A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations

Stevenson, R.; van Venetië, R.; Westerdiep, J.

DOI
10.1007/s10444-022-09930-w

Publication date
2022

Document Version
Final published version

Published in
Advances in Computational Mathematics

License
CC BY

Citation for published version (APA):
Stevenson, R., van Venetië, R., & Westerdiep, J. (2022). A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations. Advances in Computational Mathematics, 48(3), Article 17. https://doi.org/10.1007/s10444-022-09930-w
A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations

Rob Stevenson1 · Raymond van Venetië1 · Jan Westerdiep1

Received: 12 January 2021 / Accepted: 1 February 2022 / Published online: 6 April 2022
© The Author(s) 2022

Abstract
In this work, an r-linearly converging adaptive solver is constructed for parabolic evolution equations in a simultaneous space-time variational formulation. Exploiting the product structure of the space-time cylinder, the family of trial spaces that we consider are given as the spans of wavelets-in-time and (locally refined) finite element spaces-in-space. Numerical results illustrate our theoretical findings.

Keywords Space-time variational formulations of parabolic PDEs · Quasi-best approximations · Least squares methods · Adaptive approximation · Tensor product approximation · Optimal preconditioners

Mathematics Subject Classification (2010) 35K20 · 65F08 · 65M12 · 65M60 · 65T60

1 Introduction
This paper is about the adaptive numerical solution of parabolic evolution equations written in a simultaneous space-time variational formulation. In comparison to the usually applied time-marching schemes, simultaneous space-time solvers offer the following potential advantages:
• local, adaptive refinements simultaneous in space and time [21, 32, 39],
• quasi-best approximation from the selected trial space (‘Cea’s lemma’) [2, 26, 38], being a necessary requirement for proving optimal rates for adaptive routines [9, 25, 32],
• superior parallel performance [11, 22, 29, 41],
• using the product structure of the space-time cylinder, sparse tensor product approximation [9, 20, 25, 32] which allows to solve the whole time evolution at a complexity of solving the corresponding stationary problem.

Other relevant publications on space-time solvers include [10, 14, 27, 36, 40].

In any case without applying sparse tensor product approximation, a disadvantage of the space-time approach is the larger memory consumption because instead of solving a sequence of PDEs on a $d$-dimensional space, one has to solve one PDE on a $(d + 1)$-dimensional space. This disadvantage, however, disappears when one needs simultaneously the whole time evolution as for example with problems of optimal control [6, 19] or data-assimilation ([16]).

1.1 Parabolic problem in a simultaneous space-time variational formulation

For some separable Hilbert spaces $V \hookrightarrow H$ with dense embedding (e.g. $H^1_0(\Omega)$ and $L_2(\Omega)$ for the model problem of the heat equation on a spatial domain $\Omega \subset \mathbb{R}^d$), and a boundedly invertible $A(t) = A(t)^t : V \to V$ with $(A(t) \cdot \cdot) \simeq \| \cdot \|_V$ (e.g. $(A(t)\eta)(\zeta) = \int_\Omega \nabla \eta \cdot \nabla \zeta \, dx$), we consider

\[
\begin{cases}
\frac{du}{dt}(t) + A(t)u(t) = g(t) & (t \in (0, T)), \\
u(0) = u_0.
\end{cases}
\]

An application of a variational formulation of the PDE over space and time leads to an equation

$$
\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} u = \begin{bmatrix} g \\ u_0 \end{bmatrix}
$$

(1.1)

where, with $X := L_2(I; V) \cap H^1(I; V')$ and $Y := L_2(I; V)$, the operator at the left hand side is boundedly invertible $X \to Y \times H$.

1.2 Our previous work

In [9, 32], we equipped $X$ and $Y$ with Riesz bases being tensor products of wavelet bases in space in time, and $H$ with some spatial Riesz basis. Consequently, the equation (1.1) got an equivalent formulation as a bi-infinite well-posed matrix-vector equation

$$
\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} u = \begin{bmatrix} g \\ u_0 \end{bmatrix}
$$

(actually, in [32], we considered a formulation of first order, and in [9] we used a variational formulation with essentially interchanged roles of $X$ and $Y$, which however is irrelevant for the current discussion). To get a coercive bilinear form, we formed normal equations to which we applied an adaptive wavelet scheme ([8]). With such a scheme, the norm of a sufficiently accurate approximation of the (infinite) residual vector of a current approximation is used as an a posteriori error estimator. The coefficients in modulus of this vector are applied as local error
indicators in a bulk chasing (or Dörfler marking) procedure. The resulting adaptive algorithm converges at the best possible rate in linear computational complexity.

The goal of the current work is to investigate to what extent similar optimal theoretical results can be shown when we replace the wavelets-in-space by finite element spaces. In any case for locally refined partitions, a disadvantage will be that these spaces cannot be equipped with bases that are uniformly ‘stable’ w.r.t. a range of Sobolev norms. On the other hand, the application of finite elements has quantitative advantages, as provided by a very localized single-scale basis and efficient multi-grid preconditioners, and it avoids a cumbersome construction of wavelet bases on domains that are not hyperrectangles.

1.3 Least squares minimization

Without having Riesz bases for $X$ and $Y$, already the step of first discretizing and then forming normal equations does not apply, and we reverse their order. A problem equivalent to (1.1) is to compute

$$u = \arg \min_{w \in X} \| Bw - g \|_{Y'}^2 + \| \gamma_0 w - u_0 \|_H^2. \quad (1.2)$$

An obvious approach for the numerical approximation is to consider the minimization over finite dimensional subspaces $X^\delta$ of $X$, which however is not feasible because of the presence of the dual norm.

For trial spaces $X^\delta$ that are ‘full’ (or ‘sparse’) tensor products of finite element spaces in space and time, in [2] it was shown how to construct corresponding test spaces $Y^\delta \subset Y$ of similar type and dimension, such that $(X^\delta, Y^\delta)$ is uniformly inf-sup stable meaning that when the continuous dual norm $\| \cdot \|_{Y'}$ is replaced by the discrete dual norm $\| \cdot \|_{Y^\delta}$, a minimization over $X^\delta$ yields a quasi-best approximation to $u$ from $X^\delta$. Such a family of trial spaces however does not allow to create a nested sequence of trial spaces by adaptive local refinements.

1.4 Family of inf-sup stable pairs of trial and test spaces

To construct an alternative, considerably larger family, let $\Sigma$ be a wavelet Riesz basis for $L^2(0,T)$ that, after renormalization, is also a Riesz basis for $H^1(0,T)$. We equip this basis with a tree structure where every wavelet that is not on the coarsest level has a parent on the next coarser level. In space, we consider the collection of all linear finite element spaces that can be generated by conforming newest vertex bisection starting from an initial conforming partition of a polytopal $\Omega$ into $d$-simplices. The restriction to linear finite elements is not essential and is made for simplicity only. Now, we consider trial spaces $X^\delta$ that are spanned by a number of wavelets each of them tensorized with a finite element space from the aforementioned collection. In order to be able to apply the arising system matrices in linear complexity [24, 42], we impose the condition that if a wavelet tensorized with a finite element space is in the spanning set, then so is its parent wavelet tensorized with a finite element space that includes the former one.
The infinite collection of finite element spaces can be associated with a hierarchi-
cal ‘basis’ that can be equipped with a tree structure. Each hierarchical basis function,
except those on the coarsest level, is associated with a node \( v \) that was inserted
as the midpoint of an edge connecting two nodes on the next coarser level, which
nodes we call the parents of \( v \). With this definition there is a one-to-one correspon-
dence between the finite element spaces from our collection and the spans of the sets
of hierarchical basis functions that form trees. Consequently, our collection of trial
spaces \( X^\delta \) consists of the spans of sets of tensor products of wavelets-in-time and
hierarchical basis functions-in-space which sets are downwards closed, also known
as lower, in the sense that if a pair of a wavelet and a hierarchical basis function is
in the set, then so are all its parents in time and space. Spaces from this collection
can be ‘locally’ expanded by adding the span of a tensor product of a wavelet and
hierarchical basis function one-by-one.

For this family of spaces \( X^\delta \), we construct a corresponding family of spaces
\( Y^\delta \subset Y \) of similar type such that each pair \((X^\delta, Y^\delta)\) is uniformly inf-sup stable, with
the dimension of \( Y^\delta \) being proportional to that of \( X^\delta \). Furthermore, using the prop-
erties of the wavelets in time and by applying multigrid preconditioners in space, we
construct optimal preconditioners at \( X \) and \( Y \)-side which allow a fast solution of the
discrete problems \( \arg \min_{w \in X^\delta} \| Bw - g \|^2_{Y^\delta} + \| \gamma_0 w - u_0 \|^2_H \).

1.5 Adaptive algorithm

Having fixed the family of trial spaces, it remains to develop an algorithm that selects
a suitable, preferably quasi-optimal, nested sequence of spaces from the family
adapted to the solution \( u \) of (1.2). The theory on adaptive (Ritz–)Galerkin approx-
imations for such quadratic minimization problems is in a mature state. As noticed
before, however, straightforward Galerkin approximations for (1.2), i.e., approxima-
tions obtained by replacing \( X \) by \( X^\delta \) are not computable due to the presence of the
dual norm.

Therefore, given \( X^\delta \), let \( X^\tilde{\delta} \supset X^\delta \) be such that saturation holds, i.e., for some
constant \( \zeta < 1 \), it holds that \( \inf_{w \in X^\tilde{\delta}} \| u - w \|_X \leq \zeta \inf_{w \in X^\delta} \| u - w \|_X \). We now replace problem (1.2) by

\[
\tilde{u}^\delta = \arg \min_{w \in X^\tilde{\delta}} \| Bw - g \|^2_{Y^\tilde{\delta}} + \| \gamma_0 w - u_0 \|^2_H
\]

where in the notation \( \tilde{u}^\delta \) the first instance of \( \tilde{\delta} \) refers to the space \( Y^\tilde{\delta} \) and the second
to the space \( X^\tilde{\delta} \). Its (computable) Galerkin approximation from \( X^\delta \) is given by

\[
\tilde{u}^{\tilde{\delta}} = \arg \min_{w \in X^\delta} \| Bw - g \|^2_{Y^\tilde{\delta}} + \| \gamma_0 w - u_0 \|^2_H
\]

By a standard adaptive procedure, described below, we expand \( X^\delta \) to some \( X^\tilde\delta \subset X^\tilde{\delta} \)
such that \( \tilde{u}^{\tilde{\delta}} \) is nearer to \( \tilde{u}^{\tilde{\delta}} \) than \( \tilde{u}^{\tilde{\delta}} \) is. Next, we replace \( Y^\tilde{\delta} \) by \( Y^\tilde{\tilde{\delta}} \) (being the test
space corresponding to \( X^\tilde{\delta} \)) and repeat (i.e. consider (1.3) with \((\tilde{\delta}, \tilde{\delta})\) reading as \((\tilde{\delta}, \tilde{\delta})\),
and improve its Galerkin approximation \( u^{\tilde{\delta}} \) from \( X^\tilde{\delta} \) by an adaptive enlargement of
the latter space).
The adaptive expansion of the trial space $X^\delta$ to $\tilde{X}^\delta$ will be by the application of the usual solve-estimate-mark-refine paradigm, where the error indicators are given by the coefficients of the residual vector w.r.t. (modified) tensor product basis functions that were added to $X^\delta$ to create $\tilde{X}^\delta$. In order for this collection of additional tensor product basis functions to be stable in $X$-norm, for this step, we modify the hierarchical basis functions such that they get a vanishing moment, and therefore become closer to ‘real’ wavelets.

Under the aforementioned saturation assumption, we prove that the overall adaptive procedure produces an $r$-linearly converging sequence to the solution.

### 1.6 Numerical results

We tested the adaptive algorithm in several examples with a two-dimensional spatial domain. In all but one case, we observed a convergence rate equal to $1/2$, being the best that can be expected in view of the piecewise polynomial degree of the trial functions and the tensor product construction, and for non-smooth solutions improving upon usual non-adaptive approximation. Only for the case where $u_0 = 1$ and homogenous Dirichlet boundary conditions are prescribed, the observed rate was reduced to $0.4$. It is unknown whether or not this is the best non-linear approximation rate for our family of trial spaces.

Thanks to the use of optimal preconditioners and that of a carefully designed matrix-vector multiplication routine, which generalizes such a routine for sparse-grids introduced in [7] to adaptive settings, we observe that throughout the whole execution of the adaptive loop the total runtime remains proportional to the current number of unknowns.

### 1.7 Organization

This paper is organized as follows: in Section 2, the well-posed space-time variational formulation of the parabolic problem is discussed, and in Section 3, we discuss its inf-sup stable discretisation. The adaptive solution procedure is presented in Section 4, and its convergence is proven. The construction of the trial and test spaces is detailed in Section 5, and optimal preconditioners are presented. In Section 6, the definition of the enlarged space $X^\delta$ is given, and the construction of a stable basis of a stable complement space of $X^\delta$ in $X^\delta$ is outlined. Numerical results are presented in Section 7, and a conclusion is formulated in Section 8.

### 1.8 Notations

In this work, by $C \lessapprox D$ we will mean that $C$ can be bounded by a multiple of $D$, independently of parameters which $C$ and $D$ may depend on. Obviously, $C \gtrapprox D$ is defined as $D \lessapprox C$, and $C \approx D$ as $C \lessapprox D$ and $C \gtrapprox D$.

For normed linear spaces $E$ and $F$, by $\mathcal{L}(E, F)$ we will denote the normed linear space of bounded linear mappings $E \to F$, and by $\mathcal{L}(E, F)$ its subset of boundedly
invertible linear mappings \( E \to F \). We write \( E \leftrightarrow F \) to denote that \( E \) is continuously embedded into \( F \). For simplicity only, we exclusively consider linear spaces over the scalar field \( \mathbb{R} \).

## 2 Space-time formulations of a parabolic evolution problem

Let \( V, H \) be separable Hilbert spaces of functions on some ‘spatial domain’ such that \( V \leftrightarrow H \) with dense embedding. Identifying \( H \) with its dual, we obtain the Gelfand triple \( V \leftrightarrow H \cong H' \leftrightarrow V' \).

For a.e. \( t \in I := (0, T) \), let \( a(t; \cdot, \cdot) \) denote a bilinear form on \( V \times V \) such that for any \( \eta, \zeta \in V \), \( t \mapsto a(t; \eta, \zeta) \) is measurable on \( I \), and such that for some \( \varrho \in \mathbb{R} \), for a.e. \( t \in I \),

\[
|a(t; \eta, \zeta)| \lesssim \|\eta\|_V \|\zeta\|_V \quad (\eta, \zeta \in V) \quad \text{(boundedness)},
\]

\[
a(t; \eta, \eta) + \varrho \langle \eta, \eta \rangle_H \gtrsim \|\eta\|_V^2 \quad (\eta \in V) \quad \text{(Gårding inequality)}.
\]

With \( A(t) \in \text{Lis}(V, V') \) being defined by \( (A(t)\eta)(\zeta) = a(t; \eta, \zeta) \), given a forcing function \( g \) and an initial value \( u_0 \), we are interested in solving the parabolic initial value problem to finding \( u \) such that

\[
\begin{aligned}
\frac{du}{dt}(t) + A(t)u(t) &= g(t) \quad (t \in I), \\
u(0) &= u_0.
\end{aligned}
\]

In a simultaneous space-time variational formulation, the parabolic PDE reads as finding \( u \) from a suitable space of functions of time and space such that

\[
(Bw)(v) := \int_I \langle \frac{dw}{dt}(t), v(t) \rangle + a(t; w(t), v(t)) dt = \int_I \langle g(t), v(t) \rangle =: g(v)
\]

for all \( v \) from another suitable space of functions of time and space. One possibility to enforce the initial condition is by testing it against additional test functions. A proof of the following result can be found in [33], cf. [13, Ch.XVIII, §3] and [43, Ch. IV, §26] for slightly different statements.

**Theorem 2.1** With \( X := L^2(I; V) \cap H^1(I; V') \), \( Y := L^2(I; V) \), under conditions (2.1) and (2.2) it holds that

\[
\begin{bmatrix}
B \\
\gamma_0
\end{bmatrix} \in \text{Lis}(X, Y' \times H),
\]

where for \( t \in \bar{I} \), \( \gamma_t : u \mapsto u(t, \cdot) \) denotes the trace map. That is, assuming \( g \in Y' \) and \( u_0 \in H \), finding \( u \in X \) such that

\[
\begin{bmatrix}
B \\
\gamma_0
\end{bmatrix} u = \begin{bmatrix}
g \\
u_0
\end{bmatrix}
\]

is a well-posed simultaneous space-time variational formulation of (2.3).
With $\tilde{u}(t) := u(t)e^{-\varrho t}$, (2.3) is equivalent to $\frac{d}{dt}(A(t) + \varrho \text{Id})\tilde{u}(t) = g(t)e^{-\varrho t}$ ($t \in I$), $\tilde{u}(0) = u_0$. Since $(A(t) + \varrho \text{Id})\eta(\eta) \geq \|\eta\|^2_Y$, w.l.o.g. we assume that (2.2) is valid for $\varrho = 0$, i.e., $a(t; \cdot, \cdot)$ is coercive uniformly for a.e. $t \in I$.

For simplicity, additionally we assume that $a(t; \cdot, \cdot)$ is symmetric, and define $A = A' \in \mathcal{L}(Y, Y')$ by $(Aw)(v) = \int_I (A(t)w(t))v(t) \, dt$.

Because $\begin{bmatrix} A & 0 & B \\ 0 & \text{Id} & \gamma_0' \\ B' & \gamma_0 & 0 \end{bmatrix} \in \mathcal{L}(Y \times H, Y' \times H)$, an equivalent formulation of (2.4) as a self-adjoint saddle point equation reads as finding $(\mu, \sigma, u) \in Y \times H \times X$ (where $\mu = 0 = \sigma$) such that

$$
\begin{bmatrix} A & 0 & B \\ 0 & \text{Id} & \gamma_0' \\ B' & \gamma_0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \sigma \\ u \end{bmatrix} = \begin{bmatrix} g \\ u_0 \\ 0 \end{bmatrix},
$$

or equivalently

$$
\begin{bmatrix} A & B \\ B' - \gamma_0' & 0 \end{bmatrix} \begin{bmatrix} \mu \\ u \end{bmatrix} = \begin{bmatrix} g \\ -\gamma_0'u_0 \end{bmatrix},
$$

or

$$
(B'A^{-1} + \gamma_0'\gamma_0)u = B'A^{-1}g + \gamma_0'u_0.
$$

By employing $H \simeq H'$, here $\gamma_0' \in \mathcal{L}(H, X')$ is defined by $(\gamma_0'v)(w) := \langle v, \gamma_0w \rangle_H$.

We equip $Y$ and $X$ with ‘energy’-norms

$$
\| \cdot \|_X^2 := (A \cdot)(\cdot), \quad \| \cdot \|_Y^2 := \| \cdot \|_Y^2 + \| \partial_t \cdot \|_Y^2 + \| \gamma_T \cdot \|_H^2,
$$

which are equivalent to the canonical norms on $Y$ and $X$. Notice that (2.5)–(2.7) are the Euler–Langrange equations that result from the least squares problem (1.2).

**Lemma 2.2** We have $\| \cdot \|_X^2 = (S \cdot)(\cdot)$.

**Proof** It holds that

$$
\| w \|_X^2 = \sup_{0 \neq v_1 \in Y} \frac{(Bw)(v_1)}{\| v_1 \|_Y^2} + \gamma_0w_0 \| w \|_H^2 = \sup_{0 \neq (v_1, v_2) \in Y \times H} \frac{((Bw)(v_1) + \langle \gamma_0w, v_2 \rangle_H)^2}{\| v_1 \|_Y^2 + \| v_2 \|_H^2} = (Sw)(w),
$$

where the first equality is found in, e.g. [17, Thm. 2.1], and, seeing that $S$ is the Schur complement of the operator in (2.5), the last one in e.g. [23, Lem. 2.2].

### 3 Discretizations

#### 3.1 Galerkin discretization of the Schur complement equation

Let $(X^\delta)_{\delta \in \Delta}$ be a collection of closed, e.g. finite dimensional, subspaces of $X$, so equipped with $\| \cdot \|_X$. We will specify such a family in Section 5-6.1. We define a partial order on $\Delta$ by

$$
\delta \preceq \tilde{\delta} \iff X^\delta \subseteq X^{\tilde{\delta}}.
$$
For \( \delta \in \Delta \), let \( u_\delta \in X^\delta \) denote the Galerkin approximation to the solution \( u \) of (2.7), i.e., the solution of
\[
(Su_\delta)(v) = f(v) \quad (v \in X^\delta),
\]
being the best approximation to \( u \) from \( X^\delta \) w.r.t. \( \| \cdot \|_X \). This \( u_\delta \) is the solution of (1.2) with the minimization being performed over \( X^\delta \) instead of over \( X \).

For proving convergence of an adaptive solution routine, as well as for a posteriori error estimation, we shall make the following assumption.

**Assumption 3.1 (Saturation)** There exists a collection of subspaces \((\delta G \times \delta U_0)_{\delta \in \Delta} \subseteq Y' \times H\), a mapping \( \cdot : \delta \rightarrow \delta \) where \( \delta \geq \delta \), and some fixed constant \( \zeta < 1 \) such that for all \( \delta \in \Delta \), assuming that \((g, u_0) \in \delta G \times \delta U_0\),
\[
\| u - u_\delta \|_X \leq \zeta \| u - u_\delta \|_X.
\]

**Remark 3.2** Notice that above assumption cannot be valid without restrictions on the right-hand side \( f = B'A^{-1}g + \gamma_0 u_0 \in X' \). Indeed for any \( X^\delta \subseteq X^\delta \subseteq X \), consider a non-zero \( f \in X' \) that vanishes on \( X^\delta \). Then \( u_\delta = u_\delta = 0 \neq u \), violating (3.2).

For now, we will operate under the restrictive assumption that whenever we apply (3.2) (visible by the appearance of the constant \( \zeta \)) we simply assume that \((g, u_0) \in \delta G \times \delta U_0\). Later, in Section 4.3, we will remove this assumption.

The discretized problem from (3.1) only serves theoretical purposes. Indeed, since the Schur complement operator \( S \) contains the inverse of \( A \), there is no way to determine \( u_\delta \) exactly. The reason to introduce (3.1) is that \( S \) is an elliptic operator, so that for \( \delta \approx \delta \) we can make use of \( \| u - u_\delta \|_X^2 = \| u - u_\delta \|_X^2 - \| u_\delta - u_\delta \|_X^2 \), being a crucial tool for proving convergence of adaptive algorithms.

### 3.2 Uniformly stable Galerkin discretization of the saddle-point formulation

Our numerical approximations will be based on Galerkin discretizations of the saddle-point formulation (2.6). Let \((Y^\delta)_{\delta \in \Delta}\) be a collection of closed subspaces of \( Y \), so equipped with \( \| \cdot \|_Y \), such that
\[
X^\delta \subseteq Y^\delta \quad (\delta \in \Delta),
\]
and
\[
1 \geq \gamma_\Delta := \inf_{\delta \in \Delta} \inf_{0 \neq w \in X^\delta} \sup_{0 \neq v \in Y^\delta} \frac{(\partial_t w)(v)}{\| \partial_t w \|_Y \| v \|_Y} > 0.
\]

Notice that \( 1 - \gamma_\Delta \) can be made arbitrarily small by selecting, for each \( \delta \in \Delta \), \( Y^\delta \) sufficiently large in relation to \( X^\delta \).

For \( \delta, \delta' \in \Delta \) with \( Y^\delta \supseteq Y^{\delta'} \), and \( E_Y^\delta, E_X^\delta \) denoting the embeddings \( Y^\delta \rightarrow Y \), \( X^\delta \rightarrow X \), let \((\mu^{\delta, \delta'}, u^{\delta, \delta'}) \in Y^\delta \times X^\delta \) be the solution of
\[
\begin{bmatrix}
E_Y^{\delta'} E_Y^\delta & E_Y^{\delta'} E_X^\delta \\
E_X^{\delta'} B'E_Y^\delta & -E_X^{\delta'} \gamma_0' \gamma_0 E_X^\delta
\end{bmatrix}
\begin{bmatrix}
\mu^{\delta, \delta'} \\
u^{\delta, \delta'}
\end{bmatrix}
= \begin{bmatrix}
E_Y^{\delta'} g \\
-E_X^{\delta'} \gamma_0' \gamma_0 u_0
\end{bmatrix},
\]

or, equivalently,
\[
E_\delta \left( B' Y (E_\delta A E_\delta)^{-1} E_\delta B + \gamma_0 \right) E_\delta u \delta = \left( B' Y (E_\delta A E_\delta)^{-1} E_\delta \right) g + \gamma_0 u_0.
\] (3.6)

Below we show that (3.5)–(3.6) are uniquely solvable. In ‘operator language’, (3.5) is the Galerkin discretization of (2.6) on the closed subspace $Y \delta \times X \delta \subseteq Y \times X$. Unless $Y \delta = Y$, it holds that $S \delta \delta = E_\delta A E_\delta$ and $f \delta \delta \neq E_\delta f$, so generally $u \delta \delta \neq u \delta$.

As we will see, however, for $Y \delta$, and thus $Y \delta$, ‘large’ in relation to $X \delta$, $u \delta \delta$ will be ‘close’ to $u \delta$. This will allow us to show that (r-linear) convergence of a sequence of Galerkin solutions $u_\delta$ of (3.1) implies that of the corresponding sequence $u \delta \delta$.

We equip $X \delta$ with a family of ‘energy’ norms
\[
\|w\|_{X \delta}^2 := \|w\|_Y^2 + \sup_{0 \neq v \in Y \delta} \left( \frac{(\partial w)(v)^2}{\|v\|_Y^2} \right) + \|\gamma T w\|^2_H.
\]

By definition of $\gamma$, it holds that
\[
\gamma \Delta \| \cdot \|_X \leq \| \cdot \|_{X \delta \delta} \leq \| \cdot \|_X \quad \text{on} \quad X \delta.
\] (3.7)

Similar to Lemma 2.2, we have the following result.

Lemma 3.3 [38, Lem. 3.4] Thanks to (3.3) (and $Y \delta \subseteq Y \delta$), for $w \in X \delta$, we have
\[
\|w\|_{X \delta \delta}^2 = (S \delta \delta w)(w) = \sup_{0 \neq v \in Y \delta} \left( \frac{(B w)(v)^2}{\|v\|_Y^2} \right) + \|\gamma T w\|^2_H.
\]

By using additionally (3.4) this result shows that $(S \delta \delta)(\cdot)$ is coercive on $X \delta \times X \delta$ so that (3.6), and thus (3.5), has a unique solution.

Moreover, we have the following result.

Theorem 3.4 [38, Thm. 3.7] Thanks to (3.3) (and $Y \delta \subseteq Y \delta$) and (3.4), it holds that
\[
\|u - u_\delta\|_X \leq \|u - u \delta \delta\|_X \leq \gamma^{-1}_\Delta \|u - u_\delta\|_X.
\] (3.8)

3.3 Modified discretized saddle-point

In view of obtaining an efficient implementation, in the definition of $(\mu \delta \delta, u \delta \delta)$ in (3.5), and in that of $S \delta \delta$ and $f \delta \delta$ in (3.6), we replace $(E_\delta A E_\delta)^{-1}$ by some $K_\delta = K_\delta \delta \in \mathcal{L}(Y \delta, Y \delta)$ for which both
\[
\frac{(K_\delta - 1)(v)}{(A v)(v)} \in [\kappa_\Delta^{-1}, \kappa_\Delta] \quad \delta \in \Delta, \quad v \in Y \delta
\] (3.9)

for some constant $\kappa_\Delta \geq 1$ (i.e. $K_\delta$ is an optimal (self-adjoint and coercive) preconditioner for $E_\delta A E_\delta$), and which can be applied at linear cost. The resulting system
(3.6) is now amenable to the application of the (preconditioned) conjugate gradient iteration. Despite this modification, we keep using the old notations for $\mu^{\delta\delta}, u^{\delta\delta}, S^{\delta\delta}$, $\| \cdot \|_{X^{\delta\delta}} := \left( S^{\delta\delta} \cdot (\cdot)^{1/2} \right)$, and $f^{\delta\delta}$.

As shown in [38, Remark 3.8], instead of (3.8) now it holds that

$\| u - u^{\delta} \|_X \leq \| u - u^{\hat{\delta}\delta} \|_X \leq \kappa \gamma u - u^{\delta} \|_X$, (3.10)

whereas one deduces that (3.7) now should be read as

$\frac{\gamma \kappa}{\sqrt{\gamma \kappa}} \| \cdot \|_X \leq \| \cdot \|_{X^{\hat{\delta}\delta}} \leq \sqrt{\gamma \kappa \cdot \| \cdot \|_X}$ on $X^{\delta}$. (3.11)

For our forthcoming analysis, we will need $\frac{\gamma \kappa}{\gamma \kappa} - 1$ to be sufficiently small.

Remark 3.5 Later, in the proof of Proposition 4.5, temporarily we will consider the system (3.5) with $\hat{\delta} = \delta$ (i.e. $Y_{\hat{\delta}} = Y_{\delta}$), but with $X^{\delta}$ replaced by $X$, and, as we do in the current subsection, $E_{\hat{Y}} ' A E_{\hat{Y}}$ replaced by $(K_{\hat{Y}})^{-1}$. The resulting Schur operator $B' E_{\hat{Y}} K_{\hat{Y}} E_{\hat{Y}} B + \gamma_0 \gamma_0$ will be denoted as $S^{\delta\infty}$.

Note that the exact solution solves $S^{\delta\infty} u = B' E_{\hat{Y}} K_{\hat{Y}} E_{\hat{Y}} g + \gamma_0 u_0$. From $S^{\delta\delta} = E_{\hat{X}} S^{\delta\infty} E_{\hat{X}}$ we find the Galerkin orthogonality $(S^{\delta\infty} (u - u^{\delta\delta}))(X^{\delta}) = 0$. It holds that

$\| \cdot \|_{X^{\infty}} := (S^{\delta\infty} \cdot (\cdot)^{1/2} = \sqrt{(E_{\hat{Y}} ' B \cdot (K_{\hat{Y}} E_{\hat{Y}} ' B) + \| \gamma_0 \cdot \|_H)^2}$

is only a semi-norm on $X$, which is equal to $\| \cdot \|_{X^{\delta\delta}}$ on $X^{\delta}$, and

$\| \cdot \|_{X^{\infty}} \leq \sqrt{\gamma \kappa \cdot \| \cdot \|_X}$ on $X$. (3.12)

4 Convergent adaptive solution method

4.1 Preliminaries

For $\delta \in \Delta$, we consider the modified discretized saddle point problem (i.e. (3.6) with $(E_{\hat{Y}} ' A E_{\hat{Y}})^{-1}$ replaced by $K_{\hat{Y}}^{-1}$) taking $\hat{\delta} := \hat{\delta}$ from Assumption 3.1. So for a given ‘trial space’ $X^{\hat{\delta}}$, we employ $Y^{\hat{\delta}}$ as ‘test space’, which is known to be sufficiently large to give stability even when employed with trial space $X^{\hat{\delta}} \subseteq X^{\delta}$. We will use this room to (adaptively) expand $X^{\delta}$ to some $X^{\hat{\delta}} \subset X^{\hat{\delta}}$ while keeping $Y^{\hat{\delta}}$ fixed. Then in a second step, we adapt the test space to the new trial space, i.e., replace $Y^{\hat{\delta}}$ by $Y^{\hat{\delta}}$. By doing so will construct a sequence $(\delta_i) \subseteq \Delta$ with $\delta_i \leq \delta_{i+1}$ such that $(u^{\hat{\delta}\delta})_i$ converges r-linearly to $u$.

As a first step, in the next lemma it is shown that if one constructs from $w \in X^{\delta}$ a $v \in X^{\hat{\delta}}$ that is closer to the best approximation $u^{\hat{\delta}}$ to $u$ from $X^{\hat{\delta}}$, then, thanks to Assumption 3.1, $v$ is also closer to $u$.

Lemma 4.1 Let $w \in X^{\delta}, v \in X^{\hat{\delta}}$ be such that for some $\rho \leq 1$,

$\| u^{\hat{\delta}} - v \|_X \leq \rho \| u^{\hat{\delta}} - w \|_X$. 

\[ \rho \] Springer
Then,

\[\|u - v\|_X \leq \sqrt{\zeta^2 + \rho^2 (1 - \zeta^2)} \|u - w\|_X.\]

**Proof** Using \(u - u_{\delta} \perp X \quad X_{\delta}\) twice, we obtain

\[
\|u - v\|^2_X = \|u - u_{\delta}\|^2_X + \|u_{\delta} - v\|^2_X
\leq \|u - u_{\delta}\|^2_X + \rho^2 \|u_{\delta} - w\|^2_X
= \|u - u_{\delta}\|^2_X + \rho^2 (\|u - w\|^2_X - \|u - u_{\delta}\|^2_X)
= (1 - \rho^2) \|u - u_{\delta}\|^2_X + \rho^2 \|u - w\|^2_X
\leq (\zeta^2 (1 - \rho^2) + \rho^2) \|u - w\|^2_X,
\]

where we used Assumption 3.1 and \(\|u - u_{\delta}\|_X \leq \|u - w\|_X\).  

Notice that \(u_{\delta}^{\delta} \) is the Galerkin approximation from \(X_{\delta}\) to the solution \(u_{\delta}^{\delta} \in X_{\delta}\) of the system \(S_{\delta_{\delta}} u_{\delta}^{\delta} = f_{\delta_{\delta}}\), i.e. it is its best approximation from \(X_{\delta}\) w.r.t. \(\| \cdot \|_{X_{\delta}}\). In the next proposition, it is shown that an improved Galerkin approximation from an intermediate space \(X_{\delta} \supseteq X_{\tilde{\delta}} \supseteq X_{\delta}\), i.e. the function \(u_{\delta}^{\tilde{\delta}}\), is, for \(\frac{2\Lambda}{\gamma_{\Lambda}} - 1\) sufficiently small, also an improved approximation to \(u\), and that this holds true also for \(u_{\delta}^{\tilde{\delta}}\). The latter function will be the successor of \(u_{\delta}^{\delta}\) in our converging sequence.

**Proposition 4.2** Let \(\delta \leq \tilde{\delta} \leq \hat{\delta}\) be such that

\[
\|u_{\delta}^{\delta} - u_{\delta}^{\tilde{\delta}}\|_{X_{\delta}} \leq \rho \|u_{\delta}^{\delta} - u_{\delta}^{\hat{\delta}}\|_{X_{\delta}}. \tag{4.1}
\]

Then, it holds that

\[
\|u - u_{\delta}^{\tilde{\delta}}\|_X \leq \frac{\kappa_{\Lambda}}{\gamma_{\Lambda}} \sqrt{\zeta^2 + \hat{\rho}^2 (1 - \zeta^2)} \|u - u_{\delta}^{\delta}\|_X,
\]

where \(\hat{\rho} := \left(1 + \rho \frac{2\Lambda}{\gamma_{\Lambda}} \right) \sqrt{\frac{2\Lambda}{\gamma_{\Lambda}}} - 1 \sqrt{\frac{2\Lambda}{1 - \zeta^2}} + \rho \frac{2\Lambda}{\gamma_{\Lambda}} \). Notice that \(\hat{\rho}\) and \(\tilde{\rho}\) are < 1 when \(\rho < 1\) and \(\frac{2\Lambda}{\gamma_{\Lambda}} - 1\) is sufficiently small dependent on \(\rho\) with \(\frac{2\Lambda}{\gamma_{\Lambda}} - 1 \downarrow 0\) when \(\rho \uparrow 1\).

**Proof** Using that \(u - u_{\delta} \perp X \quad X_{\delta}\), it follows that \(\|u - u_{\delta}^{\delta}\|^2_X \leq \frac{\kappa^2_{\Lambda}}{\gamma_{\Lambda}} \|u - u_{\delta}\|^2_X\) (3.10) is equivalent to \(\|u_{\delta} - u_{\delta}^{\delta}\|_X \leq \sqrt{\frac{\kappa^2_{\Lambda}}{\gamma_{\Lambda}}} - 1 \|u - u_{\delta}\|_X\). Similarly, Assumption 3.1 is equivalent to \(\|u - u_{\delta}\|_X \leq \sqrt{\frac{\zeta^2}{1 - \zeta^2}} \|u_{\delta} - w\|_X\) for any \(w \in X_{\delta}\). Additionally using
\[ (3.11) \), we infer that
\[
\|u_\delta - u_\delta^\delta\|_X \leq \|u_\delta - u_\delta^\delta\|_X + \|u_\delta^\delta - u_\delta^\delta\|_X \\
\leq \|u_\delta - u_\delta^\delta\|_X + \frac{\sqrt{K_\Delta}}{\gamma_\Delta} \|u_\delta^\delta - u_\delta^\delta\|_X \\
\leq \|u_\delta - u_\delta^\delta\|_X + \rho \sqrt{K_\Delta} \|u_\delta^\delta - u_\delta^\delta\|_X \\
\leq (1 + \rho \frac{\sqrt{K_\Delta}}{\gamma_\Delta}) \|u_\delta - u_\delta^\delta\|_X + \rho \frac{\sqrt{K_\Delta}}{\gamma_\Delta} \|u_\delta^\delta - u_\delta^\delta\|_X \\
\leq \left[ (1 + \rho \frac{\sqrt{K_\Delta}}{\gamma_\Delta}) \sqrt{\frac{\zeta^2}{1 - \zeta^2}} + \rho \frac{\sqrt{K_\Delta}}{\gamma_\Delta} \right] \|u_\delta - u_\delta^\delta\|_X \\
\]

From Lemma 4.1, we conclude that \( \|u - u_\delta^\delta\|_X \leq \sqrt{\xi^2 + \tilde{\rho}^2(1 - \xi^2)} \|u - u_\delta^\delta\|_X \).

Thanks to (3.10), it holds that
\[
\|u - u_\delta^\delta\|_X \leq \frac{\kappa_\Delta}{\gamma_\Delta} \|u - u_\delta^\delta\|_X \leq \frac{\kappa_\Delta}{\gamma_\Delta} \|u - u_\delta^\delta\|_X ,
\]

which completes the proof. \( \square \)

**4.2 Bulk chasing and a posteriori error estimation**

To realize (4.1), i.e. to construct from the Galerkin approximation \( u_\delta^\delta \) to \( u_\delta^\delta \) an improved Galerkin approximation \( u_\delta^\delta \), we apply the concept of bulk chasing (or Dörfler marking) on a collection of a posteriori error indicators that constitute an efficient and reliable error estimator. We will apply an estimator of ‘hierarchical basis’ type ([45]):

Let \( \Theta_\delta = \{ \lambda : \lambda \in J_\delta \} \subseteq X_\delta \) be such that \( X_\delta + \text{span} \Theta_\delta = X_\delta \) and, for some constants \( 0 < m \leq M \), for all \( \delta \in \Delta, z \in X_\delta \) and \( c := (c_\lambda)_{\lambda \in J_\delta} \subset \mathbb{R} \),
\[
m^2 \|z + c^T \Theta_\delta\|^2_\delta \leq \|z\|^2_\delta + \|c\|^2 \leq M^2 \|z + c^T \Theta_\delta\|^2_\delta. \tag{4.2}
\]

A suitable collection \( \Theta_\delta \) will be constructed in Section 6.1.

**Proposition 4.3** Assume (4.2). Let \( r_\delta^\delta := (f_\delta^\delta - S_{\delta^\delta} u_\delta^\delta)(\Theta_\delta) \), being the residual vector of \( u_\delta^\delta \). Let \( J \subseteq J_\delta \) be such that for some constant \( \vartheta \in (0, 1] \),
\[
\|r_\delta^\delta|_J \geq \vartheta \|r_\delta^\delta\|,
\]
and, for some \( \tilde{\delta} \leq \delta \), let \( X^\delta + \text{span} \Theta_\delta|_J \subseteq X^\delta \). Then, with \( \rho := \sqrt{1 - \left( \frac{m}{M} \frac{\gamma_\Delta}{\kappa_\Delta} \vartheta \right)^2} \), (4.1) is valid, and so, when \( \frac{\kappa_\Delta}{\gamma_\Delta} - 1 \) is sufficiently small dependent on \( \vartheta \), with \( \frac{\kappa_\Delta}{\gamma_\Delta} - 1 \downarrow 0 \) when \( \vartheta \downarrow 0 \), for some constant \( \rho < \tilde{\rho} < 1 \),
\[
\|u - u_\delta^\delta\|_X \leq \tilde{\rho} \|u - u_\delta^\delta\|_X.
\]
Proof As a consequence of (4.2) and (3.11), we have
\[
\frac{\gamma}{\kappa} m^2 \|z + c^T \Theta_\delta\|^2_{\mathbb{H}^m} \leq \|z\|^2_{\mathbb{H}^m} + \|c\|^2 \leq \frac{k^2}{\gamma} M^2 \|z + c^T \Theta_\delta\|^2_{\mathbb{H}^m}.
\]
We infer that
\[
\|u_{\delta} - u_{\delta}\|_{\mathbb{H}^m} = \sup_{0 \neq (z,c) \in X^\delta \times \mathbb{R}^n} \frac{(S_{\delta}(u_{\delta} - u_{\delta}))(z + c^T \Theta_\delta)}{\|z + c^T \Theta_\delta\|^2_{\mathbb{H}^m}} \leq \frac{\kappa^2}{\gamma} M^2 \|z + c^T \Theta_\delta\|^2_{\mathbb{H}^m}.
\]
(4.3)
so that
\[
\|u_{\delta} - u_{\delta}\|_{\mathbb{H}^m} \leq \|u - u_{\delta}\|_{X} \leq \kappa^2 \gamma^2 \leq \frac{1}{\xi} ||u - u_{\delta}\||_{X}.
\]
which completes the proof of (4.1).

The final statement follows from an application of Proposition 4.2.

Moreover, \(\|r_{\delta}\|\) is an efficient and reliable a posteriori estimator for \(\|u - u_{\delta}\|_{X}\):

**Proposition 4.4** Assume (4.2). Recalling that \(\xi < 1\), let \(\frac{k^2}{\gamma} < \frac{1}{\xi}\). Then for \(\delta \in \Delta\),
\[
\frac{m \sqrt{\gamma}}{k^2} \frac{\sqrt{1/\xi - 1}}{1 + \xi} \leq \|u - u_{\delta}\||X \leq \frac{M^2 \Delta^2}{\gamma^2} \frac{1}{\sqrt{1 - \xi^2}} ||r_{\delta}\|. \tag{4.3}
\]

**Proof** Assumption 3.1 gives \(\|u - u_{\delta}\||X \leq \xi ||u - u_{\delta}\||X\). By \(u - u_{\delta} \perp X\), this yields
\[
\|u_{\delta} - u_{\delta}\| = \|u - u_{\delta}\| \leq \frac{1}{\sqrt{1 - \xi^2}} ||u_{\delta} - u_{\delta}\||X. \tag{4.4}
\]
As we already have noted in the proof of Proposition 4.2, (3.10) is equivalent to
\[
\|u_{\delta} - u_{\delta}\| \leq \sqrt{\frac{k^2}{\gamma}} - 1 ||u - u_{\delta}\||X. \]
Together with Assumption 3.1, it gives
which in combination with (4.4) and \( \zeta \frac{\kappa}{\gamma} < 1 \) yields
\[
\frac{1}{1 + \zeta \sqrt{\frac{\gamma}{\kappa}}} \| u^{\delta} - u^{\delta} \|_X \leq \| u - u^{\delta} \|_X \leq \frac{1}{\sqrt{1 - \zeta^2 - \zeta \sqrt{\frac{\gamma}{\kappa}}}} \| u^{\delta} - u^{\delta} \|_X.
\]

The proof is completed by (3.11) and \( m \frac{\sqrt{\gamma}}{\kappa} \| r^{\delta} \| \leq \| u^{\delta} - u^{\delta} \|_{X^{\bar{\delta}}} \leq M \frac{\kappa}{\sqrt{\gamma}} \| r^{\delta} \|, \)
where the latter inequalities follow from (4.3) by reading \((J, \vartheta, \tilde{\delta})\) as \((J^{\delta}, 1, \delta)\).

Next, we present an alternative a posteriori error estimator that does not rely on (4.2), that we expect to be more accurate, and that is computable at the cost of one additional inner product.

**Proposition 4.5** Let \( \frac{\kappa}{\gamma} < \frac{1}{\zeta} \). With notation from Remark 3.5, for \( v \in X^{\delta} \), we find
\[
\| u - v \|_{X^{\bar{\delta}}} = \sqrt{(E_Y^T (g - Bv))(K_Y^T E_Y^T (g - Bv)) + \| u_0 - \gamma_0 v \|_H^2}.
\]

It provides a computable, and efficient and reliable a posteriori error estimator. Indeed,
\[
\frac{\sqrt{\kappa}}{\gamma} \| u - v \|_X \leq \| u - v \|_{X^{\bar{\delta}}} \leq \sqrt{\kappa (\zeta^2 + (1 + \zeta \kappa \gamma)^2)} \| u - v \|_X.
\]

**Proof** The equality follows directly from \((Bu, \gamma_0 u) = (g, u_0)\).

Recall that \( \| \cdot \|_{X^{\bar{\delta}}} = \| \cdot \|_{X^{\bar{\delta}}} \) on \( X^{\delta} \), and that \((5^{\bar{\delta}}(u - u^{\delta}))(X^{\delta}) = 0 \), so that
\[
\| u - w \|_{X^{\bar{\delta}}} = \| u - u^{\delta} \|_{X^{\bar{\delta}}} + \| u^{\delta} - u \|_{X^{\bar{\delta}}} \quad (w \in X^{\delta}),
\]
and \( \| \cdot \|_{X^{\bar{\delta}}} \leq \sqrt{\kappa} \| \cdot \|_X \) (3.12). From (3.10) and Assum. 3.1, we have for \( v \in X^{\delta} \)
\[
\| u - u^{\delta} \|_X \leq \frac{\kappa}{\gamma} \| u - u^{\delta} \|_{X^{\bar{\delta}}} \leq \frac{\kappa}{\gamma} \| u - v \|_X.
\]

From (4.6), the triangle-inequality, (3.11), and (4.5) we obtain for \( v \in X^{\delta} \)
\[
\| u - v \|_X \leq \frac{1}{1 - \zeta \frac{\kappa}{\gamma}} \| u^{\delta} - v \|_{X^{\bar{\delta}}} \leq \frac{\sqrt{\kappa}}{\gamma - \zeta \kappa} \| u^{\delta} - v \|_{X^{\bar{\delta}}} \leq \frac{\sqrt{\kappa}}{\gamma - \zeta \kappa} \| u - v \|_{X^{\bar{\delta}}}.
\]

Conversely, we have
\[
\| u - v \|_{X^{\bar{\delta}}}^2 \overset{(4.5)}{=} \| u - u^{\delta} \|_{X^{\bar{\delta}}}^2 + \| u^{\delta} - v \|_{X^{\bar{\delta}}}^2 \overset{(4.5),(3.12)}{\leq} \| u - u^{\delta} \|_{X^{\bar{\delta}}}^2 + \kappa \| u^{\delta} - v \|_X^2 \overset{(3.12),(4.6)}{\leq} \kappa \| u - u^{\delta} \|_X^2 + \kappa (1 + \zeta \frac{\kappa}{\gamma}) \| u - v \|_X^2 \leq \kappa (\zeta^2 + (1 + \zeta \frac{\kappa}{\gamma})^2) \| u - v \|_X^2.
\]

by again applying Assumption 3.1. □
Notice that the estimator from Proposition 4.5 is exact when \( \zeta = 0 \) and \( \kappa = 1 = \gamma \Delta \), where the one from Proposition 4.3, for \( v = u^{\delta \tilde{g}} \), is exact only when additionally \( m = 1 = M \).

### 4.3 Data Oscillation

In view of the discussion following Assumption 3.1, notice that all results obtained so far that depend on the ‘saturation constant’ \( \zeta \), i.e. Lemma 4.1, and Propositions 4.2, 4.3, and 4.4, are only valid if \((g, u_0) \in \delta G \times \delta U_0\). In the following, we denote such solutions and residuals with an additional left superscript \( \tilde{\delta} \), or more generally \( \tilde{\delta} \) when a right-hand side \((\tilde{\delta} g, \tilde{\delta} u_0) \in \tilde{\delta} G \times \tilde{\delta} U_0 \) has been used for their computation.

**Proposition 4.6** Assume \((4.2)\), and let \( \vartheta \in (0, 1] \) be a constant. Then for \( \kappa = \gamma \Delta - 1 \) and a constant \( \delta \vartheta > 0 \) both being sufficiently small dependent on \( \vartheta \), with \( \max(\frac{\kappa \gamma}{\Delta} - 1, \delta \vartheta) \) and \( \eta(\delta) \leq \eta(\tilde{\delta}) \) when \( \tilde{\delta} \geq \delta \), and \( \eta(\delta(\varepsilon)) \leq \varepsilon \) when \( \delta(\varepsilon) \geq \delta \), it holds that

\[
\|\tilde{u} - u^{\delta \tilde{g}}\|_X \leq \tilde{\rho} \|\tilde{u} - u^{\delta \tilde{g}}\|_X.
\]

**Proof** In the introduced notations, the statements of Propositions 4.3 and 4.4 read

\[
\|\delta u - \delta u^{\delta \tilde{g}}\|_X \leq \tilde{\rho} \|\delta u - \delta u^{\delta \tilde{g}}\|_X,
\]

and

\[
\|\delta T^{\delta \tilde{g}}\| \approx \|\delta u - \delta u^{\delta \tilde{g}}\|_X \quad (4.7)
\]

The proof is easily completed by

\[
\frac{\|\tilde{u} - u^{\delta \tilde{g}}\|_X}{\|\tilde{u} - u^{\delta \tilde{g}}\|_X} < \|\tilde{g} - \delta \tilde{g}\|_Y + \|\delta u_0 - \delta u_0\|_H \leq 2\tilde{\omega} \|\delta T^{\delta \tilde{g}}\| \approx \tilde{\omega} \|\delta u - \delta u^{\delta \tilde{g}}\|_X.
\]

In view of the latter proposition, we make the following assumption.

**Assumption 4.7** We assume to have maps of the following types available:

\[
\Delta \to Y' \times H : \delta \mapsto (\delta g, \delta u_0) \in \delta G \times \delta U_0,
\]

\[
\eta : \Delta \to \mathbb{R} \text{ such that } \|g - \delta g\|_Y + \|u_0 - \delta u_0\|_H \leq \eta(\delta), \text{ and } \eta(\delta(\varepsilon)) \leq \eta(\delta) \text{ when } \delta(\varepsilon) \geq \delta,
\]

\[
\mathbb{R}_{>0} \to \Delta : \varepsilon \mapsto \delta(\varepsilon) \text{ such that } \eta(\delta(\varepsilon)) \leq \varepsilon.
\]

Notice that this in particular means that for any \( \varepsilon > 0 \) we are able to find a \( \delta \in \Delta \) and \((\delta g, \delta u_0) \in \delta G \times \delta U_0 \) with \( \|g - \delta g\|_Y + \|u_0 - \delta u_0\|_H \leq \varepsilon \). A specification of a suitable family \((\delta G, \delta U_0)_{\delta \in \Delta}\) will be given in Section 6.4.
Given a $\delta \in \Delta$, thinking of $(\delta g, \delta u_0)$ being a quasi-best approximation to $(g, u_0)$ from $\delta G \times \delta U_0$, the difference $(g, u_0) - (\delta g, \delta u_0)$ is often referred to as data-oscillation.

### 4.4 A convergent algorithm

In view of the statement of Proposition 4.6, in the following, we will use the shorthand notations

$$u^\delta = \delta u^\delta, \quad r^\delta = \delta r^\delta,$$

i.e. $u^\delta$ is the solution of

$$S^\delta = E_X^\delta (B' E_Y^\delta K^\delta_Y E_Y^\delta B + \gamma_0' \gamma_0) E_X^\delta u^\delta = E_X^\delta (B' E_Y^\delta K^\delta_Y E_Y^\delta g + \gamma_0' \gamma_0 u_0), \quad (4.8)$$

(cf. (3.6) and Section 3.3), and

$$r^\delta = E_X^\delta (B' E_Y^\delta K^\delta_Y E_Y^\delta g - Bu^\delta + \gamma_0 (u_0 - \gamma_0 u^\delta))(\Theta_{\delta}). \quad (4.9)$$

Instead of solving (4.8) exactly, we will allow it to be solved approximately with a sufficiently small relative tolerance by the application of an iterative method. To that end, we assume to have available a $K^\delta_X = K^\delta_X \in \mathcal{L}(X^{\delta}, X^{\delta})$ for which both

$$((K^\delta_X)^{-1} w)(w) \approx \| w \|^2_X \quad (w \in X^{\delta}) \quad (4.10)$$

(i.e. $K^\delta_X$ is an optimal (self-adjoint and coercive) preconditioner for $S^\delta$), and which can be applied at linear cost. Besides for an efficient iterative solving of (4.8), we will use this preconditioner to compute a quantity $t^\delta$ that is equivalent to the $X$-norm of the algebraic error in an approximation $\tilde{u}^\delta \in X^{\delta}$ to $u^\delta$. The residual vector $\tilde{r}^\delta$ corresponding to $\tilde{u}^\delta$ is defined as in (4.9) by replacing $u^\delta$ by $\tilde{u}^\delta$.

**Algorithm 4.8**

Let $\omega > 0$, $\vartheta \in (0, 1)$, $0 < \xi < 1$ be constants, and let $\varepsilon > 0$.

$$\delta := \delta_{\text{init}} \in \Delta, t_\delta \simeq \| g \|_Y + \| u_0 \|_H.$$

do

| do compute $\tilde{u}^\delta \in X^{\delta}$ with $\tilde{t}^\delta := \sqrt{(f^\delta - S^\delta \tilde{u}^\delta)(K^\delta_X (f^\delta - S^\delta \tilde{u}^\delta))} \leq \frac{\varepsilon}{2};$
| $t^\delta := \tilde{t}^\delta$
| if $e^\delta := \| \tilde{r}^\delta \| + \eta(\delta) + t^\delta \leq \varepsilon$ then stop
| until $t^\delta \leq \xi e^\delta$
| if $\eta(\delta) > \omega \| \tilde{r}^\delta \|$ \% i.e. data-oscillation dominates approximation error.
| then select $\tilde{\delta} \in \Delta$ s.t. $X^{\tilde{\delta}} \supseteq X^{\delta}$ is (a near-smallest) space such that $\eta(\tilde{\delta}) \leq \eta(\delta)/2$.
| else determine $\delta \leq \tilde{\delta} \leq \tilde{\delta}$ s.t. $X^{\tilde{\delta}}$ is (a near- smallest) space that for a $J \subseteq I^{\tilde{\delta}}$ contains $X^{\tilde{\delta}} + \text{span} \Theta_{\delta}|_J$ where $\| \tilde{r}^\delta |_J \geq \vartheta \| \tilde{r}^\delta \|$.
| endif
| $t^\delta := e^\delta, \delta := \tilde{\delta}$

end do
Theorem 4.9 Assume (4.2), and let the constants $\gamma_{\Delta}$ and $\kappa_{\Delta}$ be as defined in (3.4) and (3.9), respectively. For constants $\vartheta$, $\omega/\vartheta$, $\xi/\vartheta$, $(\frac{\kappa_{\Delta}}{\gamma_{\Delta}} - 1)/\omega$ sufficiently small, with additionally $\omega$ and $(\frac{\kappa_{\Delta}}{\gamma_{\Delta}} - 1)$ sufficiently small dependent on $\vartheta$ with $\max(\frac{\kappa_{\Delta}}{\gamma_{\Delta}}, 1, \omega) \downarrow 0$ when $\vartheta \downarrow 0$, there exists a constant $\bar{\rho} < 1$ such that between any two successive passings of the until-clause the value of $\vartheta u - \tilde{u}^\delta X + \eta(\delta)$ decreases with at least a factor $\bar{\rho}$. For any $\varepsilon > 0$, Algorithm 4.8 terminates, and at termination it holds that $\|u - \tilde{u}^\delta\|_X + \eta(\delta) \leq \varepsilon$.

Proof Replacing $w$ by $K^\delta_X S^\delta w$ in (4.10) yields $S^\delta v(K^\delta_X S^\delta v) \Rightarrow (S^\delta v)(K^\delta_X S^\delta v) = v^2$ for $v \in X^\delta$, so

\[
\|u^\delta - \tilde{u}^\delta\|^2_X \approx (f^\delta - S^\delta \tilde{u}^\delta)(K^\delta_X(f^\delta - S^\delta \tilde{u}^\delta)).
\]  

(4.11)

From (3.11) and (4.2), we have

\[
\|r^\delta - \tilde{r}^\delta\| = \sup_{0 \neq c \in \mathbb{R}^{X^\delta}} \|c\| \leq \frac{1}{\bar{\kappa}_{\Delta}} \|u^\delta - \tilde{u}^\delta\|_X \leq \frac{\kappa_{\Delta}}{m} \|u^\delta - \tilde{u}^\delta\|_X.
\]

(4.12)

If the algorithm stops, then

\[
\|u - \tilde{u}^\delta\|_X \leq \|u - u^\delta\|_X + \|u^\delta - \tilde{u}^\delta\|_X + \|u^\delta - \tilde{u}^\delta\|_X 
\leq \eta(\delta) + \|u - u^\delta\|_X + \|u^\delta - \tilde{u}^\delta\|_X
\]

(by Thm. 2.1 & Ass. 4.7)

\[
\approx \eta(\delta) + \|r^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X \leq \eta(\delta) + \|r^\delta\| + \|r^\delta - \tilde{r}^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X
\]

(4.7)

\[
\approx \eta(\delta) + \|\tilde{r}^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X \leq \eta(\delta) + \|\tilde{r}^\delta\| + t^\delta \leq \varepsilon.
\]

(4.10)

The inner do-loop always terminates either by passing the until-clause or by the stop-statement. Indeed, inside this loop the value of $t^\delta$ is driven to 0, so that $\tilde{r}^\delta + \eta(\delta)$ tends to $r^\delta + \eta(\delta)$. So if $\|\tilde{r}^\delta\| + \eta(\delta) \neq 0$, then at some moment $t^\delta \leq \xi(\|\tilde{r}^\delta\| + \eta(\delta) + t^\delta)$, whereas if $\|\tilde{r}^\delta\| + \eta(\delta) = 0$ then at some moment $e^\delta \leq \varepsilon$.

When passing the until-clause, it holds that $t^\delta \leq \xi(\|\tilde{r}^\delta\| + \eta(\delta) + t^\delta)$, and so by using $\xi < 1$ kicking back $t^\delta$,

\[
t^\delta \leq \xi(\|\tilde{r}^\delta\| + \eta(\delta))
\]

(4.13)

\[
\leq \xi(\|\tilde{r}^\delta\| + \|r^\delta - \tilde{r}^\delta\| + \eta(\delta))
\]

\[
\leq \xi(\|u - \tilde{u}^\delta\|_X + \|\tilde{u}^\delta - u^\delta\|_X + \eta(\delta))
\]

\[
\leq \xi(\|u - u^\delta\|_X + t^\delta + \eta(\delta)).
\]

Taking $\xi$ small enough and kicking back $t^\delta$, we obtain

\[
t^\delta \leq \xi(\|u - \tilde{u}^\delta\|_X + \eta(\delta))
\]

(4.14)

\[
t^\delta \leq \xi(\|u - u^\delta\|_X + \eta(\delta)).
\]

(4.15)

When passing the until-clause, furthermore, we have

\[
\|u - \tilde{u}^\delta\|_X \leq t^\delta + \|u - u^\delta\|_X \leq t^\delta + \|u - u^\delta\|_X + \eta(\delta) \approx t^\delta + \|r^\delta\| + \eta(\delta)
\]

(4.13)

\[
\leq t^\delta + \|r^\delta - \tilde{r}^\delta\| + \|\tilde{r}^\delta\| + \eta(\delta) \leq t^\delta + \|u^\delta - \tilde{u}^\delta\|_X + \|\tilde{r}^\delta\| + \eta(\delta)
\]

(4.10)

\[
\leq t^\delta + \|\tilde{r}^\delta\| + \eta(\delta) \leq \|\tilde{r}^\delta\| + \eta(\delta).
\]

(4.16)
Denoting \( \delta \) at the subsequent passing of the until-clause as \( \tilde{\delta} \), we have

\[
\|u - \tilde{u}\|_X \lesssim t_{\tilde{\delta}} + \|\tilde{u} - \tilde{u}\|_X + \eta(\tilde{\delta})
\]

\[
\leq t_{\tilde{\delta}} + \frac{K \Delta}{\gamma \Delta} \|\tilde{u} - \tilde{u}\|_X + \eta(\tilde{\delta})
\]

\[
\leq t_{\tilde{\delta}} + \frac{K \Delta}{\gamma \Delta} \|u - \tilde{u}\|_X + \eta(\tilde{\delta})
\]

\[
\leq t_{\tilde{\delta}} + \frac{K \Delta}{\gamma \Delta} \|u - \tilde{u}\|_X + \mathcal{O}(\eta(\tilde{\delta})�).
\]

By using \( t_{\tilde{\delta}} \lesssim \xi(\|u - \tilde{u}\|_X + \eta(\tilde{\delta})) \) (4.14) and kicking back \( \|u - \tilde{u}\|_X \), we infer that for \( \xi \) sufficiently small,

\[
\|u - \tilde{u}\|_X \leq \left(\frac{K \Delta}{\gamma \Delta} + \mathcal{O}(\xi)\right) \|u - \tilde{u}\|_X + \mathcal{O}(\eta(\tilde{\delta}))�.
\] (4.17)

In the case that

\[
\eta(\delta) > \omega \tilde{\delta}
\]

it holds that

\[
\eta(\tilde{\delta}) \leq \eta(\delta)/2,
\]

and so, thanks to (4.16) and (4.18),

\[
\|u - \tilde{u}\|_X \lesssim \eta(\delta)/\omega.
\] (4.19)

For any constant \( \rho_1 < 1 \), using (4.17) and (4.19), we have

\[
\|u - \tilde{u}\|_X \leq \rho_1 \|u - \tilde{u}\|_X + \left(\frac{K \Delta}{\gamma \Delta} + \mathcal{O}(\xi)\right) + (\rho_1/\omega + 1)C\eta(\delta),
\]

for some constant \( C > 0 \), and so

\[
\tilde{\vartheta}\|u - \tilde{u}\|_X + \eta(\tilde{\delta}) \leq \rho_1 \tilde{\vartheta}\|u - \tilde{u}\|_X + \left(\tilde{\vartheta}\left(\frac{K \Delta}{\gamma \Delta} - \rho_1 + \mathcal{O}(\xi)\right) + 1\right)C + \frac{1}{2})\eta(\delta)
\]

Now, let \( \tilde{\vartheta} > 0 \) be such that \( 2 \tilde{\vartheta} C + \frac{1}{2} < 1 \). Given a constant \( \omega \) (which later will be selected such that \( \omega/\tilde{\vartheta} \) is sufficiently small), let \( \left(\frac{K \Delta}{\gamma \Delta} - 1\right)/\omega, (1 - \rho_1)/\omega, \xi/\omega \) be sufficiently small such that the expression \( \frac{K \Delta}{\gamma \Delta} - \rho_1 + \mathcal{O}(\xi) \leq 1 \). We conclude that in this case \( \tilde{\vartheta}\|u - \tilde{u}\|_X + \eta(\delta) \) is reduced by at least a factor \( \max(\rho_1, 2 \tilde{\vartheta} C + \frac{1}{2}) < 1 \).

Next, we consider the other case that

\[
\eta(\delta) \leq \omega \|\tilde{\vartheta}\|_X.
\] (4.20)

so that

\[
\|\tilde{\vartheta}\|_X \geq \tilde{\vartheta}\|\tilde{\vartheta}\|_X.
\] (4.21)

We have

\[
\|\tilde{\vartheta}\|_X \leq \|u - \tilde{u}\|_X \lesssim \xi(\|\tilde{\vartheta}\|_X + \eta(\tilde{\delta})) \lesssim \xi\|\tilde{\vartheta}\|_X
\]

\[
\leq \xi(\|\tilde{\vartheta}\|_X + \|\vartheta\| - \tilde{\vartheta})_X),
\]

and so by taking \( \xi \) sufficiently small and kicking back \( \|\vartheta - \tilde{\vartheta}\|_X \), also

\[
\|\vartheta - \tilde{\vartheta}\|_X \lesssim \xi\|\vartheta\|_X.
\] (4.23)
which together with (4.20) implies that
\[ \eta(\delta) \lesssim \omega \| \vec{r}^\delta \|. \] (4.24)

From (4.22)–(4.23), we infer that
\[ \| \vec{r}^\delta \| \lesssim \| \vec{r}^\delta \| \lesssim \| \vec{r}^\delta \| + \| \vec{r}^\delta - \tilde{r}^\delta \| \lesssim \| \vec{r}^\delta \| + \xi \| \vec{r}^\delta \|. \] (4.23)

By taking \( \xi \) sufficiently small and kicking back \( \| \vec{r}^\delta \| \), we conclude that there exists a constant \( \vartheta > 0 \) such that
\[ \| \vec{r}^\delta \| \geq \vartheta \| \vec{r}^\delta \|. \] (4.21)

Assuming that \( \frac{\xi \alpha}{\gamma} - 1 \) and \( \omega \) are small enough, using (4.24) an application of Proposition 4.6 shows that there exists a constant \( \beta < 1 \) such that
\[ \| \tilde{u}^\delta - u^\delta \|_X \leq \rho \| \tilde{u}^\delta - u^\delta \|_X. \] (4.25)

Furthermore, we have
\[ \| \tilde{u}^\delta \| \leq \| \tilde{u}^\delta \| + \| \tilde{u}^\delta - u^\delta \| \lesssim \| u - \tilde{u}^\delta \|_X + \eta(\delta) + t^\delta \]
\[ \lesssim \| u - \tilde{u}^\delta \|_X + \eta(\delta), \] (4.13)

and so from (4.20) by kicking back \( \eta(\delta) \) and taking \( \omega \) sufficiently small,
\[ \eta(\delta) \lesssim \omega \| u - \tilde{u}^\delta \|_X. \] (4.26)

We conclude that
\[ \vartheta \| u - \tilde{u}^\delta \|_X + \eta(\delta) \leq \vartheta \| u - \tilde{u}^\delta \|_X + O(\eta(\tilde{\delta}) + t^\delta) \]
\[ \leq (\vartheta + O(\xi)) \| u - \tilde{u}^\delta \|_X + O(\eta(\tilde{\delta})) \]
\[ \leq (\vartheta + \xi \rho) \rho \| u - \tilde{u}^\delta \|_X + O(\eta(\tilde{\delta})) \]
\[ \leq (\vartheta + O(\xi)) \rho \| u - \tilde{u}^\delta \|_X + O(\eta(\tilde{\delta})) + t^\delta + \eta(\delta) \]
\[ \leq (\vartheta + \xi \rho + O(\xi)) \rho \| u - \tilde{u}^\delta \|_X + O(\eta(\tilde{\delta})) \]
\[ = \rho \| u - \tilde{u}^\delta \|_X. \] (4.26)

So for \( \omega / \vartheta \) and \( \xi / \vartheta \) sufficiently small, also in this case we established a reduction of \( \vartheta \| u - \tilde{u}^\delta \|_X + \eta(\delta) \) by at least a constant factor less than 1.

It is left to show that the algorithm terminates. We showed that the value of \( \vartheta \| u - \tilde{u}^\delta \|_X + \eta(\delta) \) at passing the until-clause is \( r \)-linearly converging. We consider the corresponding value of \( \epsilon_\delta = \| \vec{r}^\delta \| + \eta(\delta) + t^\delta. \) Arguments used multiple times show that \( \| \vec{r}^\delta \| \lesssim \| u - \tilde{u}^\delta \|_X + \eta(\delta) + t^\delta, \) and so \( \epsilon_\delta \lesssim \| u - \tilde{u}^\delta \|_X + \eta(\delta) + t^\delta. \) Using that \( t^\delta \leq \xi (\| \vec{r}^\delta \| + \eta(\delta)) \leq \xi \epsilon_\delta, \) for \( \xi \) sufficiently small kicking back \( \epsilon_\delta \) shows that
\[ \epsilon_\delta \lesssim \| u - \tilde{u}^\delta \|_X + \eta(\delta), \]
and so
\[ \epsilon_\delta \lesssim \vartheta \| u - \tilde{u}^\delta \|_X + \eta(\delta). \]

Therefore, at some moment \( \epsilon_\delta \leq \epsilon, \) meaning that the algorithm stops. \( \square \)
5 Wavelets-in-time tensorized with finite-elements-in-space

We specify the parabolic problem at hand, as well as the type of families \((X^δ)_{δ∈Δ}\) and \((Y^δ)_{δ∈Δ}\) of ‘trial’ and ‘test’ spaces. The specification of a likely harmless minor further restriction to these families that will be needed for the construction of an \(X\)-stable collection \(Θ_δ\) that spans an \(X\)-stable complement space of \(X^δ\) in \(X^δ\), specifically condition (4.2), will be postponed to Section 6.1.

5.1 Continuous problem

For some bounded domain \(Ω \subseteq \mathbb{R}^d\), we take \(H = L_2(Ω)\) and, for some closed \(Γ_D \subseteq \partial Ω\), \(V = H^1_0,Γ_D(Ω) := \text{clos}_{H^1(Ω)}\{u ∈ C^∞(Ω) \cap H^1(Ω) : u|_{Γ_D} = 0\}\), and

\[
a(t; η, ζ) := \int Ω K \nabla η \cdot \nabla ζ + c_ηζ dx,
\]

where \(K = K^T ∈ L_∞(I × I)\) with \(K(·) \approx \text{Id} \ a.e., c ∈ L_∞(I × I)\) with \(\text{ess inf } c ≥ 0\), and \(|Γ_D| > 0\) when the latter \(\text{ess inf}\) is zero. W.l.o.g. we take \(T = 1\).

5.2 Wavelets in time

We will construct the trial and test spaces as the span of wavelets-in-time tensorized with finite element spaces-in-space. In this subsection we collect some assumptions on the wavelets.

At the ‘trial side’ we consider a countable collection \(Σ = \{σ_λ : λ ∈ \mathcal{V}_Σ\}\) of functions \(I → \mathbb{R}\) known as wavelets. To each \(λ ∈ \mathcal{V}_Σ\), we associate a value \(|λ| ∈ \mathbb{N}_0\), called the level of \(λ\). We assume that the wavelets are locally supported meaning that \(\sup_{n ∈ \mathbb{N}, t ∈ \mathbb{N}_0} \#\{λ ∈ \mathcal{V}_Σ : |λ| = \ell, |\text{supp } σ_λ \cap 2^{-n}(0, 1)| > 0\} < ∞\) and \(\text{diam } \text{supp } σ_λ ≤ 2^{-|λ|}\). To each \(λ ∈ \mathcal{V}_Σ\) with \(|λ| > 0\), we associate one or more \(\tilde{λ} ∈ \mathcal{V}_Σ\) with \(|\tilde{λ}| = |λ| − 1\) and \(|\text{supp } σ_λ \cap \text{supp } σ_{\tilde{λ}}| > 0\) which we call the parent(s) of \(λ\). We denote this relation between a parent \(\tilde{λ}\) and a child \(λ\) by \(\tilde{λ} <_Σ λ\).

The definitions of parents and children give rise to obvious notions of ancestors and descendants.

To each \(λ ∈ \mathcal{V}_Σ\) we associate some neighbourhood \(S_Σ(λ)\) of \(\text{supp } σ_λ\) with \(\text{diam } S_Σ(λ) ≤ 2^{-|λ|}\) and

\[
λ <_Σ \tilde{λ} \implies S_Σ(λ) ⊇ S_Σ(\tilde{λ}).
\]

For some wavelets bases, e.g. Alpert wavelets ([1]), it suffices to take \(S_Σ(λ) = \text{supp } σ_λ\). With \(C_Σ := \sup_{λ ∈ \mathcal{V}_Σ} 2^{|λ|} \text{diam } \text{supp } σ_λ\), a neighbourhood that in any case is sufficiently large is \(\{t ∈ I : \text{dist}(t, \text{supp } σ_λ) ≤ C_Σ 2^{-|λ|}\}\). Indeed, if with this definition \(t ∈ S_Σ(λ)\) and \(\tilde{λ} <_Σ λ\), then \(\text{dist}(t, \text{supp } σ_{\tilde{λ}}) ≤ \text{dist}(t, \text{supp } σ_λ) + \text{diam}(\text{supp } σ_λ) ≤ 2C_Σ 2^{-|λ|}\), i.e. \(t ∈ S_Σ(\tilde{λ})\).

We assume that \(Σ\) is a Riesz basis for \(L_2(I)\), and, when renormalized in \(H^1(I)\)-norm, it is a Riesz basis for \(H^1(I)\). Although not essential, thinking of wavelets being
(essentially) constructed by means of dilation, we assume that
\[ \| \sigma_\lambda \|_{H^1(I)} \approx 2^{|\lambda|}. \]

At the ‘test side’ we consider a similar collection \( \Psi = \{ \psi_\mu : \mu \in \nu_\Psi \} \) of wavelets, with the difference though that this one has to be an even orthonormal basis for \( L_2(\Omega) \), whilst these wavelets do not need to be in \( H^1(I) \).

We will assume that \( \Sigma \) and \( \Psi \) are selected such that for any \( \ell \in \mathbb{N}_0 \),
\[ \text{span}\{ \sigma_\lambda : |\lambda| \leq \ell \} \cup \text{span}\{ \sigma'_\lambda : |\lambda| \leq \ell \} \subseteq \text{span}\{ \psi_\mu : |\mu| \leq \ell \}, \]
so that in particular
\[ |\mu| > |\lambda| \implies \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} = 0 = \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)}. \]  

(5.1)

### 5.3 Uniform stability

In the following proposition, we further specify the type of families of trial and test spaces that we consider, and formulate sufficient conditions for the requirements (3.3)–(3.4), which implied uniform stability of the Galerkin discretizations of our saddle-point problem (2.6).

**Proposition 5.1** For \( \delta \in \Delta \), let
\[ X^\delta = \sum_{\lambda \in \nu_\Sigma} \sigma_\lambda \otimes W^\delta_\lambda, \quad Y^\delta = \sum_{\mu \in \nu_\Psi} \psi_\mu \otimes V^\delta_\mu \]
for subspaces \( W^\delta_\lambda, V^\delta_\mu \subseteq V \) of which finitely many are non-zero. Assume
\[ \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} \neq 0 \implies V^\delta_\mu \supseteq W^\delta_\lambda, \]  and, for some constant \( \gamma_\Delta > 0 \), for any \( \mu \in \nu_\Psi \),
\[ \inf_{0 \neq w \in \sum_{\lambda \in \nu_\Sigma} W^\delta_\lambda : \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)} = 0} \sup_{0 \neq v \in V^\delta_\mu} \frac{w(v)}{\| w \|_V \| v \|_V} \geq \gamma_\Delta. \]  \[
(5.3)\]

Then \( X^\delta \subseteq Y^\delta \) and
\[ \inf_{0 \neq w \in X^\delta} \sup_{0 \neq v \in Y^\delta} \frac{(\partial_t w)(v)}{\| \partial_t w \|_Y \| v \|_Y} \geq \gamma_\Delta, \]
i.e. the conditions (3.3)–(3.4) for uniform stability are satisfied.

**Proof** For \( w_\lambda \in W^\delta_\lambda \) and \( w := \sum_{\lambda \in \nu_\Sigma} \sigma_\lambda \otimes w_\lambda \in X^\delta \), \( \sigma_\lambda = \sum_{\mu \in \nu_\Psi} \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} \psi_\mu \) shows that \( w = \sum_{\mu \in \nu_\Psi} \psi_\mu \otimes \sum_{\lambda \in \nu_\Sigma} \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} w_\lambda \in Y^\delta \) by the first assumption. Similarly \( \partial_t w = \sum_{\mu \in \nu_\Psi} \psi_\mu \otimes \tilde{v}_\mu \) where \( \tilde{v}_\mu := \sum_{\lambda \in \nu_\Sigma} \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)} w_\lambda \). For any \( \varepsilon > 0 \), the second assumption shows that for any \( \mu \in \nu_\Psi \), there exists a \( v_\mu \in V^\delta_\mu \) with \( \tilde{v}_\mu(v_\mu) \geq (\gamma_\Delta - \varepsilon) \| \tilde{v}_\mu \|_V \| v_\mu \|_V \) and \( \| \tilde{v}_\mu \|_V = \| v_\mu \|_V \). With \( v := \sum_{\mu \in \nu_\Psi} \psi_\mu \otimes v_\mu \in Y^\delta \), we infer that \( (\partial_t w)(v) = \sum_{\mu \in \nu_\Psi} \tilde{v}_\mu(v_\mu) \geq (\gamma_\Delta - \varepsilon) \| \partial_t w \|_Y \| v \|_Y \).

\( \square \)
In order to be able to apply at linear cost the arising linear operators in (4.8)–(4.9), we will restrict the type of trial spaces \( X^\delta = \sum_{\lambda \in \Sigma} \sigma_\lambda \otimes W^\delta_\lambda \) by imposing the following tree condition
\[
\tilde{\lambda} \prec \Sigma \lambda \implies W^\delta_{\tilde{\lambda}} \supseteq W^\delta_\lambda. \tag{5.4}
\]
For the same reason, the analogous condition will be needed for \( Y^\delta \). For \( X^\delta \) that satisfies (5.4), below the latter will be verified, and sufficient, more easily verifiable conditions for (5.2)–(5.3) are derived.

**Proposition 5.2** Let \( X^\delta = \sum_{\lambda \in \Sigma} \sigma_\lambda \otimes W^\delta_\lambda \) satisfy (5.4). For \( \mu \in \psi \), set
\[
\hat{W}^\delta_\mu := \sum_{\{\lambda \in \Sigma: |\lambda| = |\mu|, |S_\psi(\mu) \cap S_\Sigma(\lambda)| > 0\}} W^\delta_\lambda.
\]
Build \( Y^\delta = \sum_{\mu \in \psi} \psi_\mu \otimes V^\delta_\mu \) by taking \( V^\delta_\mu = \{0\} \) when \( \hat{W}^\delta_\mu = \{0\} \), and otherwise
\[
V^\delta_\mu \supseteq \hat{W}^\delta_\mu
\]
where
\[
\inf_{0 \neq w \in \hat{W}^\delta_\mu} \sup_{0 \neq v \in V^\delta_\mu} \frac{|w(v)|}{\|w\|_V \|v\|_V} \geq \gamma_\Delta \tag{5.5}
\]
for some constant \( \gamma_\Delta > 0 \). Then the conditions (5.2) and (5.3) from Proposition 5.1 for uniform stability are satisfied.

When \( \dim V^\delta_\mu \lesssim \dim \hat{W}^\delta_\mu \), then \( \dim Y^\delta \lesssim \dim X^\delta \), and under the natural condition that a larger \( \hat{W}^\delta_\mu \) gives rise to a larger (more precisely, not smaller) \( V^\delta_\mu \), the constructed \( Y^\delta \) satisfies the tree condition
\[
\tilde{\mu} \prec \psi \mu \implies V^\delta_{\tilde{\mu}} \supseteq V^\delta_\mu. \tag{5.6}
\]

**Proof** Let \( \langle \sigma_\lambda, \psi_\mu \rangle_{L^2(I)} \neq 0 \) or \( \langle \sigma'_\lambda, \psi_\mu \rangle_{L^2(I)} \neq 0 \). Then \( |S_\Sigma(\lambda) \cap S_\psi(\mu)| > 0 \) and \( |\lambda| \geq |\mu| \) by (5.1). When \( |\lambda| > |\mu| \), \( \lambda \) has an ancestor \( \tilde{\lambda} \) with \( |\tilde{\lambda}| = |\mu| \), \( W^\delta_{\tilde{\lambda}} \supseteq W^\delta_\lambda \), and \( S_{\Sigma}(\tilde{\lambda}) \supseteq S_{\Sigma}(\lambda) \), and thus \( |S_{\Sigma}(\tilde{\lambda}) \cap S_\psi(\mu)| > 0 \). We conclude that both \( \sum_{|\lambda| \in \Sigma: \langle \sigma_\lambda, \psi_\mu \rangle_{L^2(I)} \neq 0} W^\delta_{\lambda} \) and \( \sum_{|\lambda| \in \Sigma: \langle \sigma'_\lambda, \psi_\mu \rangle_{L^2(I)} \neq 0} W^\delta_{\lambda} \) are included in \( \hat{W}^\delta_\mu \), so that (5.2) and (5.3) are guaranteed by the selection of \( V^\delta_\mu \).

The statement \( \dim Y^\delta \lesssim \dim X^\delta \) when \( \dim V^\delta_\mu \lesssim \dim \hat{W}^\delta_\mu \) follows from \( \hat{W}^\delta_\mu \leq \sum_{|\lambda| \in \Sigma: |\lambda| = |\mu|, |S_\psi(\mu) \cap S_\Sigma(\lambda)| > 0} \dim W^\delta_\lambda \), and the fact that for any \( \lambda \in \Sigma \), the number of \( \mu \in \psi \) with \( |\mu| = |\lambda| \) and \( |S_\psi(\mu) \cap S_\Sigma(\lambda)| > 0 \) is uniformly bounded.

Let \( \tilde{\mu} \prec \psi \mu \), and so \( S_\psi(\tilde{\mu}) \supseteq S_\psi(\mu) \). For each \( \lambda \in \Sigma \) with \( |\lambda| = |\mu| \) and \( |S_\psi(\mu) \cap S_\Sigma(\lambda)| > 0 \), there exists a \( \tilde{\lambda} \prec \lambda \), thus with \( S_{\Sigma}(\tilde{\lambda}) \supseteq S_{\Sigma}(\lambda) \), and \( W^\delta_{\tilde{\lambda}} \supseteq W^\delta_\lambda \) by (5.4). We conclude that \( \hat{W}^\delta_{\tilde{\mu}} \supseteq \hat{W}^\delta_\mu \), which completes the proof of (5.6). \( \square \)

As follows from [16, Thm. 3.11] (taking \( B \) to be the Riesz map \( H \to H' \)) condition (5.5) has the following equivalent formulation.
Proposition 5.3 Condition (5.5) is equivalent to existence of uniformly bounded projectors $Q_\mu \in \mathcal{L}(V, V)$ with $Q_\mu \subseteq V^\delta_\mu$, $\text{ran } Q_\mu \supseteq \hat{W}^\delta_\mu$. Moreover, it holds that $rac{1}{\gamma_\Delta} \leq \sup_{\mu \in \mathcal{M}} \| Q_\mu \|_{\mathcal{L}(V, V)} \leq 2 + \frac{1}{\gamma_\Delta}$.

5.4 Selection of the spatial approximation spaces as finite element spaces

We will select the spaces $W^\delta_\lambda$ from a collection $\mathcal{O}$ of finite element spaces in $V$, which collection is closed under taking (finite) sums, and for which

$$\inf_{W \in \mathcal{O}} \inf_{0 \neq w \in W} \sup_{0 \neq v \in W} \frac{w(v)}{\|w\|_V \|v\|_V} > 0. \quad (5.7)$$

Consequently, the stability conditions (3.3)–(3.4) are satisfied for some $\gamma_\Delta > 0$ by simply taking in Proposition 5.2,

$$V^\mu_\delta := \hat{W}^\delta_\mu \in \mathcal{O}. \quad (5.8)$$

Proposition 5.3 shows that (5.7) is equivalent to uniform boundedness w.r.t. the $V$-norm of the $H$-orthogonal projector onto $W \in \mathcal{O}$. An example of a collection $\mathcal{O}$ for which the latter is true is given by the set of all finite element spaces $W^\delta_\lambda$ w.r.t. quasi-uniform, uniformly shape regular conforming partitions of $\Omega$ into, say, $d$-simplices.

It is known that uniform boundedness w.r.t. the $V$-norm of the $H$-orthogonal projector holds also true for finite element spaces w.r.t. locally refined partitions as long as the grading of the partitions is sufficiently mild. Recently, in [15], it was shown that in $d = 2$ spatial dimensions, and for any polynomial order, the collection of all conforming partitions that can be generated by newest vertex bisection (NVB), starting from a fixed conforming initial partition $T_\perp$ with an assignment of the newest vertices that satisfies a so-called matching condition, is sufficiently mildly graded in the above sense. Since the overlay of two conforming NVB partitions is a conforming NVB partition, this collection is closed under taking (finite) sums. In other words, with this collection of finite element spaces, which we will employ in our experiments, again the choice (5.8) guarantees uniform stability.

Also in [15], a similar result was shown in $d = 2$ for red-blue-green refinement with any polynomial order. Unfortunately, for $d > 2$ such results seem not yet to be available.

Remark 5.4 (Getting $\gamma_\Delta$ close to 1) We discussed uniform boundedness w.r.t. the $V$-norm of the $H$-orthogonal projectors onto a family of finite element spaces, which, by taking $V^\mu_\delta := \hat{W}^\delta_\mu$ in Proposition 5.2, yields the uniform inf-sup condition (3.4) for some value $\gamma_\Delta > 0$, and so uniform stability of the Galerkin discretizations of the saddle-point (2.6).

For proving convergence of our adaptive routine Algorithm 4.8, however, we needed a value of $\gamma_\Delta > 0$ that is sufficiently close to 1. Although in our numerical experiments, reported on in Section 7, with continuous piecewise linear finite element spaces generated by conforming NVB and $V^\delta_\mu = \hat{W}^\delta_\mu$, the adaptive routine is $r$-linearly converging, there is no guarantee that $1 - \gamma_\Delta$ is sufficiently small.
Restricting to quasi-uniform partitions, it can be shown that $1 - \gamma_{\Delta}$ can be made arbitrarily small by taking the mesh underlying $V^\delta_{\mu}$ to be a sufficiently deep, but fixed refinement of the mesh underlying $W^\delta_{\mu}$. One may conjecture that the same result holds true for sufficiently mildly graded locally refined meshes.

### 5.5 Best possible rates

Although we have not proved it, we expect that the sequence of approximations generated by our adaptive Algorithm 4.8 is not only $r$-linearly converging, but, ignoring data oscillation, that it is a sequence of approximations from the family $(X^\delta)_{\delta \in \Delta}$ that converges with the best possible rate. In this subsection, under some (mild) smoothness conditions on the solution $u$, we show that for our selection of the $(X^\delta)_{\delta \in \Delta}$ this best possible rate equals the rate of best approximation to the solution of the corresponding stationary problem from the spatial finite element spaces w.r.t. the $V$-norm.

Consider a family of spaces $X^\delta = \sum_{\lambda \in \Sigma} \sigma_{\lambda} \otimes W^\delta_{\lambda}$ that satisfies (5.4), with the $W^\delta_{\lambda}$ selected from a collection of finite element spaces $\mathcal{O}$ that in any case contains all spaces that correspond to uniform refinements of some initial partition of $\Omega$. Let $\Sigma$ a collection of wavelets of order $d_\lambda$, and assume that the finite element spaces are of order $d_\lambda$. When for each $X^\delta$, the space $Y^\delta$ is selected as in Proposition 5.2, then the combination of (3.10) and the analysis from [33, Section 7.1] shows that if the exact solution $u$ of our parabolic problem satisfies the mixed regularity condition $u \in H^{d_t}(I) \otimes H^{d_k}(\Omega)$, then a suitable (non-adaptive) choice of the spaces $W^\delta_{\lambda}$ yields a sequence of solutions $\hat{u}^\delta \in X^\delta$ (for arbitrary $Y^\delta \supset Y^\delta$) of the modified discretized saddle-point from Section 3.3, for which

$$\|u - \hat{u}^\delta\|_X \lesssim (\dim X^\delta - \min(d_t - 1, d_k - 1)).$$

Note that for $d_t - 1 \geq \frac{d_k - 1}{d}$, the rate $\frac{d_k - 1}{d}$ equals the best rate in the $V$-norm when the finite element spaces are employed for solving the corresponding stationary problem, which is posed on $\Omega \subset \mathbb{R}^d$ instead of on $I \times \Omega \subset \mathbb{R}^{d+1}$.

For an optimal adaptive choice of the finite element spaces $W^\delta_{\lambda}$ from a sufficiently ‘rich’ collection that contains locally refined partitions, as the collection of all conforming NVB partitions, it can be expected that the rate $\min(d_t - 1, \frac{d_k - 1}{d})$ is realized under much milder regularity conditions on $u$.

### 5.6 Preconditioners

Our adaptive algorithm for the parabolic problem requires optimal preconditioners for $E^\delta_{\lambda}AE^\delta_{\lambda'}$ and $S^\delta$, see (3.9) and (4.10). That is, for both $Z = Y$ and $Z = X$ and $\delta \in \Delta$, we need operators $K^\delta_{Z} = K^\delta_{Z} \in \mathcal{L}(Z^{\delta'}, Z^{\delta})$ with $h(K^\delta_{Z} h) \approx \|h\|_{Z^{\delta'}}^2$ ($h \in Z^{\delta'}$), which moreover should be applicable at linear cost.

To construct these preconditioners, for $Z \in \{Y, X\}$ we will select a symmetric, bounded, and coercive bilinear form on $Z \times Z$, and after selecting some basis for $Z^\delta$, we will construct a matrix $K^\delta_{Z} = K^\delta_{Z}^T$ that can be applied in linear complexity,
and that is uniformly spectrally equivalent to the inverse of the stiffness matrix corresponding to this bilinear form (being the matrix representation of the linear mapping $Z^\delta \to Z^{\delta'}$ defined by the bilinear form w.r.t. the basis for $Z^\delta$ and the corresponding dual basis for $Z^{\delta'}$). Then $K^\delta_Z \in \mathcal{L}(Z^{\delta'}, Z^\delta)$, defined as the operator whose matrix representation is $K^\delta_Z$ w.r.t. the aforementioned bases of $Z^{\delta'}$ and $Z^\delta$, is the preconditioner that satisfies our needs.

Notice that the choice of the basis for $Z^\delta$ is irrelevant. Indeed, denoting the aforementioned stiffness matrix as $C^\delta_Z$ with corresponding operator $C^\delta_Z = C^\delta_{Z'} \in \mathcal{L}(Z^\delta, Z^{\delta})$, one may verify that

$$\|K^\delta_Z\|_{\mathcal{L}(Z^{\delta'}, Z^\delta)} \| (K^\delta_Z)^{-1} \|_{\mathcal{L}(Z^\delta, Z^{\delta'})} \approx \|K^\delta_Z C^\delta_Z\|_{\mathcal{L}(Z^\delta, Z^{\delta})} \| (K^\delta_Z C^\delta_Z)^{-1} \|_{\mathcal{L}(Z^\delta, Z^{\delta})} = \frac{\lambda_{\text{max}}(K^\delta_Z C^\delta_Z)}{\lambda_{\text{min}}(K^\delta_Z C^\delta_Z)}.$$

### 5.6.1 Preconditioner at the ‘test side’

Let $Z = Y$. Since $\Psi$ is an orthonormal basis for $L_2(I)$, any $y \in Y$ equals $\sum_{\mu \in \bigvee \Psi} \psi_{\mu} \otimes v_{\mu}$ for some $v_{\mu} \in V$ with $\sum_{\mu \in \bigvee \Psi} \|v_{\mu}\|_V^2 < \infty$. Taking as bilinear form on $Y \times Y$ simply the scalar product on $Y \times Y$, we have

$$\left( \sum_{\mu_1 \in \bigvee \Psi} \psi_{\mu_1} \otimes v^{(1)}_{\mu_1} \right) \cdot \left( \sum_{\mu_2 \in \bigvee \Psi} \psi_{\mu_2} \otimes v^{(2)}_{\mu_2} \right) = \sum_{\mu \in \bigvee \Psi} \langle v^{(1)}_{\mu}, v^{(2)}_{\mu} \rangle_V.$$

Equipping $Y^\delta = \sum_{\mu \in \bigvee \Psi} \psi_{\mu} \otimes V_{\mu}^\delta$ with a basis of type $\bigcup_{\mu \in \bigvee \Psi} \psi_{\mu} \otimes \Phi_{\mu}^\delta$, the resulting stiffness matrix reads as blockdiag$[A_{\mu}^\delta]_{\mu \in \bigvee \Psi}$, where $A_{\mu}^\delta = \langle \Phi_{\mu}, \Phi_{\mu} \rangle_V$ is the stiffness matrix of $\langle \cdot, \cdot \rangle_V$ w.r.t. $\Phi_{\mu}$. Selecting $K_{\mu}^\delta \approx (A_{\mu}^\delta)^{-1}$, the matrix representation of the optimal preconditioner reads as

$$K_{Y}^\delta = \text{blockdiag}[K_{\mu}^\delta]_{\mu \in \bigvee \Psi}.$$

It is well-known that when $V_{\mu}^\delta$ is a finite element space, possibly w.r.t. a locally refined partition, suitable $K_{\mu}^\delta$ of multi-grid type are available. These $K_{\mu}^\delta$ can be applied in linear complexity, and so can $K_{Y}^\delta$.

**Remark 5.5** To show, in Theorem 4.9, $r$-linear convergence of our adaptive Algorithm 4.8, we require the constant $\kappa_{\Delta}$ from (3.9) to be sufficiently small, i.e., the eigenvalues of the $K_{Y}^\delta E_{Y}^{1/\delta} A E_{Y}^{1/\delta}$ to be sufficiently close to 1. Given an initial optimal, self-adjoint, coercive preconditioner $K_{Y}^{\delta}$, and some upper and lower bounds on the spectrum of the preconditioned system, one can satisfy the latter condition by polynomial acceleration using Chebychev polynomials of sufficiently high degree. In our numerical experiments, this ‘acceleration’ turned out to be unnecessary.
5.6.2 Preconditioner at the ‘trial side’

Let $Z = X$. The preconditioner presented in this section is inspired by constructions of preconditioners in [3, 29] for parabolic problems discretized on a tensor product of temporal and spatial spaces.

As $\Sigma$ and \( \{ 2^{-|\lambda|} \sigma_\lambda : \lambda \in \nabla \Sigma \} \) are Riesz bases for $L^2(I)$ and $H^1(I)$, any $x \in X$ equals $\sum_{\lambda \in \nabla \Sigma} \sigma_\lambda \otimes w_\lambda$ for some $w_\lambda \in V$ with $\sum_{\lambda \in \nabla \Sigma} \| w_\lambda \|_V^2 + 4|\lambda| \| w_\lambda \|_V^2 < \infty$, and

\[
\{ \sum_{\lambda \in \nabla \Sigma} \sigma_{\lambda_1} \otimes w^{(1)}_{\lambda_1}, \sum_{\lambda_2 \in \nabla \Sigma} \sigma_{\lambda_2} \otimes w^{(2)}_{\lambda_2} \} := \sum_{\lambda \in \nabla \Sigma} \langle w^{(1)}_{\lambda}, w^{(2)}_{\lambda} \rangle_V + 4|\lambda| \langle w^{(1)}_{\lambda}, w^{(2)}_{\lambda} \rangle_V,
\]

is a symmetric, bounded, coercive bilinear form on $X \times X$. Equipping $X^\delta = \sum_{\lambda \in \nabla \Sigma} \sigma_\lambda \otimes W_\lambda^\delta$ with a basis of type $\bigcup_{\lambda \in \nabla \Sigma} \sigma_\lambda \otimes \Phi_\lambda^\delta$, the resulting stiffness matrix reads

\[
\text{blockdiag[} A_\lambda^\delta + 4|\lambda| \langle \Phi_\lambda^\delta, \Phi_\lambda^\delta \rangle_V \big|_{\lambda \in \nabla \Sigma} \text{]},
\]

where $A_\lambda^\delta = \langle \Phi_\lambda^\delta, \Phi_\lambda^\delta \rangle_V$.

Thanks to our assumption (5.7), for $u \in W_\lambda^\delta$ it holds that $\| u \|_V \approx \sup_{0 \neq w \in W_\lambda^\delta} \langle u, w \rangle / \| w \|_V$. With $u$ denoting the representation of $u$ w.r.t. $\Phi_\lambda^\delta$, we have

\[
\sup_{0 \neq w \in W_\lambda^\delta} \frac{\langle u, w \rangle}{\| w \|_V} = \| (A_\lambda^\delta)^{-\frac{1}{2}} M_\lambda^\delta u \|,
\]

where $M_\lambda^\delta = \langle \Phi_\lambda^\delta, \Phi_\lambda^\delta \rangle$, so that

\[
\langle \Phi_\lambda^\delta, \Phi_\lambda^\delta \rangle_V \approx M_\lambda^\delta (A_\lambda^\delta)^{-1} M_\lambda^\delta.
\]

Since both $A_\lambda^\delta$ and $M_\lambda^\delta$ are symmetric positive definite, [31, Thm. 4] shows

\[
\frac{1}{2} (A_\lambda^\delta + 4|\lambda| M_\lambda^\delta (A_\lambda^\delta)^{-1} M_\lambda^\delta) \leq (A_\lambda^\delta + 2|\lambda| M_\lambda^\delta) (A_\lambda^\delta)^{-1} (A_\lambda^\delta + 2|\lambda| M_\lambda^\delta) \leq A_\lambda^\delta + 4|\lambda| M_\lambda^\delta (A_\lambda^\delta)^{-1} M_\lambda^\delta.
\]

Now assuming that

\[
K_\lambda^\delta \approx (A_\lambda^\delta + 2|\lambda| M_\lambda^\delta)^{-1},
\]

we infer that

\[
K_\lambda^\delta = \text{blockdiag} \left[ K_\lambda^\delta A_\lambda^\delta K_\lambda^\delta \right]_{\lambda \in \nabla \Sigma}
\]

is the matrix representation of an optimal preconditioner.

Remark 5.6 Notice that (5.9) requires an optimal preconditioner of a discretized reaction-diffusion equation that is robust w.r.t. the size of the (constant) reaction term. In [30] it was shown that, under a ‘full-regularity’ assumption, for quasi-uniform meshes multiplicative multi-grid yields such a preconditioner, whose application can even be performed at linear cost. Although we expect that this assumption can be avoided using the theory of subspace correction methods, and furthermore that the optimality, robustness and linear complexity result extends to locally refined meshes, proofs of such extensions seem not to be available.
6 A concrete realization

6.1 The collection $\mathcal{O}$ of finite element spaces, and the mapping $\delta \rightarrow \hat{\delta}$

We further specify the collection $\mathcal{O}$ of finite element spaces, construct a linearly independent set in $H^1_{0,\Gamma_D}(\Omega)$ known as the hierarchical basis, and equip it with a tree structure such that there exists a 1-1 correspondence between the finite element spaces in $\mathcal{O}$, and the spans of subsets of the hierarchical basis that form trees.

With this specification of $\mathcal{O}$, there will be a 1-1 correspondence between the spaces $X^\delta = \sum_{\lambda \in \sigma_\lambda} \sigma_\lambda \otimes W^\delta_\lambda$ with $W^\delta_\lambda \in \mathcal{O}$ that satisfy (5.4), and the spans of collections of tensor products of wavelets $\sigma_\lambda$ and hierarchical basis functions whose sets of index pairs are lower, also known as downward closed. Given such a $X^\delta$, we will define $X^\hat{\delta}$ by a certain enlargement the lower set.

For $d \geq 2$, let $\mathbb{T}$ be the family of all conforming partitions of a polytope $\Omega \subset \mathbb{R}^d$ into (closed) $d$-simplices that can be created by NVB starting from some given conforming initial partition $\mathcal{T}_\perp$ with an assignment of the newest vertices that satisfies the matching condition, see [35]. We define a partial order on $\mathbb{T}$ by writing $\mathcal{T} \preceq \hat{\mathcal{T}}$ when $\hat{\mathcal{T}}$ is a refinement of $\mathcal{T}$.

With some small adaptations that we leave to the reader, in the following the case $d = 1$ can be included by letting $\mathbb{T}$ be the family of partitions of $\Omega$ into (closed) subintervals constructed by bisections from $\mathcal{T}_\perp =$ \{\}$ such that the generations of any two neighbouring subintervals in any $\mathcal{T} \in \mathbb{T}$ differ by not more than one.

The collection $\mathcal{O}$ that we will consider is formed by the spaces $W = W_T$ of continuous piecewise linears w.r.t. $T \in \mathbb{T}$, zero on a possible Dirichlet boundary $\Gamma_D$ being the union of $\partial T \cap \partial \Omega$ for some $T \in \mathcal{T}_\perp$. We expect that generalizations to finite element spaces of higher order do not impose essential difficulties.

For $T \in \mathcal{T} := \cup_{\mathcal{T} \in \mathbb{T}}\{T: T \in \mathcal{T}\}$, we set $\text{gen}(T)$ to be the number of bisections needed to create $T$ from its ‘ancestor’ $T' \in \mathcal{T}_\perp$. With $\mathcal{N}$ being the set of all vertices (or nodes) of all $T \in \mathcal{T}$, for $v \in \mathcal{N}$ we set $\text{gen}(v) := \min\{\text{gen}(T): T \in \mathcal{T}, v \in T\}$.

Any $v \in \mathcal{N}$ with $\text{gen}(v) > 0$ is the midpoint of an edge of one or more $T \in \mathcal{T}$ with $\text{gen}(T) = \text{gen}(v) - 1$. The vertices $\check{v}$ of these $T$ with $\text{gen}(\check{v}) = \text{gen}(v) - 1$ are defined as the parents of $v$. We denote this relation between a parent $\check{v}$ and a child $v$ by $\check{v} \preceq_{\mathcal{N}} v$, see Fig. 1.

Vertices $v \in \mathcal{N}$ with $\text{gen}(v) = 0$ have no parents.

An (essentially) non-overlapping partition $\mathcal{T}$ of $\overline{\Omega}$ into $T \in \mathcal{T}$ is in $\mathbb{T}$ if and only if the set $N_\mathcal{T}$ of vertices of all $T \in \mathcal{T}$ forms a tree, meaning that it contains all $v \in \mathcal{N}$.

![Fig. 1 $\check{v}_1, \check{v}_2 \preceq_{\mathcal{N}} v$ (left picture); and $\mathcal{T}$ and its refinement $\mathcal{T}^{d+}$ (for $d = 2$) (right picture)]
with $\text{gen}(v) = 0$ as well as all parents of any $v \in N_T$ with $\text{gen}(v) > 0$, cf. [12] for the $d = 2$ case.

**Definition 6.1** For any $T \in \mathcal{T}$, we define $\mathcal{T}^{d+} \in \mathcal{T}$ (denoted as $\mathcal{T}^{++}$ in [12] for the $d = 2$ case) by replacing any $T \in \mathcal{T}$ by its $2^d$ ‘descendants’ of the $d$th generation, see Fig. 1.

Since this refinement adds exactly one vertex at the midpoint on any edge of all $T \in \mathcal{T}$, one infers that indeed $T^{d+} \in \mathcal{T}$. The corresponding tree $N_T^{d+}$ is created from $N_T$ by the addition of all descendants up to generation $d$ of all $v \in N_T$.

For any $v \in \mathcal{N}$, we set $\phi_v$ as the continuous piecewise linear function w.r.t. the uniform partition \( \{T \in \mathcal{T} : \text{gen}(T) = \text{gen}(v)\} \in \mathcal{T} \), which function is 1 at $v$ and 0 at all other vertices of this partition. Setting $\mathcal{N}_0 := \mathcal{N} \setminus \Gamma_D$ and, for any $\mathcal{T} \in \mathcal{T}$, $N_{\mathcal{T}, 0} := N_{\mathcal{T}} \setminus \Gamma_D$, the collection \( \{\phi_v : v \in \mathcal{N}_0\} \) is known as the hierarchical basis, and for any $\mathcal{T} \in \mathcal{T}$, it holds that $W_{\mathcal{T}} = \text{span}(\phi_v : v \in N_{\mathcal{T}, 0})$.

With above specification of the collection $\mathcal{O}$ of finite element spaces, there exists a 1-1 correspondence between the spaces $\sum_{\lambda \in \mathcal{V}} \sigma_\lambda \otimes W_\lambda^\delta$ with $W_\lambda^\delta \in \mathcal{O}$ that satisfy (5.4), and the spaces of the form

$$X^\delta = \text{span}\{\sigma_\lambda \otimes \phi_v : (\lambda, v) \in I_{\delta, 0} := I_\delta \setminus (\mathcal{V} \times \Gamma_D)\}$$

(6.1)

for some finite $I_\delta \subset \mathcal{V} \times \mathcal{N}$ being a lower set in the sense

$$\begin{align*}
(\lambda, v) & \in I_\delta \quad \text{and} \quad \begin{cases}
\tilde{\lambda} \ll \lambda \\
\tilde{v} \ll v \quad \text{or} \quad \text{gen}(\tilde{v}) = 0
\end{cases} \implies (\tilde{\lambda}, v) \in I_\delta, \\
(\lambda, \tilde{v}) \in I_\delta.
\end{align*}$$

(6.2)

For above specification of $X^\delta$, from Proposition 5.2 with the specification (5.8) one infers that the corresponding space $Y^\delta$ is given by

$$Y^\delta = \text{span}\{\psi_\mu \otimes \phi_v : (\mu, v) \in I_{\delta, 0}^Y\},$$

(6.3)

where

$$I_{\delta, 0}^Y := \{(\mu, v) : \exists (\lambda, \nu) \in I_{\delta, 0}, \mu \in \mathcal{V}, |\mu| = |\lambda|, |\mathcal{S}_\psi(\mu) \cap \mathcal{S}_\lambda(\lambda)| > 0\}$$

(6.4)

which index set is a lower set.

**Remark 6.2** (Complexity of matrix-vector multiplications) The fact that the index sets of the bases for $X^\delta$ and $Y^\delta$ are lower sets is the key to computing residuals of the system $S^\delta u^\delta = f^\delta$ (4.8) in $\mathcal{O}(\dim X^\delta)$ operations. The algorithm that realizes this complexity makes a clever use of multi- to single-scale transformations alternately in time and space. In a ‘uniform’ sparse-grid setting, i.e., without ‘local refinements’, this algorithm was introduced in [7], and it was later extended to general lower sets in [24]. The definition of a lower set in [24], there called multi-tree, is more restrictive than our current definition that allows more localized refinements, and details about the matrix-vector multiplication and a proof of its optimal computational complexity can be found in [42].

---

The addition of only all children of all $v \in N_T$ yields a tree only if $T$ is a uniform partition.
Definition 6.3 Given $X^\delta = \text{span}\{\sigma_\lambda \otimes \hat{\phi}_\nu : (\lambda, \nu) \in I_{\delta,0}\}$ for some lower set $I_{\delta} \subset \vee_{\Sigma} \times \mathfrak{N}$, we define the lower set $I_{\delta}^\lambda$, and with that $X^\delta$, by adding, for each $(\lambda, \nu) \in I_{\delta}$ and any child $\check{\lambda}$ of $\lambda$ and any descendant $\check{\nu}$ of $\nu$ up to generation $d$, all pairs $(\check{\lambda}, \nu)$ and $(\lambda, \check{\nu})$ to $I_{\delta}$.

6.2 The collection $\Theta_{\delta}$ such that $X^\delta = X^\delta \oplus \Theta_{\delta}$

Recall that for the bulk chasing process we need an ‘$X$-stable’ basis $\Theta_{\delta}$ that spans an ‘$X$-stable’ complement space of $X_{\delta}$ in $X^\delta$, i.e., a collection that satisfies (4.2). For that goal we define a modified hierarchical basis $\{\hat{\phi}_\nu : \nu \in \mathfrak{N}_0\}$ by $\hat{\phi}_\nu := \phi_\nu$ when $\text{gen}(\nu) = 0$, and $\hat{\phi}_\nu := \phi_\nu - \sum_{\tilde{\nu} \in \mathfrak{N}_0 \setminus \nu} \frac{\int_\Omega \phi_\nu dx \phi_{\check{\nu}}}{\# \{\tilde{\nu} \in \mathfrak{N} : \tilde{\nu} \prec_\Sigma \nu\}} \phi_{\check{\nu}}$ otherwise. Notice that for those $\nu$ with $\text{gen}(\nu) > 0$ that have all their parents not on $\Gamma_D$ it holds that $\int_\Omega \hat{\phi}_\nu dx = 0$, i.e., $\hat{\phi}_\nu$ has a vanishing moment, and furthermore that for any $T \in \mathcal{T}$, $W_T = \text{span}\{\phi_\nu : \nu \in \mathfrak{N}_T, 0\} = \text{span}\{\hat{\phi}_\nu : \nu \in \mathfrak{N}_T, 0\}$, and thus for any lower set $I_{\delta} \subset \vee_{\Sigma} \times \mathfrak{N}$, $X^\delta = \text{span}\{\sigma_\lambda \otimes \hat{\phi}_\nu : (\lambda, \nu) \in I_{\delta,0}\} = \text{span}\{\sigma_\lambda \otimes \phi_\nu : (\lambda, \nu) \in I_{\delta,0}\}$.

Moreover, for any $T \in \mathcal{T}$, the basis transformation from the modified to unmodified hierarchical basis for $W_T$ can be applied in linear complexity traversing from the leaves to the roots.

Given $\delta$, the collection $\Theta_{\delta}$ will be the set of properly normalized functions $\sigma_\lambda \otimes \hat{\phi}_\nu$ for $(\lambda, \nu) \in I_{\delta,0} \setminus I_{\delta,0}$. In order to demonstrate (4.2), we have to impose some gradedness assumption on the lower sets $I_{\delta}$.

Definition 6.4 The gradedness constant of a lower set $I_{\delta} \subset \vee_{\Sigma} \times \mathfrak{N}$ is the smallest $L_{\delta} \in \mathbb{N}$ such that for all $(\lambda, \nu) \in I_{\delta}$ for which $\nu$ has an ancestor $\tilde{\nu} \in \mathfrak{N}$ with $\text{gen}(\nu) - \text{gen}(\tilde{\nu}) = L_{\delta}$, it holds that $(\check{\lambda}, \tilde{\nu}) \in I_{\delta}$ for any child $\check{\lambda} \in \vee_{\Sigma}$ of $\lambda$.

Remark 6.5 (Uniform boundedness of the gradedness constants) Under the (unproven) assumption that our adaptive method creates a sequence of spaces $X^\delta$ which are quasi-optimal for the approximation of the solution of the parabolic PDE, one may hope that these spaces have a uniformly bounded gradedness constant, unless (locally) the solution $u$ is extremely more smooth as function of $t$ than as function of the spatial variables.

To see this, for some constant $L \in \mathbb{N}$, consider the non-adaptive sparse grid index sets of the form $\{ (\lambda, \nu) \in \vee_{\Sigma} \times \mathfrak{N} : L|\lambda| + \text{gen}(\nu) \leq N \}$, which are appropriate when the behaviour of $u$ as function of $t$ and of the spatial variables is globally similar. Their gradedness constant is $L$. In the corresponding trial spaces, the smallest spatial resolution equals the smallest temporal resolution to the power $L/d$. So only when such a polynomial decay of the spatial resolution as function of the temporal
resolution does not suffice for a proper approximation of \( u \), a uniformly bounded gradedness constant cannot be expected.

**Proposition 6.6** For \((\lambda, \nu) \in \sqrt{\Sigma} \times \mathfrak{N}_0\), let \( e_{\lambda \mu} := 1 / \sqrt{2(3^2 - 1)^{\text{gen}(\nu)}} + 4|\lambda|2(-\frac{1}{2} - 1)^{\text{gen}(\nu)} \)
and \( \theta_{\lambda \nu} := e_{\lambda \mu} \sigma_{\lambda} \otimes \hat{\phi}_\nu \). For any \( \delta \in \Delta \), let \( \Theta_\delta := \{ \theta_{\lambda \nu} : (\lambda, \nu) \in J_\delta := I_{\delta,0} \setminus I_{\delta,0} \} \).

Then, \( X^\delta \oplus \text{span} \Theta_\delta = X^\delta \), and there exist constants \( 0 < m_\delta \leq M_\delta \), only dependent on the gradedness constant \( L_\delta \), such that for any \( z \in X^\delta \) and \( c = (c_{\lambda \nu})(\lambda, \nu) \in I_{\delta,0} \setminus I_{\delta,0} \subset \mathbb{R} \),

\[
m_\delta(\|z\|^2_X + \|c\|^2) \leq \|z + c^T \Theta_\delta\|^2_X \leq M_\delta(\|z\|^2_X + \|c\|^2).
\]

So under the mild assumption that the gradedness constants of the sets \( X^\delta \) that we encounter are uniformly bounded, condition (4.2) is satisfied.

**Proof** Setting \( c_{\lambda \nu} := 0 \) when \((\lambda, \nu) \notin I_{\delta,0} \setminus I_{\delta,0} \), and writing \( z = \sum_{\lambda \in \sqrt{\Sigma}} \sigma_{\lambda} \otimes w_\lambda \)
where \( w_\lambda \in \text{span}[\hat{\phi}_\nu : (\lambda, \nu) \in I_{\delta,0}] \), from \( \Sigma \) and \( \{ 2^{-|\lambda|} \sigma_{\lambda} : \lambda \in \sqrt{\Sigma} \} \) being Riesz bases for \( L_2(I) \) and \( H^1(I) \), and \( c^T \Theta_\delta = \sum_{\lambda} \sigma_{\lambda} \otimes \sum_{\nu} e_{\lambda \mu} c_{\lambda \nu} \hat{\phi}_\nu \), an application of Lemma 6.7 given below shows that

\[
\|z + c^T \Theta_\delta\|^2_X \approx \sum_{\lambda} \|w_\lambda\|^2 + \sum_{\nu} e_{\lambda \mu} c_{\lambda \nu} \hat{\phi}_\nu ||H^1_0(\Omega) + 4|\lambda|\|w_\lambda\|^2 + \sum_{\nu} e_{\lambda \mu} c_{\lambda \nu} \hat{\phi}_\nu ||H^1_0(\Omega)\|^2 \geq \sum_{\lambda} \|w_\lambda\|^2 + \sum_{\nu} \|e_{\lambda \mu} c_{\lambda \nu} \hat{\phi}_\nu \|^2,
\]

with the \( \approx \)-symbol in the second line dependent on the gradedness constant. \( \square \)

The proof of Proposition 6.6 was based on the following lemma.

**Lemma 6.7** For \( \tilde{T} \in \mathfrak{T} \), and either \( \mathfrak{T} \ni \mathcal{T} \leq \tilde{T} \) and \( v \in W_{\mathcal{T}} \), or \( \mathcal{T} = \emptyset \), \( N_{\mathcal{T},0} := \emptyset \), and \( v = 0 \), and scalars \( (d_\nu)_{\nu \in N_{\tilde{T},0} \setminus N_{\mathcal{T},0}} \), it holds that

\[
\|v + \sum_{\nu} d_\nu \hat{\phi}_\nu \|^2_{H^1(\Omega)} \approx \|v\|^2_{H^1(\Omega)} + \sum_{\nu} 2(3^2 - 1)^{\text{gen}(\nu)}|d_\nu|^2 \quad (6.5)
\]

\[
\|v + \sum_{\nu} d_\nu \hat{\phi}_\nu \|^2_{H^1_0(\Omega)} \approx \|v\|^2_{H^1_0(\Omega)} + \sum_{\nu} 2(-\frac{1}{2} - 1)^{\text{gen}(\nu)}|d_\nu|^2 \quad (6.6)
\]

with the constants hidden in the \( \approx \)-symbols only dependent on \( M_{\tilde{T},\mathcal{T}} := \max\{ \text{gen}(\tilde{T}) - \text{gen}(\mathcal{T}) : \tilde{T} \supset \mathcal{T} \in \mathfrak{T} \} \) or \( M_{\mathcal{T},\tilde{T}} := \max\{ \text{gen}(\mathcal{T}) : \tilde{T} \in \mathfrak{T} \} \) for \( \mathcal{T} = \emptyset \).

**Proof** Once the equivalences are shown uniformly in any \( \mathcal{T} \leq \tilde{T} \) with \( M_{\tilde{T},\mathcal{T}} = 1 \), repeated application shows them for the general case, with constants only dependent on \( M_{\tilde{T},\mathcal{T}} \). So in the following, it suffices to consider the case that \( M_{\tilde{T},\mathcal{T}} = 1 \). The case \( \mathcal{T} = \emptyset \) is easy, so we will consider the case that \( \mathcal{T} \in \mathfrak{T} \).
Let $\Phi_T = \{\phi_{T,v} : v \in N_{\hat{T},0}\}$ denote the standard nodal basis for $W_{\hat{T}}$. For any weight function $0 < w_{\hat{T}} \in T \in T P_0(T)$, with $\| \cdot \|_{L_{2,w_{\hat{T}}}^x(\Omega)} := \| w_{\hat{T}}^{\frac{1}{2}} \cdot \|_{L_2(\Omega)}$ it holds that $\| \sum_v c_v \phi_{T,v} \|^2_{L_{2,w_{\hat{T}}}^x(\Omega)} \simeq \sum_v |c_v|^2 \| \phi_{T,v} \|^2_{L_{2,w_{\hat{T}}}^x(\Omega)}$ only dependent on the spectrum of the element mass matrix on a reference element, i.e., on the space dimension $d$, so independent of the weight function $w_{\hat{T}}$. We refer to this equivalence by saying that $\Phi_T$ is (uniformly) stable w.r.t. $\| \cdot \|_{L_{2,w_{\hat{T}}}^x(\Omega)}$.

Notice that for $v \in N_{\hat{T},0} \setminus N_{\hat{T},0}$, it holds that $\phi_{\hat{T},v} = \phi_v$. W.r.t. the splitting $N_{\hat{T},0} = N_{\hat{T},0} \cup N_{\hat{T},0}$, the basis transformation from $\Phi_T \cup \{\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ to $\Phi_T \cup \{\phi_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ is of the form $\begin{bmatrix} \text{Id} & \ast \\ 0 & \text{Id} \end{bmatrix}$, and the basis transformation from the latter basis to $\Phi_T$ is of the form $\begin{bmatrix} \text{Id} & 0 \\ \ast & \text{Id} \end{bmatrix}$. The entries in both non-zero off-diagonal blocks are uniformly bounded, where non-zeros can only occur for index pairs $(v, \hat{v})$ that are vertices of the same $\hat{T} \in \hat{T}$. Consequently, for a family of weight functions $(w_{\hat{T}})_{\hat{T} \in \hat{T}}$ that have uniformly bounded jumps in that

$$\sup_{\hat{T} \in \hat{T}} \sup_{T \in \hat{T} \cup T} \frac{w_{\hat{T}} |T|}{w_{\hat{T}} |T'|} < \infty, \quad (6.7)$$

all basis transformations between the $L_{2,w_{\hat{T}}}^x(\Omega)$-normalized bases $\Phi_T \cup \{\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$, $\Phi_T \cup \{\phi_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ and $\Phi_T$ are uniformly bounded.

Since, as we have seen, $\Phi_T$ is (uniformly) stable w.r.t. $\| \cdot \|_{L_{2,w_{\hat{T}}}^x(\Omega)}$, we conclude that also $\Phi_T \cup \{\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ and $\Phi_T \cup \{\phi_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ are (uniformly) stable w.r.t. $\| \cdot \|_{L_{2,w_{\hat{T}}}^x(\Omega)}$. Because of the uniform $K$-mesh property of $\hat{T} \in \hat{T}$, examples of families of weights that satisfy (6.7) are given by $(h_{\hat{T}}^x)_{\hat{T} \in \hat{T}}$ for any $s \in \mathbb{R}$, where $h_{\hat{T}} |T| := 2^{-\text{gen}(T)/d} (\sim |T|^{1/d}) (T \in \hat{T})$.

To show (6.5), let $P_T : W_{\hat{T}} \rightarrow W_T$ be the projector with ran $P_T = W_T$ and ran$(\text{Id} - P_T) = \text{span}[\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}]$. Using the form of the basis transform from $\Phi_T \cup \{\phi_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ to $\Phi_T \cup \{\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$, one infers

$$\text{Id} - P_T = J_T \circ (\text{Id} - I_T),$$

where $I_T$ is the nodal interpolator onto $W_T$, and $J_T$ is defined by $J_T \phi_v = \hat{\phi}_v$. Since both $\{\hat{\phi}_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ and $\{\phi_v : v \in N_{\hat{T},0} \setminus N_{\hat{T},0}\}$ are uniformly stable w.r.t. $\| h_{\hat{T}}^{-1} \cdot \|_{L_2(\Omega)}$, and $\| h_{\hat{T}}^{-1} \hat{\phi}_v \|_{L_2(\Omega)} \simeq \| h_{\hat{T}}^{-1} \phi_v \|_{L_2(\Omega)}$, it follows that $J_T$ is uniformly bounded w.r.t. $\| h_{\hat{T}}^{-1} \cdot \|_{L_2(\Omega)}$, i.e. $\| h_{\hat{T}}^{-1} J_T h_{\hat{T}} \|_{L(L_2(\Omega),L_2(\Omega))} \lesssim 1$, and so

$$\| h_{\hat{T}}^{-1} (\text{Id} - P_T) v \|_{L_2(\Omega)} \lesssim \| h_{\hat{T}}^{-1} (\text{Id} - I_T) v \|_{L_2(\Omega)} \lesssim |v|_{H^1(\Omega)} \quad (v \in W_{\hat{T}}).$$

With the common inverse inequality $\| \cdot \|_{H^1(\Omega)} \lesssim \| h_{\hat{T}}^{-1} \cdot \|_{L_2(\Omega)}$ on $W_{\hat{T}}$, we see that $\text{Id} - P_T$ is uniformly bounded in $H^1(\Omega)$-norm, and that on ran$(\text{Id} - P_T)$, $\| \cdot \|_{H^1(\Omega)} \simeq$
\[ h^{-1}_T \cdot \|L_2(\Omega) \cdot \) The proof of (6.5) is completed by the uniform stability of \( \{\hat{\varphi}_v : v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0} \} \) w.r.t. \( h^{-1}_T \cdot \|L_2(\Omega) \), using that \( h^{-1}_T \hat{\varphi}_v \|L_2(\Omega) \approx 2^{(-\frac{d}{2}-1)\text{gen}(v)} \).

Moving to (6.6), either by \( \int_{\Omega} \hat{\varphi}_v dx = 0 \), or otherwise using the proximity of the Dirichlet boundary \( \Gamma_D \) by an application of Poincaré’s inequality, it holds that

\[
\| \langle \hat{\varphi}_v, v \rangle_{L_2(\Omega)} \| \lesssim 2^{-\text{gen}(v)/d} \| \hat{\varphi}_v \|_{L_2(\Omega)} \| v \|_{H^1(\text{supp } \hat{\varphi}_v)} \quad (v \in \mathcal{H}_0 \setminus N_{\mathcal{T},0}).
\]

By using that for \( T \in \hat{T} \) the number of \( v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0} \) for which \( \text{supp } \hat{\varphi}_v \) has non-empty intersection with \( T \) is uniformly bounded, and furthermore that \( \Phi_T \cup \{\hat{\varphi}_v : v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0} \} \) is uniformly stable w.r.t. \( h_T \cdot \|L_2(\Omega) \), we infer that for any \( z = \sum_{v \in N_{\mathcal{T},0}} z_v \hat{\varphi}_{T,v} \in W_T \) it holds that

\[
\| \sum_{v \in N_{\mathcal{T},0}} d_v \hat{\varphi}_v \|_{H^1_{0,\Gamma_D}(\Omega)'} = \sup_{0 \neq v \in H^1_{0,\Gamma_D}(\Omega)} \frac{\| \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0} d_v \hat{\varphi}_v, v \rangle_{L_2(\Omega)}}{\| v \|_{H^1(\Omega)}}
\]

\[
\lesssim \sqrt{\sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0}} |d_v|^2 \| h_T \hat{\varphi}_v \|_{L_2(\Omega)}^2} + \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0}} |z_v|^2 \| h_T \hat{\varphi}_{T,v} \|_{L_2(\Omega)}^2 \approx \| h_T (z + \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0}} d_v \hat{\varphi}_v) \|_{L_2(\Omega)} \lesssim \| z + \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0}} d_v \hat{\varphi}_v \|_{H^1_{0,\Gamma_D}(\Omega)'} \quad (v \in \mathcal{H}_0 \setminus N_{\mathcal{T},0}).
\]

the last inequality by application of a less common inverse inequality which proof can be found in [37, Lemma 3.4] for general dimensions \( d \). From (6.8) it follows that \( \text{Id} - P_T \) is uniformly bounded in the \( H^1_{0,\Gamma_D}(\Omega) \)-norm, and also that \( \| \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0} d_v \hat{\varphi}_v \|_{H^1_{0,\Gamma_D}(\Omega)'} \approx \sum_{v \in N_{\mathcal{T},0} \setminus N_{\mathcal{T},0}} |d_v|^2 \| h_T \hat{\varphi}_v \|_{L_2(\Omega)}^2 \| \frac{1}{2} \| h_T - h_T \|_{L_2(\Omega)}^2 \approx 2^{(-\frac{d}{2}-1)\text{gen}(v)} \), where we used

The proof of (6.6) is completed. \( \square \)

6.3 The wavelet collections \( \Sigma \) and \( \Psi \)

As wavelet basis \( \Sigma = \{\sigma_{\lambda} : \lambda \in \mathring{\Sigma} \} \) we select the three-point hierarchical basis illustrated in Fig. 2. This basis is known to be a Riesz basis for \( L_2(I) \), and, after renormalization, for \( H^1(I) \) (see [34]). It also satisfies the other assumptions made in Section 5.2. The wavelets up to level \( \ell \) span the space of continuous piecewise linear functions on \( I \) w.r.t. the uniform partition into \( 2^\ell \) subintervals.

As wavelet basis \( \Psi = \{\psi_{\mu} : \mu \in \mathring{\Psi} \} \) we take the orthonormal (discontinuous) piecewise linear wavelets, see Fig. 3. The wavelets up to level \( \ell \) span the space of (discontinuous) piecewise linear functions on \( I \) w.r.t. the uniform partition into \( 2^\ell \) subintervals.
Three-point hierarchical basis $\Sigma$. On level 0 there are two wavelets, and on level 1 there is one wavelet, whose parents are both wavelets on level 0. On each level $\ell > 1$, there are $2^{\ell-1}$ wavelets, among them near each boundary one boundary-adapted wavelet, where each wavelet has one parent being the wavelet on level $\ell - 1$ whose support includes the support of its child (so $S_{\Sigma}(\lambda)$ can be taken equal to $\text{supp} \sigma_\lambda$). All but one wavelets have one (bdr. wav) or two vanishing moments.

6.4 The family $\left(\delta_G, \delta U_0\right)_{\delta \in \Delta}$

The index set $\bigvee \Sigma$ is naturally identified with the set of ‘nodal dyadic’ points, see Fig. 4, which is the natural index set for the one-dimensional hierarchical basis that we denote by $\{\phi_\lambda : \lambda \in \bigvee \Sigma\}$.

Recalling that for $\delta \in \Delta$, $X^\delta = \text{span}\{\sigma_\lambda \otimes \phi_\nu : (\lambda, \nu) \in I_\delta, 0 = I_\delta \setminus (\bigvee \Sigma \times \mathbb{N})\}$ for some lower set $I_\delta \subset \bigvee \Sigma \times \mathbb{N}$, we define

$$\delta G := \text{span}\{\phi_\lambda \otimes \phi_\nu : (\lambda, \nu) \in I_\delta\}, \ \delta U_0 := \text{span}\{\phi_\nu : (\lambda, \nu) \in I_\delta, \phi_\lambda(0) \neq 0\}.$$ 

Since the level of resolution of these spaces is comparable to that of $X^\delta$, based on our experiences with wavelet and finite element methods we expect that with this choice of $(\delta G, \delta U_0)$ and the definition of $X^\delta$, that saturation holds, i.e., that Assumption 3.1 assumption is valid.
Given $g \in Y'$ and $u_0 \in L_2(\Omega)$, it remains to define their approximations $(\delta g, \delta u_0) \in (\delta G, \delta U_0)$. In general, the construction of these approximations depends on the data at hand. Below we give a construction that applies to general continuous $g$ and $u_0$, and that avoids quadrature issues.

For $\nu \in \mathbb{N}$ with $\text{gen}(\nu) = 0$, let $\tilde{\phi}_\nu := \delta_\nu$. Each $\nu \in \mathbb{N}$ with $\text{gen}(\nu) > 0$ is the midpoint of an edge of a $T \in T$ with $\text{gen}(T) = \text{gen}(\nu) - 1$. Denoting the endpoints of this edge as $\nu_1, \nu_2 \in \mathbb{N}$, let $\tilde{\phi}_\nu := \frac{1}{2}(\delta_{\nu_1} + \delta_{\nu_2})$. Then $\{\tilde{\phi}_\nu : \nu \in \mathbb{N}\} \subset C(\bar{\Omega})$ is biorthogonal to $\{\phi_\nu : \nu \in \mathbb{N}\}$.

For $\nu \in \mathbb{N}$ with $\text{gen}(\nu) = 0$, let $\tilde{\phi}_\nu := \delta_\nu$. Each $\nu \in \mathbb{N}$ with $\text{gen}(\nu) > 0$ is the midpoint of an edge of a $T \in T$ with $\text{gen}(T) = \text{gen}(\nu) - 1$. Denoting the endpoints of this edge as $\nu_1, \nu_2 \in \mathbb{N}$, let $\tilde{\phi}_\nu := \frac{1}{2}(\delta_{\nu_1} + \delta_{\nu_2})$. Then $\{\tilde{\phi}_\nu : \nu \in \mathbb{N}\} \subset C(\bar{\Omega})$ is biorthogonal to $\{\phi_\nu : \nu \in \mathbb{N}\}$.

Since we expect that for sufficiently smooth $g$ and $u_0$, the errors $\|g - \delta g\|_{Y'}$ and $\|u_0 - \delta u_0\|_{L_2(\Omega)}$ are of higher order than the approximation error $\inf_{w \in X^\delta} \|u - w\|_X$, for our convenience in the adaptive Algorithm 4.8 we ignore errors caused by data-oscillation by setting $\eta(\cdot) \equiv 0$.

Notice that setting up the matrix vector formulation of the system (4.8) that defines our approximation $u^\delta$ requires computing the vectors

$$
\begin{align*}
[\delta g, \psi_\mu \otimes \phi_\nu]_{L_2(\Omega)} & \in I^Y_{\frac{\delta}{2},0}, \\
[\delta u_0, \phi_\nu]_{L_2(\Omega)} & \in \{\nu : (\lambda, \nu) \in I_0, \sigma_\mu(0) \neq 0\}
\end{align*}
$$

which can be performed in $O(\dim X^\delta)$ operations because $I_\delta$ and $I^Y_{\frac{\delta}{2},0}$ are lower sets (and $\# I^Y_{\frac{\delta}{2},0} \lesssim \# I_\delta$).

## 7 Numerical results

We test our algorithm on the heat equation, i.e. the parabolic problem with $a(t; \eta, \zeta) = \int_\Omega \nabla \eta \cdot \nabla \zeta d\mathbf{x}$, posed on a two-dimensional polygonal spatial domain $\Omega$, and Dirichlet boundary $\Gamma_D = \partial \Omega$. Recall from Section 6.3 the three-point continuous piecewise linear temporal wavelet basis $\Sigma$, the orthonormal discontinuous piecewise linear temporal wavelet basis $\Psi$, and the hierarchical continuous piecewise linear spatial basis $\Xi := \{\phi_\nu : \nu \in \mathbb{N}_0\}$.

We consider ‘trial’ spaces $X^\delta$ spanned by finite subsets of $\Sigma \otimes \Xi$ whose index sets are lower sets (cf. (6.1)–(6.2)), and corresponding ‘test’ spaces $Y^\delta$ spanned by finite...
subsets of $\Psi \otimes \Xi$ as defined in (6.3)–(6.4). We construct the enlarged trial space $X_\delta$ as defined in Def. 6.3, with corresponding test space $Y_\delta$.

For a given level $N \in \mathbb{N}$, $\text{span}\{\sigma_\lambda : |\lambda| \leq N\}$ coincides with the span of the continuous piecewise linears on an $N$-times recursive dyadic refinement of $I$, and $\text{span}\{\phi_\nu \in \Xi : \text{gen}(\nu) \leq 2N\}$ coincides with that of the continuous piecewise linears, zero at $\partial \Omega$, on a $2N$-times recursive bisection refinement of an initial partition $\mathcal{T}_\perp$. Therefore, the span of the ‘full’ tensor product $\{\sigma_\lambda : |\lambda| \leq N\} \otimes \{\phi_\nu : \text{gen}(\nu) \leq 2N\}$ equals a space of lowest order continuous finite elements w.r.t. a quasi-uniform shape regular product mesh into prismatic elements.

Taking only those index pairs $(\lambda, \nu)$ for which $2|\lambda| + \text{gen}(\nu) \leq 2N$ produces a ‘sparse’ tensor product on level $N$. Sparse tensor products overcome the curse of dimensionality in the sense that for smooth solutions they achieve a rate in $X$-norm equal to the best rate in $H^1(\Omega)$-norm that can be expected for the corresponding stationary problem on the spatial domain, here the Poisson equation; cf. Section 5.5.

We run our adaptive Algorithm 4.8 with $\theta = 0.5$ and $\xi = \frac{1}{2}$, computing $\delta g$ and $\delta u_0$ as in Section 6.4. Since we envisage that in our experiments data-oscillation errors are not dominant, for our convenience we took $\omega = \infty$. We solve the arising linear system of (4.8) using Preconditioned CG, using the previous solution as initial guess. We then perform Dörfler marking on the residual, yielding a minimal set $J$, and finally choose $I_\delta$ as the smallest lower set containing $J \cup I_\delta$. Due to this constraint generally we add index pairs outside of the marked set, i.e. $I_\delta \setminus I_\delta \supseteq J$. Still, in our experiments, we observe $\#I_\delta - \#I_\delta \lesssim \#J$ with a moderate constant.

Remark 7.1 Rather we would have applied an algorithm that produces a $I_\delta$ such that $I_\delta \setminus I_\delta$ is guaranteed to have an, up to a multiplicative factor, smallest cardinality among all lower sets $I_\delta \supseteq I_\delta$ that realize the bulk criterion. Such an algorithm was introduced in [4, 5] for ‘single-tree’ approximation, but seems not to be available for the ‘double-tree’ (i.e. lower set) constraint that we need here.

We compare adaptive refinement with non-adaptive full- and sparse tensor products, and monitor the $X$-norm error estimator from Proposition 4.5, the residual error estimator from Proposition 4.3, and the $L^2(\Omega)$ trace error at $t = 0$.

7.1 Condition numbers of preconditioner

For the calibration of our preconditioners, we consider $\Omega := (0, 1)^2$, and compare uniformly refined space-time meshes with locally refined meshes with refinements towards $[0] \times \partial \Omega$.

The replacement of the nonlocal operator $(E_\delta^Y A E_\delta^Y)^{-1}$ in the forward application of $S^\delta \delta$ by the block-diagonal preconditioner $K_\delta^Y$ from Section 5.6.2 is only guaranteed to result in a convergent algorithm when the eigenvalues of $K_\delta^Y E_\delta^Y \delta A_\delta^Y \delta E_\delta^Y$ are sufficiently close to one.

In Table 1, we investigate the values $\kappa_\delta := \max\{\lambda_{\max}(K_\delta^Y A_\delta^Y), 1/\lambda_{\min}(K_\delta^Y A_\delta^Y)\}$ with $A_\delta^Y$ the matrix representation of $E_\delta^Y A_\delta^Y E_\delta^Y$, and $K_\delta^Y$ built from spatial multigrid
preconditioners \( K_{\mu}^\delta \) corresponding to \( n \) V-cycles. In each V-cycle we applied one pre- and one post Gauss–Seidel smoother. In case of a locally refined spatial mesh, on each level these Gauss–Seidel updates were restricted to the vertices whose generation is equal to that level as well as both endpoints of the edge on which these vertices were inserted ([44]). We see that for both uniform and locally refined space-time meshes, \( \kappa_\delta \) converges to 1 rapidly in \( n \), and is essentially independent of \( \text{dim} \ X^\delta \). In our examples, \( \kappa_\delta \) is close enough to one already for \( n = 1 \).

Fixing \( n = 1 \) for the forward application of \( S^\delta_{\mu} \), we want to precondition \( S^\delta_{\mu} \) itself as well. Following Section 5.6.2, we build a block-diagonal preconditioner taking \( K_{\mu}^\delta \) to correspond to \( m \) V-cycles of the aforementioned multigrid method now applied to \( A_{\lambda}^\delta + 2|\lambda| M_{\lambda}^\delta \) with \( A_{\lambda}^\delta \) and \( M_{\lambda}^\delta \) being stiffness- and mass-matrices. Table 2 shows the condition numbers of the preconditioned matrix. We again see fast stabilization in \( m \) as well as in \( \text{dim} \ X^\delta \). We fix \( m = 3 \) in the sequel. Most interestingly, in every of our example problems, only one or two PCG iterations suffice to fulfill the stopping criterion \( \tilde{t}_\delta \leq \frac{\delta}{2} \) on the algebraic error in Algorithm 4.8.

### 7.2 Smooth problem

We consider the square domain \( \Omega := (0, 1)^2 \) and prescribe

\[
u(t, x, y) := (1 + t^2)x(1 - x)y(1 - y)
\]

Table 2  Spectral condition numbers of \( K_{\lambda}^\delta S_{\mu}^\delta \), using spatial multigrid with \( m \) V-cycles

| \( \text{dim} \ X^\delta \) | \( m = 1 \) | \( m = 2 \) | \( m = 3 \) | \( m = 4 \) | \( m = 5 \) | \( m = 6 \) |
|--------------------------|----------|----------|----------|----------|----------|----------|
| uniform                  |          |          |          |          |          |          |
| \( 4913 \)               | 9.196    | 6.119    | 6.048    | 6.042    | 6.041    | 6.041    |
| \( 35937 \)              | 9.718    | 6.315    | 6.263    | 6.260    | 6.260    | 6.260    |
| \( 274625 \)             | 9.991    | 6.750    | 6.749    | 6.751    | 6.752    | 6.752    |
| \( 2146689 \)            | 10.115   | 7.080    | 7.087    | 7.088    | 7.088    | 7.088    |
| local                    |          |          |          |          |          |          |
| \( 3520 \)               | 5.707    | 5.132    | 5.110    | 5.111    | 5.111    | 5.111    |
| \( 30151 \)              | 6.355    | 5.734    | 5.706    | 5.704    | 5.704    | 5.704    |
| \( 244870 \)             | 7.619    | 6.879    | 6.843    | 6.841    | 6.841    | 6.841    |
| \( 1964797 \)            | 9.353    | 8.734    | 8.703    | 8.701    | 8.701    | 8.701    |
with derived data \( u_0 \) and \( g \). For this smooth solution, full and sparse tensor products are expected to yield the best possible error decays proportional to \((\dim X^\delta)^{-1/3}\) and \((\dim X^\delta)^{-1/2}\), respectively.

The left side of Fig. 5 shows the error progressions for the smooth problem. We plot the error estimator \( \| \delta u - \tilde{u}^\delta \|_{X^\delta} \approx \| u - \tilde{u}^\delta \|_X \) from Proposition 4.5, the residual error estimator \( \| r^\delta \| \), and \( \| \gamma_0 (\delta u - \tilde{u}^\delta) \|_{L_2(\Omega)} \). We see that the error progressions are as expected. For this solution, adaptive refinement yields no advantage over sparse grid refinement. We observe a higher order of convergence for the trace at \( t = 0 \) measured in \( L_2(\Omega) \).

### 7.3 Moving peak problem [28]

Again for \( \Omega := (0, 1)^2 \), we select

\[
u(t, x, y) := x(1 - x)y(1 - y) \exp(-100[(x - t)^2 + (y - t)^2]).\]

The solution is smooth, and almost zero everywhere except on a small strip near the diagonal from \((0, 0, 0)\) to \((1, 1, 1)\) of the space-time cylinder. As \( u \) is smooth, we expect sparse grid refinements to asymptotically yield the optimal error decay proportional to \((\dim X^\delta)^{-1/2}\), albeit with a terrible constant. Adaptive refinement should be able to achieve the same rate at quantitatively smaller doubletrees.

The right of Fig. 5 shows that the sparse grid rate is not (yet) optimal, while our adaptive routine is able to find the optimal rate from \( \dim X^\delta \approx 10^3 \) onwards. Figure 6 shows the number of basis functions \( \sigma_\lambda \otimes \phi_\nu \) whose supports intersect given points in the time-space cylinder. We see the adaptation to the moving peak.
Fig. 6 Moving peak problem, adaptive lower set with dim $X_\delta = 89401$. Shown: $\# \{ (\lambda, \nu) \in I_\delta : (t, x, y) \in \text{supp } \sigma_\lambda \otimes \phi_\nu \}$ for a selection of times $t$

7.4 Cylinder problem

Selecting the L-shaped domain $\Omega := (-1, 1)^2 \setminus (-1, 0)^2$ with data $u_0 = 0$ and $g(t, x, y) := t \cdot 1_{\{x^2+y^2<1/4\}}$, the true solution is known to be singular at the re-entrant corner and at the wall of the cylinder $\{(t, x, y) : x^2 + y^2 = 1/4\}$. We took this example from [18]. The left side of Fig. 7 shows the error progression for this cylinder problem. We see that the full grid error decay proportional to $(\text{dim } X_\delta)^{-1/4}$ is improved to $(\text{dim } X_\delta)^{-1/3}$ by considering sparse grids. Adaptive refinement, however, achieves the best possible error decay proportional to $(\text{dim } X_\delta)^{-1/2}$, recovering the rate for a smooth solution.

7.5 Singular problem

Again for $\Omega := (-1, 1)^2 \setminus (-1, 0)^2$, we select data $u_0=1$, and $g = 0$. The solution has a strong singularity along $\{0\} \times \partial \Omega$ due to the incompatibility of initial- and boundary conditions, in addition to the singularity at the re-entrant corner $(0, 0)$. At the right of Fig. 7, for uniform refinement, we see the extremely slow error decay proportional to $(\text{dim } X_\delta)^{-1/11}$, already found in [18]. Interestingly, sparse grid refinement offers no rate improvement over full grid refinement. The adaptive algorithm yields a much better error decay proportional to $(\text{dim } X_\delta)^{-2/5}$. We observed that increasing the Dörfler marking parameter to $\theta = 0.7$ decreases the convergence rate to $-1/3$, whereas a $\theta$ smaller than 0.5 did not improve the rate beyond $-2/5$. Looking at Fig. 8, we see strong adaptivity towards $\{0\} \times \partial \Omega$ and $I \times \{(0, 0)\}$, even seeing basis functions $\sigma_\lambda \otimes \phi_\nu$ of $X_\delta$ whose barycenter is at $t = 2^{-14} \approx 10^{-4}$. 
7.6 Gradedness and error reduction

In Section 4, we used (4.2) to demonstrate proportionality of $\|r^\delta\|$ and $\|u - u^\delta\|_X$, as well as a constant error reduction in each iteration of the adaptive algorithm. In Proposition 6.6, we showed that (4.2) holds when the gradedness $L_\delta$ of Definition 6.4 is uniformly bounded.

In the left picture of Fig. 9, we see however a more than expected increase in gradedness, where in particular for the singular problem we observe a logarithmic increase in terms of dim $X^\delta$. However, this turns out not to be a problem in practice:
Figures 5 and 7 demonstrate that the residual error $\|r^\delta\|$ and the estimated $X$-norm error $\|\delta u - \tilde{u}^\delta\|_{X^{\infty}}$ are very close, and even converge for the singular problem. Moreover, in the right picture of Fig. 9, we see a constant error reduction of $\bar{\rho} \approx 0.89$ at every step of the adaptive algorithm, and hence, that the conclusion of Theorem 4.9 holds in practice.

Fig. 10 Total runtime and peak memory consumption as function of dim $X^\delta$, measured after every iteration of the adaptive loop, for the four different model problems.
7.7 Total runtime and memory consumption

Figure 10 shows the total runtime and peak memory consumption after every iteration of the adaptive algorithm. The top row shows absolute values, and the bottom row values relative to \( \dim X^\delta \).

The left of the figure shows that the adaptive algorithm runs in optimal linear time in the dimension of the current trial space.

The right of the figure shows that the peak memory is linear as well, stabilizing to around 15kB per degree of freedom. This is relatively high, mainly because our implementation uses trees rather than hash maps to represent vectors to ensure a linear-time implementation of the matrix-vector products (cf. Rem. 6.2).

8 Conclusion

We constructed an adaptive solver for a space-time variational formulation of parabolic evolution problems. The collection of trial spaces are given by the spans of sets of tensor products of wavelets-in-time and hierarchical basis functions-in-space. Compared to our previous works [9, 32] where we employed ‘true’ wavelets also in space, the theoretical results are weaker. We demonstrated \( r \)-linear convergence of the adaptive routine, but have not shown optimal rates at linear complexity. On the other hand, the runtimes we obtained with the current approach are much better.

Funding The second and third authors have been supported by the Netherlands Organization for Scientific Research (NWO) under contract. no. 613.001.652

Declarations

Conflict of interest The authors declare no competing interests.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Alpert, B.K.: A class of bases in \( L^2 \) for the sparse representation of integral operators. SIAM J. Math. Anal. 24, 246–262 (1993)
2. Andreev, R.: Stability of sparse space-time finite element discretizations of linear parabolic evolution equations.IMA J. Numer. Anal. 33(1), 242–260 (2013). https://doi.org/10.1093/imanum/drs014
3. Andreev, R.: Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations. SIAM J. Sci. Comput. 38(1), A216–A242 (2016). https://doi.org/10.1137/140998639
4. Binev, P., DeVore, R.: Fast computation in adaptive tree approximation. Numer. Math. 97(2):193–217 (2004)
5. Binev, P., Fierro, F., Veeser, A.: Near-best adaptive approximation on conforming meshes. arXiv:1912.13437 (2019)
6. Beranek, N., Reinhold, M.A., Urban, K.: A space-time variational method for optimal control problems. arXiv:2010.00345 (2020)
7. Balder, R., Zenger, Ch.: The solution of multidimensional real Helmholtz equations on sparse grids. SIAM. J. Sci. Comput. 17(3), 631–646 (1996)
8. Cohen, A., Dahmen, W., DeVore, R.: Adaptive wavelet methods for elliptic operator equations – Convergence rates. Math. Comp. 70, 27–75 (2001)
9. Chegini, N.G., Stevenson, R.P.: Adaptive wavelet schemes for parabolic problems: Sparse matrices and numerical results. SIAM J. Numer. Anal. 49(1), 182–212 (2011)
10. Devaud, D.: Petrov-Galerkin space-time hp-approximation of parabolic equations in $H^{1/2}$. IMA J. Numer. Anal. 40(4), 2717–2745 (2020). https://doi.org/10.1093/imanum/drz036
11. Dyja, R., Ganapathysubramanian, B., van der Zee, K.G.: Parallel-in-space-time, adaptive finite element framework for nonlinear parabolic equations. SIAM J. Sci Comput. 40(3), C283–C304 (2018). https://doi.org/10.1137/16M108985X
12. Diening, L., Kreuzer, Ch., Stevenson, R.P.: Instance optimality of the adaptive maximum strategy. Found Comput. Math., 1–36. https://doi.org/10.1007/s10208-014-9236-6 (2015)
13. Dautray, R., Lions, J.-L.: Mathematical analysis and numerical methods for science and technology. Vol. 5. Springer, Berlin. Evolution problems I. https://doi.org/10.1007/978-3-642-58090-1 (1992)
14. Diening, L., Storn, J.: A space-time DPG method for the heat equation. Comput. Math. Appl., 105:41–53. https://doi.org/10.1016/j.camwa.2021.11.013 (2022)
15. Diening, L., Storn, J., Tcherpel, T.: On the Sobolev and $L^p$-stability of the $L^2$-projection. SIAM J. Numer. Anal. 59(5), 2571–2607 (2021). https://doi.org/10.1137/20M1358013
16. Dahmen, W., Stevenson, R.P., Westerdiep, J.: Accuracy controlled data assimilation for parabolic problems (2021) 40 pages Accepted for publication in. Math. Comp. arXiv:2105.05836
17. Ern, A., Smears, I., Vohralík, M.: Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. SIAM J. Numer. Anal. 55(6), 2811–2834 (2017). https://doi.org/10.1137/16M1097626
18. Führer, T., Karkulik, M.: Space-time least-squares finite elements for parabolic equations. Comput. Math. Appl. 92, 27–36 (2021). https://doi.org/10.1016/j.camwa.2021.03.004
19. Gunzburger, M.D., Kunoth, A.: Space-time adaptive wavelet methods for control problems constrained by parabolic evolution equations. SIAM J. Contr. Optim. 49(3), 1150–1170 (2011)
20. Griebel, M., Oeltz, D.: A sparse grid space-time discretization scheme for parabolic problems. Computing 81(1), 1–34 (2007)
21. Gimperlein, H., Stocek, J.: Space-time adaptive finite elements for nonlocal parabolic variational inequalities. Comput. Methods Appl. Mech Engrg. 352, 137–171 (2019). https://doi.org/10.1016/j.cma.2019.04.019
22. Hofer, Ch., Langer, U., Neumüller, M., Schneckenleitner, R.: Parallel and robust preconditioning for space-time isogeometric analysis of parabolic evolution problems. SIAM J. Sci. Comput. 41(3), A1793–A1821 (2019). https://doi.org/10.1137/18M1208794
23. Kondratyuk, Y., Stevenson, R.P.: An optimal adaptive finite element method for the Stokes problem. SIAM J. Numer. Anal. 46(2), 747–775 (2008)
24. Kestler, S., Stevenson, R.P.: Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions. J. Comput. Appl. Math. 260, 103–116 (2014)
25. Kestler, S., Steih, K., Urban, K.: An efficient space-time adaptive wavelet Galerkin method for time-periodic parabolic partial differential equations. Math. Comp. 85(299), 1309–1333 (2016). https://doi.org/10.1090/mcom/3009
26. Larsson, S., Molteni, M.: Numerical solution of parabolic problems based on a weak space-time formulation. Comput. Methods Appl. Math. 17(1), 65–84 (2017). https://doi.org/10.1515/cmam-2016-0027
27. Langer, U., Moore, S.E., Neumüller, M.: Space-time isogeometric analysis of parabolic evolution problems. Comput. Methods Appl. Mech. Engrg. 306, 342–363 (2016). https://doi.org/10.1016/j.cma.2016.03.042
28. Langer, U., Schafelner, A.: An optimal adaptive finite element method for Non-autonomous Parabolic Problems with Distributional Sources. Comput. Methods Appl. Math. 20(4), 677–693 (2020). https://doi.org/10.1515/cmam-2020-0042
29. Neumüller, M., Smears, I.: Time-parallel iterative solvers for parabolic evolution equations. SIAM J. Sci. Comput. 41(1), C28–C51 (2019). https://doi.org/10.1137/18M1172466
30. Olshanskii, M.A., Reusken, A.: On the convergence of a multigrid method for linear reaction-diffusion problems. Computing 65(3), 193–202 (2000). https://doi.org/10.1007/s006070070006
31. Pearson, J.W., Wathen, A.J.: A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. Numer. Linear Algebra Appl. 19(5), 816–829 (2012). https://doi.org/10.1002/nla.814
32. Rekatsinas, N., Stevenson, R.P.: An optimal adaptive tensor product wavelet solver of a space-time FOSLS formulation of parabolic evolution problems. Adv. Comput. Math. 45(2), 1031–1066 (2019). https://doi.org/10.1007/s10444-018-9644-2
33. Schwab, Ch., Stevenson, R.P.: A space-time adaptive wavelet method for parabolic evolution problems. Math Comp. 78, 1293–1318 (2009). https://doi.org/10.1090/S0025-5718-08-02205-9
34. Stevenson, R.P.: The frequency decomposition multi-level method: a robust additive hierarchical basis preconditioner. Math. Comp. 65(215), 983–997 (1996)
35. Stevenson, R.P.: The completion of locally refined simplicial partitions created by bisection. Math. Comp. 77, 227–241 (2008)
36. Steinbach, O.: Space-time finite element methods for parabolic problems. Comput. Methods Appl. Math. 15(4), 551–566 (2015). https://doi.org/10.1515/cmam-2015-0026
37. Stevenson, R.P., van Venetië, R.: Uniform preconditioners for problems of negative order. Math. Comp. 89(322), 645–674 (2020). https://doi.org/10.1090/mcom/3481
38. Stevenson, R.P., Westerdiep, J.: Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. IMA J. Numer. Anal. 41(1), 28–47 (2021). https://doi.org/10.1093/imanum/drz069
39. Steinbach, O., Yang, H.: Comparison of algebraic multigrid methods for an adaptive space-time finite-element discretization of the heat equation in 3D and 4D. Numer. Linear Algebra Appl. 25(3), e2143, 17 (2018). https://doi.org/10.1002/nla.2143
40. Steinbach, O., Zank, M.: Coercive space-time finite element methods for initial boundary value problems. Electron. Trans. Numer. Anal. 52:154–194. https://doi.org/10.1553/etna_vol52s154 (2020)
41. van Venetië, R., Westerdiep, J.: A parallel algorithm for solving linear parabolic evolution equations. 33–50, Springer Proc. Math. Stat., 356, Springer, Cham (2021)
42. van Venetië, R., Westerdiep, J.: Efficient space-time adaptivity for parabolic evolution equations using wavelets in time and finite elements in space. arXiv:2104.08143 (2021)
43. Wloka, J.: Partielle Differentialgleichungen. B. G. Teubner, Stuttgart. Sobolevräume und Randwertaufgaben (1982)
44. Wu, J., Zheng, H.: Uniform convergence of multigrid methods for adaptive meshes. Appl. Numer Math. 113, 109–123 (2017). https://doi.org/10.1016/j.apnum.2016.11.005
45. Zitelli, J., Muga, I., Demkowicz, L., Gopalakrishnan, J., Pardo, D., Calo, V.M.: A class of discontinuous Petrov-Galerkin methods. Part IV: the optimal test norm and time-harmonic wave propagation in 1D. J. Comput. Phys. 230(7), 2406–2432 (2011)

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.