = 0$ and $t = 1$. For fixed level j we divide the time interval $(0, 1]$ in 2^j subintervals

$$I_k^j := ((k-1)2^{-j}, k2^{-j}], \tag{10}$$

$1 \le k \le 2^j$. In the following we make use of three different one-dimensional multilevel splittings.

1. We denote by $V_{j,1}^T$ the space of functions which are linear on each interval I_k^j and globally continuous. As a basis of this space, we use the simple hat functions $\varphi_{j,k}^T$, $1 \le k < 2^j$,

$$\varphi_{j,k}^T(x) := \begin{cases} 1 - |(k - 2^j x)| \text{ for all } x \in I_k^j \cup I_{k+1}^j, \\ 0 \text{ elsewhere}, \end{cases} \tag{11}$$

   i.e. $V_{j,1}^T = \text{span}\{\varphi_{j,k}^T \mid 1 \le k < 2^j\}$. We then obtain with $V_{j,1}^T := V_{j-1,1}^T \oplus W_{j,1}^T$ the hierarchic increment spaces $W_{j,1}^T$ as the span of the classical hierarchical basis, $\psi_{j,k}^T := \varphi_{j,2k-1}^T$, $1 \le k < 2^{j-1}$, i.e.

$$W_{j,1}^T = \text{span}\{\psi_{j,k}^T \mid 1 \le k < 2^{j-1}\}. \tag{12}$$

2. We define the space $V_{j,0}^T$ as the set of all functions which are constant on each interval I_k^j. As a basis of this space, we use

$$\varphi_{j,k}^T(x) := \begin{cases} 1 & \text{for all } x \in I_k^j, \\ 0 & \text{else}, \end{cases} \tag{13}$$

   i.e. $V_{j,0}^T = \text{span}\{\varphi_{j,k}^T \mid 1 \le k < 2^j\}$. Here, we use the hierarchic increments $W_{j,0}^T$, $V_{j,0}^T := V_{j-1,0}^T \oplus W_{j,0}^T$, which are the span of the basis functions $\psi_{j,k}^T := \varphi_{j,2k-1}^T$, $1 \le k < 2^{j-1}$, i.e.

$$W_{j,0}^T = \text{span}\{\psi_{j,k}^T \mid 1 \le k < 2^{j-1}\}. \tag{14}$$

3. Finally, we use the spaces $V_{j,1}^{T,disc}$ of functions which are linear on each interval I_j^k but not necessarily continuous. The functions $\varphi_{j,k}^T$ and $\tilde{\varphi}_{j,k}^T$,

$$\begin{aligned} \varphi_{j,k}^T(x) &= \begin{cases} 1 - \frac{|k \cdot 2^{-j} - x|}{2^{-j}} & \text{if } x \in I_k^j, \\ 0 & \text{otherwise}, \end{cases} \\ \tilde{\varphi}_{j,k}^T(x) &= \begin{cases} 1 - \frac{|(k-1) \cdot 2^{-j} - x|}{2^{-j}} & \text{if } x \in I_k^j, \\ 0 & \text{elsewhere}, \end{cases} \end{aligned}$$

form a basis of $V_{j,1}^{T,disc}$, i.e.

$$V_{j,1}^{T,disc} = \mathrm{span}\{\varphi_{j,k}^T, \tilde{\varphi}_{j,\tilde{k}}^T \mid 1 \leq k < 2^j,\ 1 < \tilde{k} \leq 2^j\}$$

and we use the hierarchic increments $W_{j,1}^{T,disc}$, $V_{j,1}^{T,disc} := V_{j-1,1}^{T,disc} \oplus W_{j,1}^{T,disc}$ where

$$W_{j,1}^{T,disc} = \mathrm{span}\{\psi_{j,k}^T, \tilde{\psi}_{j,\tilde{k}}^T \mid 1 \leq k < 2^{j-1},\ 1 < \tilde{k} \leq 2^{j-1}\} \tag{15}$$

with $\psi_{j,k}^T := \varphi_{j,2\cdot k-1}^T$ and $\tilde{\psi}_{j,\tilde{k}}^T := \tilde{\varphi}_{j,2\cdot\tilde{k}}^T$.

Figure 1 shows examples for functions of the spaces $V_{3,1}^T$, $V_{3,0}^T$ and $V_{3,1}^{T,disc}$ as well as the basis functions of the respective increment spaces.

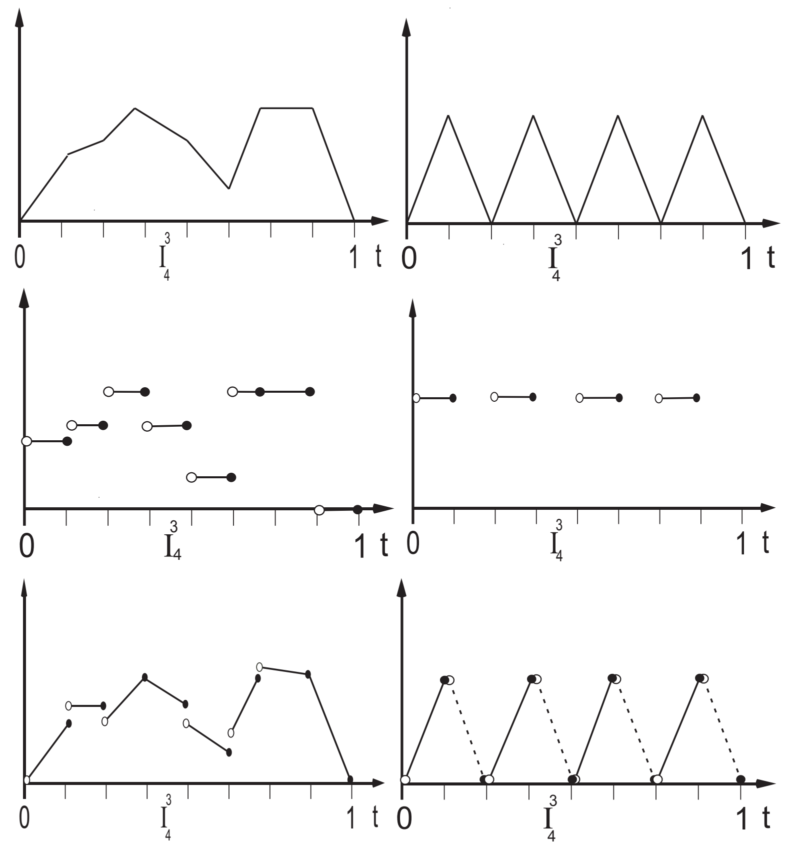


Figure 1: Functions of the one-dimensional discrete function spaces (left column) and hierarchical basis elements of the increment spaces (right column): $V_{3,1}^T$ and basis of $W_{3,1}^T$ (top), $V_{3,0}^T$ and basis of $W_{3,0}^T$ (middle), $V_{3,0}^{T,disc}$ and basis of $W_{3,1}^{T,disc}$ (bottom).

3.2 The Hierarchical Basis in Space

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

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

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 $P_j^\Omega : C^0(\Omega) \to V_j^\Omega$. Then each $u_l \in V_l^\Omega$ has the representation

$$u_l = P_0^\Omega u_l + \sum_{j=0}^{l-1}(P_{j+1}^\Omega u_l - P_j^\Omega u_l), \tag{16}$$

with the increment spaces

$$W_{j+1}^\Omega := \mathrm{range}(P_{j+1}^\Omega - P_j^\Omega) \tag{17}$$

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

$$P_j^\Omega u(x) = u(x) \text{ for all } x \in \mathcal{N}_j, \tag{18}$$

we obtain a splitting into increment spaces which are just spanned by the hierarchical basis as described in [61]. For the choice $P_j^\Omega u$ 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 or approximate L^2-projections which lead to a stable multilevel basis.

In the following we focus on the choice of P_j as the interpolation operator (18) onto V_j. If we denote by $\{\varphi_{j,k}^\Omega\}$ the nodal basis of the finite element space V_j we directly obtain the basis $\{\psi_{j,k}^\Omega\} := \{\varphi_{j,k}^\Omega \mid \varphi_{j,k}^\Omega(n_i) = 0 \text{ for all } n_i \in \mathcal{N}_{j-1}\}$ of the increment space W_j^Ω, i.e. $W_j^\Omega = \mathrm{span}\{\psi_{j,k}^\Omega\}$.

3.3 Interpolation Error Estimates

Along the lines of [35, 43], we now discuss the interpolation properties of the space-time sparse grid spaces constructed from the d-linear hierarchical basis in spatial finite element space as described in section 3.2 and the three one-dimensional multilevel splittings in time as introduced in section 3.1.

Since we are dealing with a regular triangulation, the classical interpolation theory of finite elements spaces, c.f. [9, 14], shows that

$$\|u - P_j^\Omega u\|_{L^2(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^2(\Omega)} \tag{19}$$

for all $u \in H^2(\Omega)$ with constant $c > 0$ independent of j. Therefore, using the hierarchical decomposition (16) we obtain for the parts w_j of the splitting $u = P_0^\Omega u + \sum_{j\in\mathbb{N}}(P_{j+1}^\Omega u_l - P_j^\Omega u_l) = \sum_{j\in\mathbb{N}} w_j$, $u \in H^2(\Omega)$, the estimate

$$\|w_j\|_{L^2(\Omega)} \leq c \cdot 2^{-2j} \|u\|_{H^2(\Omega)} \tag{20}$$

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 P_j^T and thus for the associated multilevel splitting of a temporal function with increment space $W_{j,1}^T$ from (12).

Now, let us consider a function $u \in H_{mix}^{2,2}(\Omega_T)$ which depends on space and time. Then, for the splitting $u = \sum_{\mathbf{j}\in\mathbb{N}^2} w_{\mathbf{j}}$, $w_{\mathbf{j}} \in W_{\mathbf{j}} = W_{j_1}^\Omega \otimes W_{j_2,1}^T$, arguments on the tensor product of operators [36] lead to the estimate

$$\|w_{\mathbf{j}}\|_{L^2(\Omega_T)} \leq c \cdot 2^{-2\cdot|\mathbf{j}|_1} \|u\|_{H_{mix}^{2,2}(\Omega_T)} \ \forall \mathbf{j} \in \mathbb{N}^2, \tag{21}$$

with c independent of $\mathbf{j}$.

The application of Lemma 2.3 with $\|\cdot\|_\star = \|\cdot\|_{L^2(\Omega_T)}$ and $\|\cdot\|_\Delta = \|\cdot\|_{H_{mix}^{2,2}(\Omega_T)}$ leads to the existence of a constant $c > 0$ independent of l such that

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

holds true for all $u \in H_{mix}^{2,2}(\Omega)$. Analogously, we obtain estimates for the approximation errors in other norms than just $\|\cdot\|_{L^2(\Omega_T)}$ and other smoothness properties. Here and in the following we denote the space-time sparse grid interpolant in the space V_l^0 or $\tilde{V}_l^0$ of a continuous function u by $P_{V_l^0} u$ and $P_{\tilde{V}_l^0} u$ and omit any index concerning the specific choice of increment space used in time to simplify notation. If we assume that the function to be approximated is continuous, the results directly carry over to the space-time sparse grid constructed from the piecewise linear hierarchical increment spaces $W_{j,1}^{disc}$ from (15). Using interpolation error estimates for the piecewise constant interpolation in one dimension we can show similar estimates for the space-time sparse grid spaces where we use the increments $W_{j,0}^T$ from (14) in time. Overall, we obtain the following estimates:

Lemma 3.1 *Let the sequence of finite element spaces V_j^Ω fulfill (19) with a constant c independent of j for all $u \in H^2(\Omega)$. Then, for the space-time sparse grid $\tilde{V}_l^0$ constructed from the hierarchical basis increment spaces (17) in space and the continuous linear hierarchical increment spaces $W_{j,1}^T$ from (12), we have the estimates*

$$\begin{aligned} \|u - P_{\tilde{V}_l^0} u\|_{L^2(\Omega_T)} &\le c \cdot 2^{-2l} \cdot l \cdot \|u\|_{H^{2,2}_{mix}(\Omega_T)}, & (22)\\ \|u - P_{\tilde{V}_l^0} u\|_{H^1(\Omega)\otimes L^2((0,T))} &\le c \cdot 2^{-l} \|u\|_{H^{2,2}_{mix}(\Omega_T)}, & (23)\\ \|u - P_{\tilde{V}_l^0} u\|_{L^2(\Omega)\otimes L^\infty((0,T))} &\le c \cdot 2^{-2l} \cdot l \cdot \|u\|_{H^2(\Omega)\otimes H^{2,\infty}(0,T)}. & (24) \end{aligned}$$

Here, we denote by $H^{2,\infty}(0,T)$ the Sobolev space with weak derivatives up to the order second which are bounded in L^∞. The same estimates hold true for the space $\tilde{V}_l^0$ if we use the piecewise linear discontinuous increment spaces $W_{l,1}^{T,disc}$ from (15). The use of the hierarchic increments $W_{l,0}^T$ from (14) with piecewise constant functions in time in the construction of $\tilde{V}_l^0$ leads to the estimates

$$\begin{aligned} \|u - P_{\tilde{V}_l^0} u\|_{L^2(\Omega_T)} &\le c \cdot 2^{-l} \cdot \|u\|_{H^{2,1}_{mix}(\Omega_T)}, & (25)\\ \|u - P_{\tilde{V}_l^0} u\|_{H^1(\Omega)\otimes L^2((0,T))} &\le c \cdot 2^{-l} \cdot l \cdot \|u\|_{H^{2,1}_{mix}(\Omega_T)}, & (26)\\ \|u - P_{\tilde{V}_l^0} u\|_{L^2(\Omega)\otimes L^\infty((0,T))} &\le c \cdot 2^{-l} \cdot l \cdot \|u\|_{H^2(\Omega)\otimes H^{1,\infty}(0,T)}. & (27) \end{aligned}$$

If we use the space V_l^0 instead of the space $\tilde{V}_l^0$ we get the estimates

$$\begin{aligned} \|u - P_{V_l^0} u\|_{L^2(\Omega_T)} &\le c \cdot 2^{-2l} \cdot l \cdot \|u\|_{H^{2,1}_{mix}(\Omega_T)}, & (28)\\ \|u - P_{V_l^0} u\|_{H^1(\Omega)\otimes L^2((0,T))} &\le c \cdot 2^{-l} \cdot \|u\|_{H^2(\Omega)\otimes H^{1,\infty}(0,T)}, & (29)\\ \|u - P_{V_l^0} u\|_{L^2(\Omega)\otimes L^\infty((0,T))} &\le c \cdot 2^{-2l} \cdot l \cdot \|u\|_{H^2(\Omega)\otimes H^{1,\infty}(0,T)}. & (30) \end{aligned}$$

For a detailed discussion of the proofs we refer to [43].

The above results show that if we interpolate functions with bounded mixed space-time derivatives using our space-time sparse grid spaces we obtain nearly the same approximation rates as for conventional full grid spaces in space-time. Note however that space-time sparse grids have substantially less degrees of freedom than space-time full grid spaces and thus reduce the overall memory and computational complexity substantially.

4 Classical Regularity Theory for Parabolic Problems

Although the sparse grid concept seems to be quite natural for space-time problems due to the tensor product structure between space and time, it is not directly obvious that parabolic problems lead to solutions which fulfill the regularity assumptions needed in Lemma 3.1. This will be the topic of this section where we briefly recall some classical regularity results for parabolic problems following the lines of [60]. These results show that under suitable smoothness assumptions on the domain, the coefficients of the elliptic operator, the boundary conditions and the right hand side, the resulting solutions are contained in $H^{m,k}_{mix}(\Omega_T)$ and thus space-time sparse grids can indeed be applied.

In the following let $\Omega \subset \mathbb{R}^d$ be a bounded domain and let $\Omega_T := \Omega \times (0,T]$ be the associated space-time domain for fixed $T > 0$. Let L be a linear elliptic operator of second order, i.e. for $u : \Omega_T \to \mathbb{R}$

$$Lu = -\nabla \cdot (A \nabla u) + b \cdot \nabla u + c \cdot u \tag{31}$$

with $A = (a_{ij})_{i,j=1}^d : \Omega_T \mapsto \mathbb{R}^{d\times d}$, $b = (b_i)_{i=1}^d : \Omega_T \mapsto \mathbb{R}^d$ and $c : \Omega_T \mapsto \mathbb{R}$. The derivatives are here only taken w.r.t. the spatial variables. To simplify the presentation, we will assume that L does not depend on the time, i.e. $A = (a_{ij})_{i,j=1}^d : \Omega \mapsto \mathbb{R}^{d\times d}$, $b = (b_i)_{i=1}^d : \Omega \mapsto \mathbb{R}^d$ and $c : \Omega \mapsto \mathbb{R}$.

Now, for given $f : \Omega_T \mapsto \mathbb{R}$, $u_0 : \Omega \mapsto \mathbb{R}$ and $g : \partial\Omega \times (0,T]$ we search for a solution $u : \Omega_T \mapsto \mathbb{R}$ such that

$$\begin{aligned} \partial_t u + Lu &= f \text{ in } \Omega_T, && (32)\\ u &= g \text{ on } \partial\Omega \times (0,T], && (33)\\ u(\cdot,0) &= u_0 \text{ on } \Omega, && (34) \end{aligned}$$

is fulfilled. If we multiply equation (32) with a test function $\varphi \in H_0^1(\Omega)$, integration by parts leads to the weak formulation of the problem

$$\begin{aligned} (\tfrac{d}{dt}u, \varphi)_{L^2(\Omega)} + a(u,\varphi) &= (f,\varphi)_0 \text{ for all } \varphi \in H_0^1(\Omega),\ \forall t \in (0,T]\\ u(\cdot,0) &= u_0(\cdot), \text{ a.e. in } \Omega,\\ u(t,x) &= g(t,x) \text{ a.e. } t \in (0,T) \text{ a.e. in } \partial\Omega, \end{aligned} \tag{35}$$

where

$$a(u,v) := \int_\Omega (\nabla u)^T A \nabla v + bv \cdot \nabla u + cuv\, dx. \tag{36}$$

We need the following definition.

Definition 4.1 *For $k,m \in \mathbb{N}$ we define the spaces $W^k((0,T);H^m(\Omega))$ as the union of all functions $u : (0,T) \mapsto H^m(\Omega)$ which satisfy*

$$\frac{d^n u}{dt^n} \in L^2((0,T);H^m(\Omega)) \text{ for all } 0 \le n \le k, \tag{37}$$

where the derivatives have to be understood in the distributional sense. We define the norm on $W^k((0,T);H^m(\Omega))$ by

$$\|u\|^2_{W^k((0,T);H^m(\Omega))} := \sum_{n=0}^k \int_0^T \|\frac{d^n u}{dt^n}\|^2_{H^m(\Omega)}\, dt. \tag{38}$$

Let us now assume that the following conditions on the bilinear form $a(\cdot,\cdot)$ from (36) hold true.

1. The bilinear form is continuous, i.e. there is a constant $C > 0$ such that

$$|a(\psi,\varphi)| \leq C \cdot \|\psi\|_{H_0^1(\Omega)} \|\varphi\|_{H_0^1(\Omega)}, \text{ for all } \varphi,\psi \in H_0^1(\Omega). \tag{39}$$

2. The bilinear form is coercive, i.e. there is a $c > 0$ such that

$$a(\varphi,\varphi) \geq c \cdot \|\varphi\|^2_{H_0^1(\Omega)}, \text{ for all } \varphi \in H_0^1(\Omega). \tag{40}$$

3. The coefficients are smooth i.e. $a_{ij}, b_i, c \in C^k(\bar{\Omega})$.

Then, we have the following result on the regularity of the solution of the weak problem (35).

Theorem 4.2 *Let the conditions 1, 2 and 3 hold true for the bilinear form $a(\cdot,\cdot)$ and let $k > 0$ be fixed. Furthermore, let us assume that there is an extension of the function g on the boundary to the interior of the domain, again denoted by g, such that*

$$g \in W^k((0,T);H^1(\Omega)) \quad \text{and} \quad \partial_t^{k+1} g \in L^2((0,T);H^{-1}(\Omega))$$

and

$$\partial_t^l u(\cdot,0) - \partial_t^l g(\cdot,0) \in H_0^1(\Omega),\ \forall 0 \leq l \leq k-1 \text{ and } \partial_t^k u(\cdot,0) - \partial_t^k g(\cdot,0) \in L^2(\Omega). \tag{41}$$

Furthermore let $f \in W^k((0,T);H^{-1}(\Omega))$. Then there is a unique solution u of the problem (35) and we have

$$u \in W^k((0,T);H^1(\Omega)) \quad \text{and} \quad \partial_t^{k+1} u \in L^2((0,T);H^{-1}(\Omega)). \tag{42}$$

A similar result concerning higher smoothness with respect to the spatial direction imposes stronger conditions to the smoothness of the coefficients of the elliptic part of the parabolic operator and the right hand side. This is expressed in the following theorem.

Theorem 4.3 *For $q > 0$ fixed, we assume that the condition 3 is satisfied with $k+q$ instead of k. Let the domain Ω be bounded and of smoothness class $C^{1+q+k,1}$. Furthermore, let the conditions of the previous theorem be satisfied and let*

$$f \in W^{k-1}((0,T);H^q(\Omega)) \quad \text{and} \quad \partial_t^k f \in L^2((0,T);H^{-1}(\Omega)) \tag{43}$$

and

$$g \in W^k((0,T); H^{2+\tilde{q}}(\Omega)) \quad \text{and} \quad \partial_t^{k+1} g \in L^2((0,T); H^{-1}(\Omega)), \tag{44}$$

where $\tilde{q} = \max\{\min(q, m_j) \mid 1 \leq j \leq k\}$ *and* m_j *is recursively defined by* $m_1 = 2 + \min(1, q)$, $m_{j+1} = 2 + \min(q, m_j)$. *Then, the problem (35) admits a unique solution* u *with*

$$u \in W^{k-j}((0,T); H^{2+\min(q,m_j)}) \ \forall\, 1 \leq j \leq k. \tag{45}$$

For the proofs of the Theorems we refer to [60].

The above results show the following: For every fixed $k, m \in \mathbb{N}$, we can achieve the solution u of the parabolic problem to be contained in the space $W^k((0,T); H^m(\Omega))$ if we enforce certain smoothness and compatibility conditions. The next statement shows that the spaces $W^k((0,T); H^m(\Omega))$ and $H^{k,m}_{mix}(\Omega_T)$ are isomorphic identical.

Theorem 4.4 *For every fixed pair* $k, m \in \mathbb{N}$, *we have*

$$W^k((0,T); H^m(\Omega)) = H^{m,k}_{mix}(\Omega_T). \tag{46}$$

Proof. Let $u \in W^k((0,T); H^m(\Omega))$. We choose an orthonormal basis $\{e_n(\cdot)\}$ of the space $H^m(\Omega)$. Now we define

$$u_n(t) := (u(t), e_n)_{H^m(\Omega)}, \tag{47}$$

and for fixed t we get $u(t)(\cdot) = \sum_n u_n(t) e_n(\cdot)$. Moreover, we have

$$\int_0^T |u_n(t)|^2 \, dt \leq \int_0^T \|u(t)\|^2_{H^m(\Omega)} \, dt < \infty$$

and therefore $u_n \in L^2((0,T))$. Since for $1 \leq l \leq k$ we have

$$\partial_t^l u_n(t) = (\frac{d^l}{dt^l} u(t), e_n)_{H^m(\Omega)} \in L^2(0,T), \tag{48}$$

we get $u_n \in H^k(0,T)$ and $w_m \in H^{m,k}_{mix}(\Omega_T)$. Furthermore, for $0 \leq l \leq k$, it holds

$$\lim_{m \to \infty} \int_0^T \|\partial_t^l w_m - \partial_t^l u\|^2_{H^m(\Omega)} \, dt =$$
$$\lim_{m \to \infty} \int_0^T \| \sum_{n>m} (\partial_t^l u, e_n)_{H^m(\Omega)} e_n \|^2_{H^m(\Omega)} \, dt = 0$$

for the sequence $w_m(x,t) := \sum_{n=1}^m u_n(t) e_n$ and w_m is a Cauchy sequence in $W^k((0,T); H^m(\Omega))$ converging to u. Therefore, since

$$\|w_m - w_l\|_{W^k((0,T);H^m(\Omega))} = \|w_m - w_l\|_{H^{m,k}_{mix}(\Omega_T)},$$

w_m is a Cauchy sequence in $H^{m,k}_{mix}(\Omega_T)$ and since $H^{m,k}_{mix}(\Omega)$ is complete we conclude that $u = \lim_{t\to\infty} w_m \in H^{m,k}_{mix}(\Omega)$.

On the other hand, let $u \in H^{m,k}_{mix}(\Omega)$. Then, there are sequences $u_n^\Omega \in H^m(\Omega)$, $u_n^T \in H^k((0,T))$ and $\lambda_n \in \mathbb{R}$, such that

$$u_n(x,t) := \sum_{i=1}^{n} \lambda_i u_i^\Omega(x) \otimes u_i^T(t)$$

converges to u with respect to $\|\cdot\|_{H^{m,k}_{mix}(\Omega)}$. Therefore, u_n is a Cauchy sequence in $H^{m,k}_{mix}(\Omega)$ and especially a Cauchy sequence in $W^k((0,T);H^m(\Omega))$.

Since $W^k((0,T);H^m(\Omega))$ is complete we have

$$u = \lim_{n\to\infty} u_n \in W^k((0,T);H^m(\Omega)).$$

□

This result together with Theorems 4.2 and 4.3 directly shows that we can expect the solution of parabolic problems to be in the spaces $H^{2,2}_{mix}(\Omega_T)$ and $H^{2,1}_{mix}(\Omega_T)$ assuming that the coefficients of the parabolic operator, the right–hand side, the boundary conditions and the domain are sufficiently smooth. We can then apply Lemma 3.1 and thus obtain interpolation errors of the space-time sparse grids for the solution of such a parabolic problem which are (up to a logarithmic factor) the same as for space-time full grids.

5 Sparse Grid Discretization

The previous results showed that space-time sparse grids significantly reduce the overall degrees of freedom in comparison to full grid discretizations, while they (nearly) lead to the same interpolation rates provided that only slightly stronger regularity assumptions are valid. From classical regularity theory it followed that under some reasonable conditions on the domain, the right hand side and the boundary- and initial conditions, the solutions of parabolic problems indeed fulfill these required stronger smoothness assumptions. Thus, space-time sparse grids are good candidates for an efficient discretization of parabolic problems.

In this section we now introduce and analyze three space-time sparse grid discretization schemes for parabolic problems. The first method is a sparse counterpart of the classical Crank-Nicolson scheme which can be interpreted as a Petrov-Galerkin scheme with piecewise linear continuous trial functions and piecewise constant test functions in time, see [24, 30], where it is also referred to as the *continuous Galerkin* (cG) method. The other two methods are based on the discontinuous Galerkin method (dG) [25] which is in contrast to the cG discretization a Galerkin scheme, i.e. the trial and ansatz spaces are the same. Here we investigate the cases with either piecewise constant or piecewise linear functions in time.

Throughout this section we use the spatial increment spaces W_j^Ω from (17) constructed from a sequence of finite element spaces and the nodal prolongation operator (18). Thus, the following discretization schemes differ only in the choice of the underlying bilinear form and the multilevel splittings in time used within the underlying space-time sparse grid trial or test spaces.

5.1 Continuous Galerkin (Crank-Nicolson)

Let us denote by $\tilde{V}_{l,tr}^0$ the space-time sparse grid space $\tilde{V}_l^0$ where we use the multilevel splitting $\{W_{j,1}^T\}$ of (12) in time within the construction. Furthermore, $\tilde{V}_{l,te}^0$ denotes the space $\tilde{V}_l^0$ where we use the piecewise constant multilevel splitting $\{W_{j,0}^T\}$ of (14).

We search for an $u_l \in \tilde{V}_{l,tr}^0$ such that u_l satisfies the boundary and initial conditions and

$$\int_0^T (\partial_t u_l, \varphi)_{L^2(\Omega)} + a(u_l, \varphi)\, dt = (f, \varphi)_{L^2(\Omega_T)}\, \forall \varphi \in \tilde{V}_{l,te}^0. \tag{49}$$

Note that the use of the full grids $\tilde{V}_{l,tr}^\infty$ and $\tilde{V}_{l,te}^\infty$ results in the classical Crank-Nicolson time discretization with a finite element discretization with respect to the spatial variable. Therefore we will refer to the sparse counterpart (49) of the Crank-Nicolson discretization as the space-time sparse grid Crank-Nicolson method (sCN).

Classical a priori error estimates for the full grid Crank-Nicolson scheme show second order convergence with respect to the spatial mesh width and the time step size for the error measured in the $\|\cdot\|_{L(\Omega)\otimes L^\infty(0,T)}$-norm, c.f. [59]. As we see from (24) we obtain the same order of convergence (up to a logarithmic factor) for the interpolation error of the trial space $V_{l,tr}^0$. Therefore we expect the same order of convergence for the sCN discretization itself and we really observe this error reduction in the numerical results of the next section. The proof for the classical Crank-Nicolson discretization is based on the separation of the error introduced by the time discretization and the error introduced by the spatial finite element discretization. This makes it very difficult to carry it over to the space-time sparse grid case and an a priori error estimate for the sCN discretization is a subject of further research. In the following we just give a simple stability result.

Lemma 5.1 *Let us assume that $a(\psi(\cdot,T), \psi(\cdot,T)) \geq 0$ for all $\psi \in \tilde{V}_{l,tr}^0$. For $u_l \in V_{l,tr}^0$ satisfying (49) with $u_l(x,0) = u_0(x) = 0$ for all $x \in \Omega$ we have*

$$\|\partial_t u_l\|_{L^2(\Omega_T)} \leq \|f\|_{L^2(\Omega_T)}. \tag{50}$$

Proof. We use $\partial_t u_l \in \tilde{V}_{l,const}^0$ as test function and obtain from (49) after partial integration

$$\int_0^T (\partial_t u_l, \partial_t u_l)_{L^2(\Omega)}\, dt + \frac{1}{2} a(u_l(\cdot,T), u_l(\cdot,T)) = \int_0^T f \partial_t u_l\, dt.$$

With $\frac{1}{2}a(u_l,u_l) \geq 0$ the statement follows.
□

From the above result it directly follows that the discrete problem is uniquely solvable. To this end, let u_l and $\tilde{u}_l$ be two solutions of (49) which fulfill the boundary and initial conditions. Then, $u_l - \tilde{u}_l$ is a solution of the discrete problem (49) with zero initial condition and $f = 0$. From Lemma 5.1 we obtain that $\partial_t(u_l - \tilde{u}_l) = 0$ and therefore $u_l - \tilde{u}_l = 0$.

5.2 Discontinuous Galerkin

We now investigate the Discontinuous Galerkin method [25] in the context of our space-time sparse grid approach. Here, for given $l \in \mathbb{N}$, we split the time domain $(0,T]$ into intervals $I_k = (t_{k-1}, t_k]$, $0 = t_0 < t_1 < \ldots < t_N = T$, $1 \leq k \leq N := 2^l$, of length $\delta t = 2^{-l} \cdot T$. For a given spatial discretization V_h^Ω with finite elements and $p \in \mathbb{N}$ we define the space

$$V_{N,p} := \{v : v|_{I_k} = \sum_{j=0}^{p} t^j v_j \text{ with } v_j \in V_l^\Omega\} \tag{51}$$

of functions which are piecewise polynomials with maximal degree p in time. Note that this space is generally discontinuous at the boundaries of the time intervals I_k. To be able to deal with these discontinuities we define the jump term

$$[v]_k := v_k^+ - v_k^- \text{ with } v_k^{+(-)} := \lim_{s \to 0^{+(-)}} v(t_k + s),$$

for every $v \in V_{N,p}$. Now, in the dG formulation for problem (35), we search for $u_l \in V_{N,p}$ such that $u_l(\cdot, 0) = u_0(\cdot)$ and

$$B(u_l, v) = \int_0^T (f, v)_{L^2(\Omega)}\, dt \text{ for all } v \in V_{N,p}, \tag{52}$$

where

$$B(u_l, v) := \sum_{k=1}^{N} \int_{I_k} (\partial_t u_l, v)_{L^2(\Omega)} + a(u_l, v) dt + \sum_{k=1}^{N} ([u_l]_{k-1}, v_{k-1}^+)_{L^2(\Omega)}. \tag{53}$$

We obtain with $V_{N,p}$ from (51) for $p = 0$ the implicit Euler method and for $p = 1$ a method similar to the subdiagonal Padé scheme of order $(2,1)$. For a deeper introduction to dG methods for parabolic problems, see [25, 26, 27, 28, 29].

Instead of the space $V_{N,p}$ we will now use our space-time sparse grid spaces $\tilde{V}_l^0$ and V_l^0 from (3) where we either employ the piecewise linear hierarchical spaces $\{W_{j,1}^{T,disc}\}$ from (15) or the piecewise constant hierarchical spaces $\{W_{j,0}^T\}$ from (14) in time. Here, we denote by sdG0 and sdG1 the resulting discretization method using the space-time sparse grid space $\tilde{V}_l^0$

with hierarchic increment spaces $\{W_{j,0}^T\}$ and $\{W_{j,1}^{T,disc}\}$, respectively. Furthermore, sdG0a denotes the dG method using the space V_l^0 with hierarchic increment spaces $\{W_{j,0}^T\}$ in time. In the following we denote by $P_{V_l^{sdG0}}$ ($P_{V_l^{sdG0a}}$ or $P_{V_l^{sdG1}}$) the interpolation operator which belongs to the underlying space-time sparse grid space of the sdG0 (sdG0a or sdG1) method.

In the following we will make use of the inequality

$$a \cdot b \leq \frac{\varepsilon}{2} a^2 + \frac{1}{2\varepsilon} b^2 \tag{54}$$

for arbitrary $a, b \in \mathbb{R}$ and $\varepsilon > 0$.

Lemma 5.2 *For all $v \in H^{1,0}_{mix}(\Omega_T)$ with $\partial_t v \in L^2(\Omega \times I_k)$ for all $1 \leq k \leq N$ and $v_0^- = 0$ we have*

$$B(v, v) = \int_0^T a(v, v) dt + \frac{1}{2} D(v), \tag{55}$$

with $D(v) := \|v_N^-\|^2_{L^2(\Omega)} + \sum_{k=1}^{N-1} \|[v]_k\|^2_{L^2(\Omega)}$.

In [31] a proof is given for the spaces $V_{N,p}$ which can directly be carried over to our space-time sparse grid settings. For positive definite $a(\cdot,\cdot)$ this result shows that the associated dG operator $B(\cdot,\cdot)$ is also positive definite. Therefore, the linear systems arising from the sdG0, sdG1 and sdG0a discretization are uniquely solvable.

The following theorem gives an a priori error estimate for the sdG0 and the sdG0a method.

Theorem 5.3 *Let us assume that there are constants $c > 0$ and $C > 0$ such that the continuity and the ellipticity conditions (1), (2) are fulfilled for $a(\cdot,\cdot)$. Let the solution u of problem (35) be sufficiently smooth. Then there exists a constant $\tilde{c}$ independent of l such that for the error $e_l := u - u_l$ between the solution u and the discrete solution u_l of the sdG0 method there holds*

$$\begin{aligned} \|\nabla(e_l)\|^2_{L^2(\Omega_T)} + D(e_l) \quad &\leq \quad \tilde{c} \cdot \Big(\|u - P_{V_l^{sdG0}} u\|^2_{H^1(\Omega)\otimes L^2((0,T))} \\ &\qquad + 2^l \|u - P_{V_l^{sdG0}} u\|^2_{L^2(\Omega)\otimes L^\infty((0,T))} \Big). \end{aligned}$$

The same estimate holds true for the sdG0a discretization error and the interpolant $P_{V_l^{sdG0a}}$.

Proof. With the ellipticity of $a(\cdot,\cdot)$ we obtain

$$\|\nabla(e_l)\|^2_{L^2(\Omega_T)} + D(e_l) \leq c \left(\int_0^T a(e_l, e_l) dt + \frac{1}{2} D(e_l) \right) = c \cdot B(e_l, e_l), \tag{56}$$

and it therefore suffices to give an upper estimate for $B(e_l, e_l)$. We assume that u is continuous at the discrete time points t_k, $1 \leq k \leq 2^l$. Then $[u]_k = 0$ for all $1 \leq k \leq 2^l$ and u fulfills (52). Therefore $B(e_l, \varphi) = 0$ for all $\varphi \in V_l^0$. Hence we have $B(e_l, e_l) = B(e_l, u - P_{V_l^{sdG0}} u)$ and we obtain

$$\begin{aligned} B(e_l, e_l) = B(e_l, u - P_{V_l^{sdG0}} u) &= \sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG0}} u)_{L^2(\Omega)} dt \quad (57) \\ + \int_0^T a(e_l, u - P_{V_l^{sdG0}} u) dt &+ \sum_{k=1}^{N} ([e_l]_{k-1}, (u - P_{V_l^{sdG0}} u)_{k-1}^{+})_{L^2(\Omega)}. \end{aligned}$$

With the abbreviation $\| \cdot \|_a^2 := a(\cdot, \cdot)$ we get

$$\begin{aligned} \int_0^T a(e_l, u - P_{V_l^{sdG0}} u) dt &\leq C \cdot \int_0^T \|e_l\|_a \|u - P_{V_l^{sdG0}} u\|_a \\ &\leq \frac{1}{2} \int_0^T \|e_l\|_a^2 \, dt + c \cdot \int_0^T \|u - P_{V_l^{sdG0}} u\|_a^2 \, dt, \end{aligned}$$

where we used (54) with suitably chosen ε to derive the last inequality. Furthermore, using (54) again, we obtain

$$\begin{aligned} &\sum_{k=1}^{N} ([e_l]_{k-1}, (u - P_{V_l^{sdG0}} u)_{k-1}^{+})_{L^2(\Omega)} \leq \\ &\sum_{k=1}^{N} \left(\frac{1}{4} \|[e_l]_{k-1}\|_{L^2(\Omega)}^2 + c^2 \cdot \|(u - P_{V_l^{sdG0}} u)_{k-1}^{+}\|_{L^2(\Omega)}^2 \right) \leq \\ &\frac{1}{4} D(e_l) + c^2 2^l \cdot \|u - P_{V_l^{sdG0}} u\|_{L^2(\Omega) \otimes L^\infty((0,T))}^2. \end{aligned}$$

Since u_l is constant on each interval I_k and therefore $\partial_t u_l \mid_{I_k} = 0$, we have

$$\begin{aligned} \sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG0}} u)_{L^2(\Omega)} dt &= \sum_{k=1}^{N} \int_{I_k} (\partial_t u, u - P_{V_l^{sdG0}} u)_{L^2(\Omega)} dt \\ &\leq \|\partial_t u\|_{L^2(\Omega_T)} \cdot \|u - P_{V_l^{sdG0}} u\|_{L^2(\Omega_T)}. \end{aligned}$$

By plugging these three estimates into relation (57) and using the relation (55) the assertion follows. The proof for the sdG0a method is completely analogous.
$\square$

In this proof we needed the fact that the basis functions in time are piecewise constant only for the estimation of the term $\sum_{k=1}^{N} \int_{I_k} (\partial_t u - u_l, u - P_{V_l^{sdG0}} u)_{L^2(\Omega)} dt$. Therefore it is easy to carry the above result over to the case of piecewise linear functions in time and thus to the sdG1 discretization. We then have the following theorem.

Theorem 5.4 *Let us assume that there are constants $c > 0$ and $C > 0$ such that the continuity and ellipticity conditions (1), (2) for $a(\cdot,\cdot)$ are fulfilled. Let the solution u of problem (35) be sufficiently smooth. Then there exists a constant $\tilde{c}$, independent of l, such that for the discrete solution u_l of the sdG1 discretization there holds*

$$\begin{aligned}
&\|\nabla e_l\|^2_{L^2(\Omega_T)} + D(e_l) \leq \\
&\tilde{c} \cdot \Big(\|u - P_l u\|^2_{H^1(\Omega)\otimes L^2((0,T))} + 2^{2l}\|u - P_l u\|^2_{L^2(\Omega)\otimes L^\infty((0,T))} + \\
&\qquad \|u - P_l u\|_{L^2(\Omega_T)} \cdot \|\partial_t (u - P_l u)\|_{L^2(\Omega_T)} \Big) .
\end{aligned}$$

Proof. The proof is analogous to that of Theorem 5.3, i.e. we have to estimate the three terms

$$\begin{aligned}
&\sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG1}} u)_{L^2(\Omega)} dt, \\
&\int_0^1 a(e_l, u - P_{V_l^{sdG1}} u) dt, \\
&\sum_{k=1}^{N} ([e_l]_{k-1}, (u - P_{V_l^{sdG1}} u)^+_{k-1})_{L^2(\Omega)}.
\end{aligned}$$

For the last two terms we can proceed completely analogously to Theorem 5.3. There, we did not make use of the fact that the space-time sparse grid functions were piecewise constant in time. Thus the estimates directly carry over to the case of the sdG1 discretization. It remains to estimate the first term. To this end, we use the fact that $\|\partial_t v\|_{L^2(\Omega_T)}$ is independent of t on each interval I_k for every $v \in V_l^0$. Therefore there is a constant $c > 0$ with

$$\|\partial_t v\|_{L^2(\Omega\times I_k)} \leq c 2^l \|v\|_{L^2(\Omega\times I_k)} \tag{58}$$

which is independent of l and $v \in V_l$. Now, using (58) we derive with $\Omega_k := \Omega \times I_k$

$$\begin{aligned}
&\sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG1}} u)_{L^2(\Omega)} dt \leq \sum_{k=1}^{N} \|\partial_t e_l\|_{L^2(\Omega_k)} \cdot \|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_k)} \leq \\
&\sum_{k=1}^{N} \|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_k)} \cdot \\
&\qquad \Big(\|\partial_t (u - P_{V_l^{sdG1}} u)\|_{L^2(\Omega_k)} + \|\partial_t (P_{V_l^{sdG1}} u - u_l)\|_{L^2(\Omega_k)} \Big) \leq \\
&\sum_{k=1}^{N} \|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_k)} \cdot \\
&\qquad \Big(\|\partial_t (u - P_{V_l^{sdG1}} u)\|_{L^2(\Omega_k)} + c2^l \|P_{V_l^{sdG1}} u - u_l\|_{L^2(\Omega_k)} \Big) .
\end{aligned}$$

The application of the Cauchy-Schwarz inequality to the sum leads to

$$\sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG1}} u)_{L^2(\Omega)} dt \leq$$
$$\|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_T)} \cdot \Big(\|\partial_t (u - P_{V_l^{sdG1}} u)\|_{L^2(\Omega_T)}$$
$$+ c2^l \|P_{V_l^{sdG1}} u - u_l\|_{L^2(\Omega_T)} \Big) \leq$$
$$\|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_T)} \cdot \Big(\|\partial_t (u - P_{V_l^{sdG1}} u)\|_{L^2(\Omega_T)}$$
$$+ c2^l (\|P_{V_l^{sdG1}} u - u\|_{L^2(\Omega_T)} + \|e_l\|_{L^2(\Omega_T)}) \Big) .$$

With the help of (54) and the Poincaré inequality we finally obtain

$$\sum_{k=1}^{N} \int_{I_k} (\partial_t e_l, u - P_{V_l^{sdG1}} u)_{L^2(\Omega)} dt \leq$$
$$\frac{1}{4} \|\nabla (u - u_l)\|_{L^2(\Omega_T)} + c \|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_T)} \cdot$$
$$\Big(\|\partial_t (u - P_{V_l^{sdG1}} u)\|_{L^2(\Omega_T)} + 2^{2l} \|u - P_{V_l^{sdG1}} u\|_{L^2(\Omega_T)} \Big) .$$

□

The above results show that the discretization errors of the sdG1, the sdG0 and the sdG0a method can be estimated by the interpolation errors of the respective trial spaces. Now, we can apply Lemma 3.1 to obtain direct estimates of the discretization errors. We then have

$$\|\nabla e_l\|^2 + D(e_l) \leq c_1(u) 2^{-2l} l^2$$

for the sdG1 method,

$$\|\nabla e_l\|^2 + D(e_l) \leq c_2(u) 2^{-l} l^2$$

for the sdG0 method and finally

$$\|\nabla e_l\|^2 + D(e_l) \leq c_3(u) 2^{-l}$$

for the sdG0a method. All constants are positive and depend on u.

5.3 Numerical Results

In this section we present numerical results for the proposed space-time sparse grid methods. Here, we are especially interested in the achieved accuracy and the associated computational cost and storage requirements.

In the following, besides the error measured in the norms $\|\cdot\|_{L^2(\Omega_T)}$, $\|\cdot\|_{L^\infty(\Omega_T)}$, $\|\cdot\|_{H^1\otimes L^\infty}$ and $\|\cdot\|_{H^1\otimes L^2}$, we are interested in the convergence rates

$$\rho_{L^\infty} := \frac{\|e_{l+1}\|_{L^\infty}}{\|e_l\|_{L^\infty}}, \qquad \rho_{L^2} := \frac{\|e_{l+1}\|_{L^2}}{\|e_l\|_{L^2}},$$
$$\rho_{H^1\otimes L^\infty} := \frac{\|e_{l+1}\|_{H^1\otimes L^\infty}}{\|e_l\|_{H^1\otimes L^\infty}}, \quad \rho_{H^1\otimes L^2} := \frac{\|e_{l+1}\|_{H^1\otimes L^2}}{\|e_l\|_{H^1\otimes L^2}},$$

of the different discretization methods. In all experiments we set $T = 1.0$ and we employ in the spatial finite element space the hierarchic increments constructed from a nested hierarchy of d-linear finite elements which are spanned by the d-dimensional hierarchical basis, i.e. in (17) we use the nodal prolongation operator.

Example 1.

In our first example we consider the problem

$$\begin{aligned} \partial_t u - \Delta u &= f \text{ for all } x \in \Omega = (0,1)^d, 0 < t \leq 1, \\ u(x,0) &= g(x) \text{ for all } x \in \Omega, \\ u(x,t) &= u_0(x) \text{ for all } t \in (0,1] \text{ and } x \in \partial\Omega, \end{aligned} \tag{59}$$

for $d = 1, 2, 3$. Here, we select the right hand side f, the boundary condition g and the initial condition u_0 so that

$$u(x,t) = \prod_{i=1}^{d} \sin(\pi \cdot \frac{x_i}{t+0.5}) \tag{60}$$

is the solution of the problem. Table 1 shows the results of the different space-time sparse grid discretization schemes for $d = 1$. We see that the sdG1 as well as the sCN method result in error reduction factors of about 0.3 for the L^∞- and the L^2-norms. This is the same behavior as predicted for the interpolation error of the space-time sparse grid spaces by our theory in Section 3. Note that the error for the sdG1 method is larger than the error for the sCN discretization, although the sCN Discretization involves only half the number of unknowns. For the $H^1 \otimes L^\infty$- and the $H^1 \otimes L^2$-norms both methods show a reduction rate of 0.5.

The sdG0 method provides a reduction rate of 0.5 for the L^2- and the L^∞-norms whereas the sdG0a method results in the same behavior as for the sdG1 and sCN discretizations. Here, the use of the trial space V_l^0 improves the rates, i.e. the reduction rate with respect to the L^2- and the L^∞-norms becomes about 0.3 and approaches 0.25 as the level index is increased, whereas for the other two norms a rate of 0.5 can be observed.

Note that although the sCN and the sdG1 discretizations provide similar error reduction rates, the sdG1 method leads to twice the number of degrees of freedom of the sCN method. Altogether, the sCN method is more efficient than the sdG0, the sdG0a and the sDg1 method with respect to the cost-benefit ratio in the case of a smooth solution.

Table 1: Results of the different discretization methods for problem (59), $d = 1$, with exact solution $u(x,t) = \sin(\pi \cdot \frac{x}{t+0.5})$.

l	e_{L^∞}	ρ_{L^∞}	e_{L^2}	ρ_{L^2}	$e_{H^1\otimes L^\infty}$	$\rho_{H^1\otimes L^\infty}$	$e_{H^1\otimes L^2}$	$\rho_{H^1\otimes L^2}$
				sCN				
4	1.079_{-2}	-	5.750_{-3}	-	1.755_{-1}	-	7.975_{-2}	-
5	2.856_{-3}	0.26	1.637_{-3}	0.28	9.012_{-2}	0.51	3.880_{-2}	0.49
6	8.636_{-4}	0.30	5.091_{-4}	0.31	4.536_{-2}	0.50	1.922_{-2}	0.50
7	2.518_{-4}	0.29	1.590_{-4}	0.31	2.272_{-2}	0.50	9.582_{-3}	0.50
8	7.303_{-5}	0.29	4.856_{-5}	0.31	1.136_{-2}	0.50	4.788_{-3}	0.50
9	2.076_{-5}	0.28	1.446_{-5}	0.30	5.680_{-3}	0.50	2.394_{-3}	0.50
10	5.817_{-6}	0.28	4.211_{-6}	0.29	2.840_{-3}	0.50	1.197_{-3}	0.50
11	1.611_{-6}	0.28	1.205_{-6}	0.29	1.420_{-3}	0.50	5.986_{-4}	0.50
				sdG1				
4	2.292_{-2}	-	8.730_{-3}	-	1.394_{-1}	-	6.907_{-2}	-
5	7.546_{-3}	0.33	2.736_{-3}	0.31	7.213_{-2}	0.52	3.158_{-2}	0.46
6	2.333_{-3}	0.31	8.292_{-4}	0.30	3.639_{-2}	0.50	1.507_{-2}	0.48
7	6.904_{-4}	0.30	2.446_{-4}	0.29	1.824_{-2}	0.50	7.378_{-3}	0.49
8	1.988_{-4}	0.29	7.065_{-5}	0.29	9.126_{-3}	0.50	3.657_{-3}	0.50
9	5.537_{-5}	0.28	7.527_{-5}	0.28	8.006_{-3}	0.50	4.925_{-3}	0.50
10	1.544_{-5}	0.28	2.112_{-5}	0.28	4.003_{-3}	0.50	2.459_{-3}	0.50
11	4.281_{-6}	0.28	5.820_{-6}	0.28	2.002_{-3}	0.50	1.229_{-3}	0.50
				sdG0				
4	3.296_{-2}	-	1.343_{-2}	-	1.159_{-1}	-	7.013_{-2}	-
5	1.879_{-2}	0.57	7.078_{-3}	0.53	7.709_{-2}	0.67	4.042_{-2}	0.58
6	1.117_{-2}	0.59	3.602_{-3}	0.51	4.706_{-2}	0.61	2.240_{-2}	0.55
7	6.180_{-3}	0.55	1.809_{-3}	0.50	2.726_{-2}	0.58	1.215_{-2}	0.54
8	3.276_{-3}	0.53	9.043_{-4}	0.50	1.527_{-2}	0.56	6.497_{-3}	0.53
9	1.693_{-3}	0.52	4.515_{-4}	0.50	8.374_{-3}	0.55	3.445_{-3}	0.53
10	8.624_{-4}	0.51	2.254_{-4}	0.50	4.524_{-3}	0.54	1.814_{-3}	0.53
				sdG0a				
4	2.250_{-2}	-	7.453_{-3}	-	1.042_{-1}	-	4.693_{-2}	-
5	8.005_{-3}	0.36	2.343_{-3}	0.31	5.414_{-2}	0.52	2.183_{-2}	0.47
6	2.606_{-3}	0.33	7.058_{-4}	0.30	2.731_{-2}	0.50	1.053_{-2}	0.48
7	8.027_{-4}	0.31	2.066_{-4}	0.29	1.367_{-2}	0.50	5.187_{-3}	0.49

The results for $d = 2$ in Table 2 and $d = 3$ in Table 3 show the same behavior. Here the sdG1 and the sCN method result in reduction rates above 0.25 in the L^2 and L^∞ norms, whereas we get a reduction rate of about 0.5 in the $H^1 \otimes L^2$- and $H^1 \otimes L^\infty$-norms. Again, the sdG1 method provides about the same convergence rate as the sCN method while it has more degrees of freedom. The sdG0 method provides convergence rates of about 0.5 in the L^2- and the L^∞-norms whereas the sdG0a discretization results in a convergence rate of about 0.3 with respect to the L^2- and L^∞-norms.

Overall, the presented results show that the behavior of the discretization error is analogous to the behavior of the interpolation error which was investigated theoretically in Section 3. For smooth solutions we see that the sCN, the sdG1 and the sdG0a discretizations show about the same error reduction. However, as Lemma 2.2 shows, the trial space of the sdG0a

Table 2: Results of the different discretization methods for problem (59), $d = 2$, with exact solution $u(x,t) = \prod_{i=1}^{2} \sin(\pi \cdot \frac{x_i}{t+0.5})$.

l	e_{L^∞}	ρ_{L^∞}	e_{L^2}	ρ_{L^2}	$e_{H^1\otimes L^\infty}$	$\rho_{H^1\otimes L^\infty}$	$e_{H^1\otimes L^2}$	$\rho_{H^1\otimes L^2}$
				sCN				
4	3.100_{-3}	-	7.043_{-4}	-	1.715_{-2}	-	8.187_{-3}	-
5	1.086_{-3}	0.35	2.526_{-4}	0.36	8.959_{-3}	0.52	3.973_{-3}	0.49
6	3.652_{-4}	0.34	8.482_{-5}	0.34	4.531_{-3}	0.51	1.947_{-3}	0.49
7	1.129_{-4}	0.31	2.689_{-5}	0.32	2.272_{-3}	0.50	9.636_{-4}	0.49
8	3.323_{-5}	0.29	8.175_{-6}	0.30	1.136_{-3}	0.50	4.798_{-4}	0.50
9	9.508_{-6}	0.29	2.411_{-6}	0.29	5.680_{-4}	0.50	2.396_{-4}	0.50
10	2.703_{-6}	0.28	6.954_{-7}	0.29	2.840_{-4}	0.50	1.197_{-4}	0.50
				sdG1				
4	3.652_{-3}	-	9.779_{-4}	-	1.352_{-2}	-	7.540_{-3}	-
5	1.246_{-3}	0.34	3.384_{-4}	0.35	7.164_{-3}	0.53	3.367_{-3}	0.45
6	4.056_{-4}	0.33	1.083_{-4}	0.32	3.637_{-3}	0.51	1.557_{-3}	0.46
7	1.239_{-4}	0.31	3.301_{-5}	0.30	1.825_{-3}	0.50	7.478_{-4}	0.48
8	3.646_{-5}	0.29	9.737_{-6}	0.29	9.128_{-4}	0.50	3.675_{-4}	0.49
9	1.047_{-5}	0.29	2.806_{-6}	0.29	4.564_{-4}	0.50	1.825_{-4}	0.50
10	2.956_{-6}	0.28	7.946_{-7}	0.28	2.282_{-4}	0.50	9.102_{-5}	0.50
				sdG0				
4	5.152_{-2}	-	1.399_{-2}	-	1.395_{-1}	-	8.176_{-2}	-
5	2.956_{-2}	0.57	7.417_{-3}	0.53	7.927_{-2}	0.57	4.652_{-2}	0.57
6	1.843_{-2}	0.62	3.787_{-3}	0.51	4.682_{-2}	0.59	2.544_{-2}	0.55
7	1.059_{-2}	0.57	1.907_{-3}	0.50	2.770_{-2}	0.59	1.362_{-2}	0.54
8	5.762_{-3}	0.54	9.551_{-4}	0.50	1.569_{-2}	0.57	7.208_{-3}	0.53
9	3.026_{-3}	0.53	4.774_{-4}	0.50	8.637_{-3}	0.55	3.785_{-3}	0.53
10	1.556_{-3}	0.51	2.385_{-4}	0.50	4.668_{-3}	0.54	1.977_{-3}	0.52
				sdG0a				
4	3.543_{-2}	-	8.461_{-3}	-	1.018_{-1}	-	5.504_{-2}	-
5	1.353_{-2}	0.38	2.793_{-3}	0.33	5.417_{-2}	0.53	2.397_{-2}	0.44
6	4.603_{-3}	0.34	8.680_{-4}	0.31	2.741_{-2}	0.51	1.099_{-2}	0.46
7	1.455_{-3}	0.32	2.597_{-4}	0.30	1.371_{-2}	0.50	5.274_{-3}	0.48
8	4.393_{-4}	0.30	7.563_{-5}	0.29	6.844_{-3}	0.50	2.595_{-3}	0.49

method has more degrees of freedom than the trial spaces for the sCN and sdG1 method. Moreover, the hierarchic increment spaces $\{W_{j,1}^{T,disc}\}$ from (15) which are used within the construction of the trial space for the sdG1 method have about twice the number of degrees of freedom of the $\{W_{j,1}^{T}\}$ increment spaces from (12). Hence, the underlying trial space of the sdG1 method has twice as many degrees of freedom as the trial space of the sCN method. Furthermore the sdG0 approach shows a reduction rate of about 0.5 for all norms but involves about the same number of degrees of freedom (dof) as the sCN method. Thus, in summary the sCN method shows the best cost-benefit ratio.

Example 2.

One advantage of our space-time sparse grids over classical sparse grids is the treatment of general domains. While due to the tensor product structure of classical sparse grids in space the treatment of more complicated domains

Table 3: Results of the different discretization methods for problem (59), $d = 3$, with exact solution $u(x,t) = \prod_{i=1}^{3} \sin(\pi \cdot \frac{x_i}{t+0.5})$.

l	e_{L^∞}	ρ_{L^∞}	e_{L^2}	ρ_{L^2}	$e_{H^1\otimes L^\infty}$	$\rho_{H^1\otimes L^\infty}$	$e_{H^1\otimes L^2}$	$\rho_{H^1\otimes L^2}$
				sCN				
3	1.220_{-2}	-	1.869_{-3}	-	2.389_{-2}	-	1.559_{-2}	-
4	4.901_{-3}	0.40	7.752_{-4}	0.41	1.452_{-2}	0.61	7.652_{-3}	0.49
5	1.635_{-3}	0.33	2.867_{-4}	0.37	7.714_{-3}	0.53	3.623_{-3}	0.47
6	5.506_{-4}	0.34	9.619_{-5}	0.34	3.921_{-3}	0.51	1.730_{-3}	0.48
7	1.704_{-4}	0.31	3.032_{-5}	0.32	1.967_{-3}	0.50	8.437_{-4}	0.49
				sdG1				
3	1.034_{-2}	-	2.349_{-3}	-	2.214_{-2}	-	1.608_{-2}	-
4	4.898_{-3}	0.47	9.670_{-4}	0.41	1.134_{-2}	0.51	7.383_{-3}	0.46
5	1.721_{-3}	0.35	3.460_{-4}	0.36	6.154_{-3}	0.54	3.197_{-3}	0.43
6	5.606_{-4}	0.33	1.126_{-4}	0.33	3.147_{-3}	0.51	1.418_{-3}	0.44
7	-	-	-	-	-	-	-	-
				sdG0				
3	1.105_{-1}	-	2.198_{-2}	-	2.455_{-1}	-	1.348_{-1}	-
4	6.705_{-2}	0.61	1.321_{-2}	0.60	1.516_{-1}	0.62	8.413_{-2}	0.62
5	3.965_{-2}	0.59	7.043_{-3}	0.53	8.307_{-2}	0.55	4.722_{-2}	0.56
6	2.534_{-2}	0.64	3.599_{-3}	0.51	4.482_{-2}	0.54	2.545_{-2}	0.54
7	1.495_{-2}	0.59	1.813_{-3}	0.50	2.438_{-2}	0.54	1.347_{-2}	0.53
				sdG0a				
3	1.084_{-1}	-	2.122_{-2}	-	2.353_{-1}	-	1.303_{-1}	-
4	4.783_{-2}	0.44	8.162_{-3}	0.38	8.902_{-2}	0.38	5.615_{-2}	0.43
5	1.889_{-2}	0.39	2.757_{-3}	0.34	4.688_{-2}	0.53	2.321_{-2}	0.41
6	6.567_{-3}	0.35	8.692_{-4}	0.32	2.384_{-2}	0.51	1.008_{-2}	0.43

Table 4: Results for Example 2 with exact solution $u(x,t) = u(x,t) = \sin(\pi \cdot \frac{x_1}{t+0.5}) \sin(\pi \cdot \frac{x_2}{t+0.5})$ on the domain depicted in Figure 2 for different numbers of unknowns (dof).

dof	e_{L^∞}	ρ_{L^∞}	e_{L^2}	ρ_{L^2}	$e_{H^1\otimes L^\infty}$	$\rho_{H^1\otimes L^\infty}$	$e_{H^1\otimes L^2}$	$\rho_{H^1\otimes L^2}$
				sCN				
5613	3.829_{-3}	-	7.469_{-4}	-	2.469_{-2}	-	1.333_{-2}	-
23532	9.401_{-4}	0.25	1.829_{-4}	0.24	1.317_{-2}	0.53	6.090_{-3}	0.46
95595	2.564_{-4}	0.27	5.010_{-5}	0.27	6.713_{-3}	0.51	2.924_{-3}	0.48
383850	7.057_{-5}	0.28	1.471_{-5}	0.29	3.376_{-3}	0.50	1.439_{-3}	0.49
1535337	2.192_{-5}	0.31	4.398_{-6}	0.30	1.691_{-3}	0.50	7.157_{-4}	0.50
				sdG1				
8313	4.682_{-3}	-	1.039_{-3}	-	1.945_{-2}	-	1.218_{-2}	-
35526	1.344_{-3}	0.29	3.690_{-4}	0.36	1.054_{-2}	0.54	5.288_{-3}	0.43
145650	4.126_{-4}	0.31	1.186_{-4}	0.32	5.395_{-3}	0.51	2.388_{-3}	0.45
587229	1.198_{-4}	0.29	3.543_{-5}	0.30	2.714_{-3}	0.50	1.130_{-3}	0.47
2352711	3.449_{-5}	0.29	1.013_{-5}	0.29	1.359_{-3}	0.50	5.513_{-4}	0.49
				sdG0				
5613	3.056_{-2}	-	6.703_{-3}	-	1.242_{-1}	-	7.455_{-2}	-
23532	1.729_{-2}	0.57	3.654_{-3}	0.55	8.656_{-2}	0.70	4.616_{-2}	0.62
95595	8.659_{-3}	0.50	1.849_{-3}	0.51	5.607_{-2}	0.65	2.694_{-2}	0.58
383850	4.204_{-3}	0.49	9.166_{-4}	0.50	3.400_{-2}	0.61	1.520_{-2}	0.56
1535337	2.051_{-3}	0.49	4.543_{-4}	0.50	1.974_{-2}	0.58	8.380_{-3}	0.55

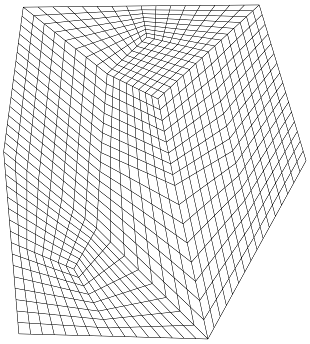


Figure 2: Domain with sample grid of the example 2.

is very difficult (or even impossible without losing convergence orders), our space-time sparse grids can be applied whenever a multilevel basis in space is present. In this example we consider such a situation. We employ the same parabolic operator as in the previous example with $d = 2$ but on a spatial domain Ω as depicted in Figure 2.

Again, we select the right–hand side, the boundary conditions and the initial condition so that the exact solution is

$$u(x,t) = \sin(\pi \cdot \frac{x_1}{t+0.5}) \cdot \sin(\pi \cdot \frac{x_2}{t+0.5}). \tag{61}$$

Table 4 shows the results for the different discretization schemes. For the norms which involve spatial derivatives, we obtain error reduction rates of about 0.5 for the sCN and the sdG1 scheme. For the L^2- and the L^∞-norms we obtain rates of 0.24-0.31 for the sCN discretization and for the sdG1 method. Note that in this example the difference between the cost-benefit ratio of the sCN and the sdG1 scheme is even more drastic than in the previous example. While we achieve an L^2-error of about $4.4 \cdot 10^{-6}$ with $1.5 \cdot 10^6$ degrees of freedom for the sCN scheme, the sDG1 scheme needs $2.4 \cdot 10^6$ degrees of freedom to provide an error of only about 10^{-5}.

The discretization with the sdG0 method on $\tilde{V}_l^0$ also leads to similar results as in the previous example. For the norms involving the spatial derivatives, we obtain convergence rates which start with 0.70 and 0.62 and which seem to approach 0.5 from above, whereas the reduction rate for the L^2- and L^∞-norms is about 0.5 on all levels.

6 Adaptivity

The numerical results of the last section show that the discontinuous Galerkin and the Crank-Nicolson discretization schemes provide the same rate of convergence as the respective interpolation error of the ansatz spaces. All the previous examples above had solutions which fulfill the smoothness requirements necessary for the interpolation properties in subsection 3.2. In general, however, if these regularity requirements are not fulfilled, we will no longer obtain the same interpolation and discretization errors on our space-time sparse grids. For the interpolation error this was already numerically observed in [35]. But there, with additional adaptive refinement of the space-time sparse grids, it was possible to recover the same cost-benefit ratio, i.e. the ratio of degrees of freedom versus error, for non-smooth functions as for the regular sparse grid case with smooth functions.

In the following we now apply an analogous adaptive refinement technique in space-time to obtain efficient discretization schemes for parabolic problems with non-smooth solutions.

Recall that with a basis $\{\psi^{\Omega}_{j,i}\}$ of the spatial increment space W^{Ω}_j and a basis $\{\psi^{T}_{j,i}\}$ of the increment space W^{T}_j we immediately obtain a basis $\{\psi_{\mathbf{j},\mathbf{i}} := \psi^{\Omega}_{j_1,i_1} \cdot \psi^{T}_{j_2,i_2}\}$ of the space $W_{\mathbf{j},\mathbf{i}}$. We may describe space-time sparse grids simply by a set of multi-indices associated to the basis functions which span the respective space. For the standard case of the regular space-time spare grid space $\tilde{V}^0_l$ we have the index set

$$G := \{(\mathbf{j},\mathbf{i}) \mid j_1 + j_2 \leq l,\ 1 \leq i_1 \leq \dim(W^{\Omega}_{j_1}),\ 1 \leq i_2 \leq \dim(W^{T}_{j_2})\}.$$

For an adaptive space-time sparse grid we will have more general index sets which contain the indices $(\mathbf{j},\mathbf{i})$ of the "active" basis functions. Such sets must be properly constructed in an adaptive fashion steered by local error indicators. To this end, let us define the spatial hierarchical relation $>$ between two indices (j_1, i_1) and $(\bar{j}_1, \bar{i}_1)$ as $(j_1, i_1) > (\bar{j}_1, \bar{i}_1) :\Leftrightarrow \operatorname{supp} \psi^{\Omega}_{\bar{j}_1,\bar{i}_1} \subset \operatorname{supp} \psi^{\Omega}_{j_1,i_1}$. We use the analogous definition for indices (j_2, i_2) and $(\bar{j}_2, \bar{i}_2)$ of the multilevel basis in time. For a basis function $\psi_{\mathbf{j},\mathbf{i}} = \psi^{\Omega}_{j_1,i_i} \cdot \psi^{T}_{j_2,i_2}$ we then define the set $\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}})$ of hierarchical sons as

$$\begin{aligned} \mathcal{H}(\psi_{\mathbf{j},\mathbf{i}}) := \{\psi_{\bar{\mathbf{j}},\bar{\mathbf{i}}} \mid (\bar{j}_1, \bar{j}_2) &= (j_1 + 1, j_2) \text{ or } (\bar{j}_1, \bar{j}_2) = (j_1, j_2 + 1), \\ \psi^{T}_{j_1,i_i} &> \psi^{T}_{\bar{j}_1,\bar{i}_1} \text{ and } \psi^{\Omega}_{j_2,i_2} > \psi^{\Omega}_{\bar{j}_2,\bar{i}_2}\}. \end{aligned} \tag{62}$$

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 has been already successfully used in several different application areas, e.g. in visualization [33], in numerical integration [34] and for the solution of elliptic partial differential equations [3, 11, 12, 40]. It results in a reliable but, in general, not perfectly efficient error estimator.[5] Nevertheless, numerical results show that the proposed adaptivity criterion provides quite good results.

[5]Note that the upper bound (which implies reliability) in the norm equivalency for the hierarchical basis is (after proper weighting) independent of the number of levels whereas the lower bound (which relates to efficiency) depends on the number of involved levels.

We proceed as follows: Given a sparse grid approximation

$$u^{sp} = \sum_{(\mathbf{j},\mathbf{i})\in G} u_{\mathbf{j},\mathbf{i}} \cdot \psi_{\mathbf{j},\mathbf{i}},$$

with associated set G of indices we add $\mathcal{H}(\psi_{\mathbf{j},\mathbf{i}})$ to the actual sparse grid space and thus the indices $(\tilde{\mathbf{j}}, \tilde{\mathbf{i}})$ of the hierarchical sons to the actual set G whenever

$$\gamma_{\mathbf{j},\mathbf{i}} |u_{\mathbf{j},\mathbf{i}}| > \varepsilon, \tag{63}$$

for a given threshold ε. Here $\gamma_{\mathbf{j},\mathbf{i}} := \|\psi_{\mathbf{j},\mathbf{i}}\|$ depends on the chosen norm $\|\cdot\|$. In the following we will simply use $\|\cdot\| = \|\cdot\|_{L^\infty}$ which led to good results in all our numerical experiments. Then we compute a discrete solution on the new, refined space-time sparse grid. We repeat this adaptive iteration process, the refinement of the space-time sparse grid and the computation of a discrete solution, until a predetermined fixed number of refinement steps or unknowns is reached.[6]

It remains to answer the question of how to choose the parameter ε. We follow an approach proposed in [52] where the parameter ε is derived from the current discrete solution. Here, we first determine ε_{max} from

$$\varepsilon_{max} := \max_{\substack{(\mathbf{j},\mathbf{i})\in G \wedge \\ \mathcal{H}(\psi_{\mathbf{j},\mathbf{i}}) \not\subset G}} \gamma_{\mathbf{j},\mathbf{i}} |u_{\mathbf{j},\mathbf{i}}|.$$

Then, in a second step for a fixed θ, $0 < \theta \leq 1$, we set $\varepsilon = \theta \varepsilon_{max}$.

For the following numerical experiments we use the parameter $\theta = 0.5$. To measure the error we replace the exact solution by its full grid interpolant on a level, which is two levels finer than the finest level index contained in the adaptive space-time sparse grid. Since the sCN method provided the best cost-benefit ratio in the previous experiments, we now focus our investigations on the sCN discretization only.

Example 3. In our first adaptive example we examine the problem

$$\begin{aligned} \partial_t u - \partial_{xx} u &= f \text{ in } \Omega = f \text{ in } \Omega_T = (0,1) \times (0,1) \\ u(x,0) &= u_0(x) \text{ for all } x \in \Omega = (0,1) \\ u(x,t) &= g(x,t) \text{ for all } t \in (0,1] \text{ and } x \in \partial\Omega, \end{aligned}$$

where the right–hand side f, the initial condition u_0 and the boundary condition g are such that the solution u is given by

$$u(x,t) = (x^2 + t^2)^{1/3}.$$

[6] We restricted ourselves for reasons of simplicity to this heuristic approach. Of course, more elaborate refinement schemes in the spirit of [6] or [17] may be envisioned which involve also global error estimation and thus may guarantee the reduction of the error to a prescribed tolerance.

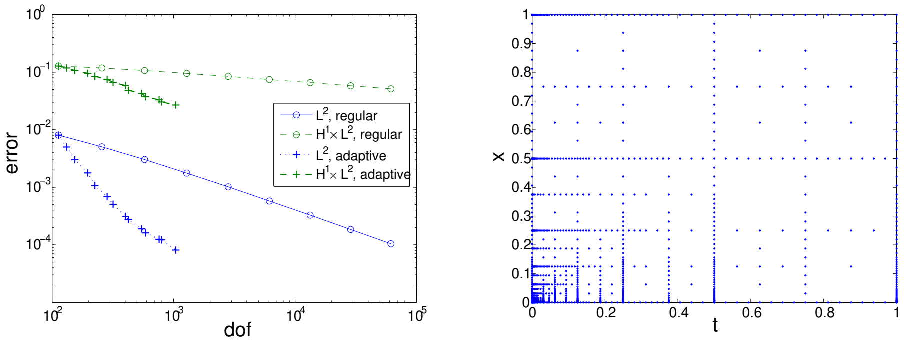


Figure 3: Error of the regular and adaptive sCN discretization, starting the refinement on the regular space-time sparse grid on level 4 (left), of Example 3 and the adaptive space-time sparse grid after 14 refinement steps (right).

Note that $u \in H^{5/3-\varepsilon}(\Omega_T)$ for every $\varepsilon > 0$, c.f. [40]. The results of the regular and adaptive sCN discretizations are shown in Figure 3. We started the refinement process on the regular space-time sparse grid on level $l = 4$, i.e. $\tilde{V}_4^0$, and allowed for 14 adaptive refinement steps. One can clearly see that the adaptive method is much more efficient than the discretization on a regular sparse grid. Here, the sCN discretization on a regular space-time sparse grid on level $l = 12$ with 61441 unknowns leads to an L^2-error of 6.47_{-5} and a $H^{1,0}_{mix}$-error of 4.59_{-2} whereas the adaptive discretization after 14 refinement steps provides an L^2-error of 8.88_{-6} and a $H^{1,0}_{mix}$-error of 5.61_{-3} with only 1675 unknowns.

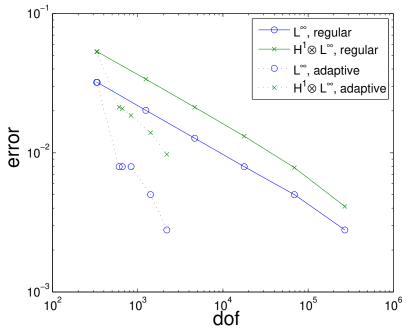
The adaptive grid shows a strong refinement towards the point $(0,0)$ where the singularity of the function u is located. Note that we directly obtain so-called *local time stepping*, i.e. different time steps are used in different parts of the spatial domain. Due to the time stepping structure of classical discretization methods, local time stepping is complicated to implement for full–grid discretizations, whereas the space-time sparse grid discretizations result in this type of adaptivity in a natural way and we do not have to invest additional implementational work.

Example 4. In this example we examine the two-dimensional problem

$$\begin{aligned}
\partial_t u - \partial_{xx} u &= f \text{ in } \Omega = f \text{ in } \Omega_T = (0,1)^2 \times (0,1) \\
u(x,0) &= u_0(x) \text{ for all } x \in \Omega = (0,1)^2 \\
u(x,t) &= g(x,t) \text{ for all } t \in (0,1] \text{ and } x \in \partial\Omega,
\end{aligned}$$

where the right-hand side f, the initial condition u_0 and the boundary condition g are such that the solution u represented in polar coordinates (r,ϕ) is given by

$$u(r,\phi,t) = \exp(-t) r^{2/3} \sin(\frac{2\phi - \pi}{3}).$$



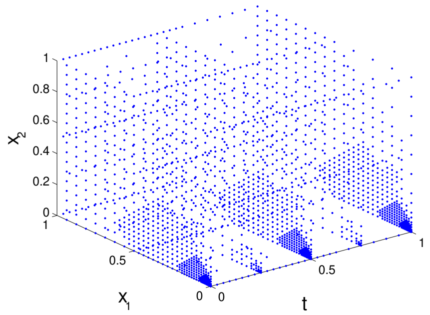


Figure 4: Error of the regular and adaptive sCN discretization (left) of Example 4 and an adaptive space-time sparse grid (right) resulting after 5 refinement steps starting from a regular grid on level $l = 3$.

Here a singularity is located for all times at the spatial point $(0, 0)$. Such a behavior of the solution of parabolic problems can typically be observed at re-entrant corners of the underlying spatial domain. Results for the regular and adaptive sCN discretization as well as a picture of an adaptive space-time sparse grid can be found in Figure 4. Here we started the refinement process with the regular space-time sparse grid $\tilde{V}_3^0$ and allowed for five adaptive refinement steps. Again, one can clearly see that the adaptive space-time sparse grid is much more efficient than the regular discretization. Concerning the L^∞-norm we obtain an error of 2.79_{-3} with only 2194 degrees of freedom, while we need 269822 degrees of freedom to obtain the same error for the discretization for $\tilde{V}_8^0$. As we can clearly see in Figure 4 the error indicator leads to a refinement towards the spatial singular point $(0, 0)$ at all time points. Since the function is smooth with respect to the time direction, local time stepping is now not necessary and adaptivity takes place in the spatial direction only.

7 Concluding Remarks

In this paper we presented new space-time sparse grid discretization schemes for parabolic problems based on the discontinuous Galerkin and Crank-Nicolson methods. These discretization methods save a substantial amount of degrees of freedom compared to the classical discontinuous Galerkin and Crank-Nicolson discretization schemes on full space-time grids. The additional order of complexity due to the time coordinate is completely avoided and only the complexity of the spatial problem remains. For the space-time sparse grid discontinuous Galerkin method we showed that we obtain nearly the same error reduction rates as in the full grid case if the solution fulfills certain smoothness assumptions. This theoretical result was reflected by the numerical experiments presented. Furthermore, our numerical results indicate that the same holds true for the space-time sparse grid Crank-Nicolson method for which we were however not able to give an a priori error estimate yet. This will be a subject of future research.

From classical regularity theory it follows that the smoothness requirements necessary for our space-time sparse grids to obtain the same rates as for full grids are really fulfilled if we impose proper assumptions on the initial and boundary conditions, on the right-hand side and on the spatial domain of the parabolic problem. For problems with non-smooth solutions we proposed an adaptive scheme which uses the norm of the hierarchic coefficients as error indicators. Numerical examples for the space-time sparse grid Crank-Nicolson method, which was the most efficient method in the smooth case, showed that the adaptive approach improves the results of the regular discretizations in the case of singular problems significantly and that nearly the same cost-benefit ratio could be obtained for non-smooth solutions as with regular space-time sparse grids in the smooth case.

Note finally that it is possible to solve the resulting linear systems with optimal complexity by a space-time multigrid method which will be presented in a forthcoming paper.

Altogether this makes our space-time sparse grid approach interesting as a simple parabolic extension to existing elliptic multilevel discretization and solution packages without involving any additional complexity order for the time part. Thus various parabolic problems and especially instationary optimization problems with three spatial dimensions in general geometries can be handled on standard workstations for the first time, see also [43].

References

[1] S. Adjerid, J. Flaherty, and Y. Wang, *Adaptive method-of-lines techniques for parabolic systems*, tech. rep., Rensselaer Polytechnic Institute, 1993.

[2] D. Arney and J. Flaherty, *An adaptive local mesh refinement method for time-dependent partial differential equations*, Appl. Numer. Math., (1989), pp. 257–274.

[3] R. Balder, *Adaptive Verfahren für elliptische und parabolische Differentialgleichungen*, Dissertation, Technische Universität München, 1994.

[4] R. Balder and C. Zenger, *The solution of the multidimensional real Helmholtz equation on sparse grids*, SIAM J. Sci. Comp., 17 (1996), pp. 631–646.

[5] G. Baszenski and F. Delvos, *Multivariate Boolean midpoint rules*, in Numerical Integration IV, H. Brass and G. Hämmerlin, eds., vol. 112 of ISNM, Birkhäuser, 1993, pp. 1–11.

[6] R. Becker and R. Rannacher, *A feed-back approach to error control in finite element methods: basic analysis and examples*, East-West J. Numer. Math., 4 (1996), pp. 237–264.

[7] M. Bieterman and I. Babuska, *An adaptive method of lines with error control for parabolic problems*, J. Comp. Phys., 63 (1986), pp. 33–66.

[8] T. Bonk, *Ein rekursiver Algorithmus zur adaptiven numerischen Quadratur mehrdimensionaler Funktionen*, Dissertation, Institut für Informatik, Technische Universität München, 1994.

[9] D. Braess, *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*, Cambridge University Press, 2001.

[10] H.-J. Bungartz, *An adaptive Poisson solver using hierarchical bases and sparse grids*, in Iterative Methods in Linear Algebra, Elsevier, 1992, pp. 293–310.

[11] H.-J. Bungartz, *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*, Dissertation, Technische Universität München, 1992.

[12] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica, 13 (2004), pp. 1–121.

[13] C. Chui and Y. Wang, *A general framework for compactly supported splines and wavelets*, J. Approx. Theory, 71 (1992), pp. 263–304.

[14] P. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics in Applied Mathematics, SIAM, 2002.

[15] A. Cohen, *Numerical Analysis of Wavelet Methods*, Studies in Mathematics and its Applications, Vol. 32, North Holland, 2003.

[16] A. Cohen and L. Echeverry, *Finite element wavelets*, tech. rep., Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, 2000.

[17] 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. 21–48.

[18] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, (1997), pp. 55–228.

[19] 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.

[20] I. Daubechies, *Orthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.

[21] T. Dornseifer, *Diskretisierung allgemeiner elliptischer Differentialgleichungen in krummlinigen Koordinatensystemen auf dünnen Gittern*, Dissertation, Technische Universität München, 1997.

[22] T. Dornseifer and C. Pflaum, *Discretization of elliptic differential equations on curvilinear bounded domains with sparse grids*, Computing, 56 (1996), pp. 197–213.

[23] J. Elf, P. Lötstedt, and P. Sjöberg, *Problems of high dimension in molecular biology*, in 17th Gamm Seminar Leipzig 2001, W. Hackbusch, ed., 2001, pp. 1–10.

[24] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge University Press, 1996.

[25] K. Eriksson and C. Johnson, *Adaptive finite element methods for parabolic problems I: A linear model problem*, SIAM J. Numer. Anal., 28 (1991), pp. 43–77.

[26] ——, *Adaptive finite element methods for parabolic problems II: Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$*, SIAM J. Numer. Anal., 32 (1995), pp. 706–740.

[27] ——, *Adaptive finite element methods for parabolic problems IV: Nonlinear problems*, SIAM J. Numer. Anal., 32 (1995), pp. 1729–1749.

[28] ——, *Adaptive finite element methods for parabolic problems V: Long-time integration*, SIAM J. Numer. Anal., 32 (1995), pp. 1750–1763.

[29] K. Eriksson, C. Johnson, and S. Larsson, *Adaptive finite element methods for parabolic problems VI: Analytic semigroups*, SIAM J. Numer. Anal., 35 (1998), pp. 1315–1325.

[30] D. Estep and D. French, *Global error control for the continuous Galerkin finite element method for ordinary differential equations*, Math. Anal. Numer., 28 (1994), pp. 815–852.

[31] D. French, *Discontinuous Galerkin finite element methods for a forward-backward heat equation*, Appl. Numer. Math., 27 (1998).

[32] J. Garcke and M. Griebel, *On the computation of the eigenproblems of hydrogen and helium in strong magnetic and electric fields with the sparse grid combination technique*, Journal of Computational Physics, 165 (2000), pp. 694–716.

[33] 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.

[34] T. Gerstner and M. Griebel, *Numerical integration using sparse grids*, Numerical Algorithms, 18 (1998), pp. 209–232.

[35] M. Griebel, D. Oeltz, and P. Vassilevski, *Space-time approximation with sparse grids*, SIAM J. Sci. Comput., 28 (2005), pp. 701–727.

[36] M. Griebel, P. Oswald, and T. Schiekofer, *Sparse grids for boundary integral equations*, Numer. Mathematik, 83 (1999), pp. 279–312.

[37] C. Grossmann and H.-G. Roos, *Numerik partieller Differentialgleichungen*, Teubner Verlag, 1992.

[38] W. Hackbusch, *The efficient computation of certain determinants arising in the treatment of Schrödinger's equation*, Computing, 67 (2001), pp. 35–56.

[39] S. Knapek, *Approximation und Kompression mit Tensorprodukt-Multiskalen-Approximationsäumen*, Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, 2000.

[40] F. Koster, *Multiskalen-basierte Finite-Differenzen-Verfahren auf adaptiven dünnen Gittern*, Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, 2001.

[41] J. Lang, *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems*, Springer, 2001.

[42] U. Mitzlaff, *Diffusionsapproximation von Warteschlangensystemen*, Dissertation, Institut für Mathematik, Technische Universität Clausthal, 1997.

[43] D. Oeltz, *Ein Raum-Zeit Dünngitterverfahren zur Diskretisierung parabolischer Differentialgleichungen*, Dissertation, Universität Bonn, 2006.

[44] S. Paskov, *Average case complexity of multivariate integration for smooth functions*, J. Complexity, 9 (1993).

[45] T. Petersdorff and C. Schwab, *Numerical solution of parabolic equations in high dimensions*, Mathematical Modelling and Numerical Analysis, 38 (2004), pp. 93–128.

[46] C. Pflaum, *Anwendung von Mehrgitterverfahren auf dünnen Gittern*, Diplomarbeit, Technische Universität München, 1992.

[47] J. Prakash, *Rouse chains with excluded volume interactions: Linear viscoelasticity*, Tech. Rep. Nr. 221, Berichte der Arbeitsgruppe Technomathematik, Universität Kaiserslautern, 2000.

[48] J. Prakash and H. Öttinger, *Viscometric functions for a dilute solution of polymers in a good solvent*, Macromolecules, 32 (1999), pp. 2028–2043.

[49] C. Reisinger, *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*, Dissertation, Universität Heidelberg, 2003.

[50] P. Rouse, *A theory of the linear viscoelastic properties of dilute solutions of coiling polymers*, J. Chem. Phys., 21 (1953), pp. 1272–1280.

[51] S. Schneider, *Extrapolationsmethoden zur Lösung parabolischer Gleichungen*, Diplomarbeit, Technische Universität München, 1994.

[52] S. Schneider and C. Zenger, *Multigrid methods for hierarchical adaptive finite elements*, tech. rep., Technische Universität München, 1999.

[53] C. Schwab and R. Todor, *Sparse finite elements for stochastic elliptic problems*, Tech. Rep. 2002-05, Semimar für Angewandte Mathematik, ETH Zürich, 2002.

[54] C. Schwab and R. Todor, *Sparse finite elements for stochastic elliptic problems-higher order moments*, Computing, 71 (2003), pp. 43–63.

[55] X. Shen, H. Chen, J. Dai, and W. Dai, *The finite element method for computing the stationary distribution on an SRBM in a hypercube with applications to finite buffer queueing networks*, Queuing Systems, 42 (2002), pp. 33–62.

[56] P. Sjöberg, *Numerical Solution of the Master Equation in Molecular Biology*, Master Thesis, Department for Scientific Computing, Uppsala Universität, 2002.

[57] R. Stevenson, *Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes*, Constr. Approx., 19 (2003), pp. 477–508.

[58] W. Sweldens, *The lifting scheme: A construction of second generation wavelets*, SIAM J. Math. Anal., 29 (1997), pp. 511–546.

[59] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, 1997.

[60] J. Wloka, *Partielle Differentialgleichungen*, Teubner, 1982.

[61] H. Yserentant, *On the multi-level splitting of finite element spaces*, Numer. Math., 58 (1986), pp. 379–412.

[62] ——, *Large Scale Scientific Computing*, Birkhäuser, 1987, ch. Hierarchical Bases in the Numerical Solution of Parabolic Problems, pp. 22–36.

[63] ——, *On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives*, Numer. Math., 98 (2004), pp. 731–759.

Michael Griebel
Institut für Numerische Simulation, Universität Bonn
Wegelerstr. 6
53115 Bonn
Germany
griebel@ins.uni-bonn.de

Daniel Oeltz
Institut für Numerische Simulation, Universität Bonn
Wegelerstr. 6
53115 Bonn
Germany
oeltz@ins.uni-bonn.de