AN EFFICIENT SPACE-TIME ADAPTIVE WAVELET
GALERKIN METHOD FOR TIME-PERIODIC PARABOLIC
PARTIAL DIFFERENTIAL EQUATIONS

SEBASTIAN KESTLER, KRISTINA STEIH, AND KARSTEN URBAN

ABSTRACT. We introduce a multitree-based adaptive wavelet Galerkin algo-
rithm for space-time discretized linear parabolic partial differential equations,
 focusing on time-periodic problems. It is shown that the method converges
with the best possible rate in linear complexity and can be applied for a wide
range of wavelet bases. We discuss the implementational challenges arising
from the Petrov-Galerkin nature of the variational formulation and present
numerical results for the heat and a convection-diffusion-reaction equation.

1. Introduction

In recent years, space-time variational approaches for linear parabolic partial
differential equations (PDEs) of the form
\[ u_t + A(t)[u] = g(t) \quad \text{on } \Omega \subset \mathbb{R}^n, \text{ for } t \in [0,T], \]
have been considered in various contexts. These methods treat both temporal and
spatial variables simultaneously, allowing e.g. for targeted adaptive refinement of
the numerical discretization in the full space-time domain or efficient paralleliza-
tion. On the other hand, this in general amounts to solving an \((n+1)\)-dimensional
problem. This differs from standard time-stepping techniques for time-dependent
PDEs, which are usually based on semi-discretization schemes: The vertical method
of lines requires the solution of a system of coupled ordinary differential equations
that arise from a discretization in space. Within the horizontal method of lines
and the discontinuous Galerkin method, the temporal variable is discretized first,
leading to a (coupled) sequence of elliptic problems in the spatial domain. Such
time-stepping schemes have some drawbacks: The sequential treatment of the time
variable often does not allow for parallelization in time. Furthermore, adaptive
schemes typically focus either on the spatial or on the temporal variable or are
based on local error estimators (e.g. [Raa07]), thus forfeiting optimality. Moreover,
a posteriori error estimators – needed e.g. for adaptive schemes or model reduction
approaches – are usually increasing functions in time, therefore losing efficiency over
long time horizons. These issues are amplified when considering time-periodic prob-
lems, i.e., when searching for solutions \( u \) with \( u(0) = u(T) \). Such problems arise

Received by the editor August 2, 2013 and, in revised form, October 28, 2014.
2010 Mathematics Subject Classification. 35B10, 41A30, 41A63, 65N30, 65Y20.
Key words and phrases. Time-periodic problems, tensor product approximation, wavelets,
adaptivity, optimal computational complexity.

This work has partly been supported by the Deutsche Forschungsgemeinschaft within the
Research Training Group (Graduiertenkolleg) GrK1100 Modellierung, Analyse und Simulation in
der Wirtschaftsmathematik at Ulm University.

©2015 American Mathematical Society
naturally in different physical, biological or chemical models, e.g. flows around a rotor or propeller, biological models or chemical engineering [JPSU07, SZ99, KB06]. Standard numerical methods for periodic problems require either an additional fixed-point scheme (when using a temporal semi-discretization) or the solution of a system of boundary value problems (in case of the method of lines), both entailing non-negligible additional computational effort. In the present work, we will therefore consider a space-time variational formulation for time-periodic problems.

Space-time variational formulations for initial value problems in particular include space-time multigrid methods [HV95], space-time sparse grids [And13, GO07] or space-time wavelet collocation methods [AKV06]. Other space-time formulations based on special test bases or discontinuous Galerkin methods are e.g. [MV07, UP12, UP14]. These approaches exploit the space-time approach mainly for theoretical considerations and allow the use of a time-stepping scheme – thus effectively circumventing the main drawback of space-time methods, i.e., the additional dimension introduced by the temporal variable. However, optimality has not yet been investigated in such a framework.

Here, we follow the approach proposed by [SS09], where a space-time adaptive scheme using tensorized wavelet bases is proven to be optimal for initial value problems. In this setting, the partial differential equation is reformulated as an equivalent non-symmetric bi-infinite matrix-vector problem of the form $Bu = f$ and is numerically approximated by employing an adaptive wavelet Galerkin method (AWGM) to the corresponding normal equations. As opposed to standard algorithms for time-periodic problems, the upshot of this approach is that time-periodic boundary conditions can be incorporated into the underlying ansatz basis.

AWGMs may be described as follows [CDD01, GHS07]: Consider a bi-infinite linear system $Cu = g$ in $\ell_2$ with a symmetric positive definite (s.p.d.) stiffness matrix $C : \ell_2 \rightarrow \ell_2$, an infinite right-hand side $g \in \ell_2$ and a unique solution $u \in \ell_2$ which arises e.g. from the wavelet discretization of an elliptic operator problem. In each iteration, these bi-infinite problems are approximated on a finite-dimensional index set $\Lambda_k$ steering the local refinement. This means that a finite vector $u_{\Lambda_k}$ satisfying $C_{\Lambda_k \times \Lambda_k} u_{\Lambda_k} = g|_{\Lambda_k}$ is computed. The (infinite-dimensional) residual $r_{\Lambda_k} := g - Cu_{\Lambda_k}$ is then approximated to serve as an error estimator and to identify an update, i.e., an (usually refined) index set $\Lambda_{k+1}$ (typically $\Lambda_{k+1}$ corresponds to the significant coefficients of $r_{\Lambda_k}$).

Space-time variational approaches to parabolic problems lead to non-symmetric Petrov-Galerkin formulations and hence do not directly fit into this framework. In particular, the bi-infinite associated stiffness matrix is no longer s.p.d., which is, however, a crucial ingredient for the convergence analysis of AWGMs. Moreover, the residual belonging to a test space which is not identical to the trial space does not directly convey information for an update of the trial space. So, working with the normal equations, i.e., with the s.p.d. operator $C := B^\top B$ and right-hand side $g = B^\top f$, is a natural approach for initial value problems (e.g. [CS11, SS09]) and also for time-periodic problems, as we will show in this article.

The treatment of normal equations by adaptive wavelet methods was first discussed in [CDD02]. The main difficulty lies in the (approximate) evaluation of $B^\top B$ and $B^\top f$. There are several approaches in the literature that address these issues. The techniques proposed in [CDD01] rely on so-called wavelet matrix compression schemes, whereas [CS11, CS12] use special wavelet constructions leading
to truly sparse matrices $B$. In the present work, we use another approach based on multirule-structured index sets as introduced in [KS13,KS14]. This allows for the exact application of $B$ for wavelet discretizations of linear differential operators with polynomial coefficients within linear complexity when the underlying wavelet basis is of tensor product type. It basically consists of the so-called unidirectional principle first introduced in sparse grid algorithms (e.g. [BG04,Zen91]) where coordinate directions may be treated separately. The evaluation is then based on multirule-structured index sets that permit a tree structure when all but one coordinate direction are frozen.

We stress that, to the best of our knowledge, no quantitative results on the numerical solution of parabolic operator problems using the multirule concept within an AWGM are available. So far, only numerical results for elliptic operator problems are presented in [KS13, KS14]. Even though we focus on time-periodic problems, we expect that our results can be extended to non-periodic settings as well.

The outline of this article is as follows. In Section 2, we introduce parabolic problems with periodic boundary conditions in time. The derivation of the equivalent $\ell_2$-problem by means of tensor product wavelet bases is explained in Section 3. Next, in Section 4, we define quasi-optimal algorithms showing what can be expected in terms of convergence rates and computational work. Some details on wavelet bases are then given in Section 5. In Section 6, we explain AWGMs for elliptic problems and extend it to our parabolic problem. The realization and analysis of an efficient, multirule-based AWGM is then presented in Sections 7 and 8. We underline our theoretical findings by numerical experiments in Section 9.

2. Time-periodic parabolic problems

Let $\Omega := \Omega_1 \times \cdots \times \Omega_n \subset \mathbb{R}^n$ be a product domain and $V$ be a real separable Hilbert space with dual $V'$ such that $V \hookrightarrow H := L_2(\Omega) \hookrightarrow V'$ is a Gelfand triple. For $A(t) \in \mathcal{L}(V,V')$ and $g \in L_2(0,T;V')$ we consider the time-periodic equation

\begin{equation}
(2.1) \quad u_t + A(t)[u(t)] = g(t) \text{ in } V' \text{ for a.e. } t \in [0,T], \quad u(0) = u(T) \text{ in } H.
\end{equation}

Denoting by $\langle \cdot, \cdot \rangle_{V \times V'}$ the duality pairing on $V \times V'$, we assume that $t \mapsto \langle v, A(t)[u] \rangle$ is measurable on $[0,T]$ and that $A(t)$ is uniformly coercive and bounded in time; i.e., there exist $0 < \alpha \leq \alpha(t), \infty > \gamma \geq \gamma(t)$ such that for a.e. $t \in [0,T]$,

\begin{equation}
\langle v, A(t)[w] \rangle_{V \times V'} \leq \gamma \|w\|_{V'} \|v\|_V, \quad \langle v, A(t)[v] \rangle_{V \times V'} \geq \alpha \|v\|_V^2 \quad \forall v,w \in V.
\end{equation}

Moreover, we assume that the space $V$ is a Sobolev space of non-negative order

\begin{equation}
(2.3) \quad V := \bigcap_{i=1}^n \bigotimes_{j=1}^n W_{ij}, \text{ where } W_{ij} := \begin{cases} L_2(\Omega_i), & i \neq j, \\ V^{(i)}, & i = j, \end{cases}
\end{equation}

and, for a fixed $m \in \mathbb{N}$, $V^{(i)}$ is either $H^m(\Omega_i)$ or a subspace incorporating essential boundary conditions. Note that several partial differential operators allow such a structure. As an example, think of $V = H_0^1(\Omega)$ and $V^{(i)} = H_0^1(\Omega_i)$ (see [GOG95]).

2.1. Space-time formulation. We derive a variational formulation where the temporal periodicity can be integrated into the function spaces and is therefore ultimately incorporated into the basis of a discrete approximation space. To this end, we derive a space-time variational formulation: Defining

\begin{equation}
(2.4) \quad H^1_{\text{per}}(0,T) := \{v \in H^1(0,T) : v(0) = v(T)\},
\end{equation}
we consider the spaces \( \mathcal{Y} := L_2(0, T; V) \) and \( \mathcal{X} := L_2(0, T; V) \cap H^1_{\text{per}}(0, T; V') \), i.e.,
\[
(2.5) \quad \mathcal{X} = \{ v \in L_2(0, T; V) : v_t \in L_2(0, T; V') \}, \quad v(0) = v(T) \text{ in } H,
\]
where \( \mathcal{X} \) is equipped with the norm \( \|v\|_\mathcal{X}^2 := \|v\|_{L_2(0,T;V)}^2 + \|v_t\|_{L_2(0,T;V')}^2 \), \( v \in \mathcal{X} \). Note that \( v(0), v(T) \) are well-defined due to \( H^1(0, T) \subset C([0, T]) \) and \{ \( v \in L_2(0, T; V) : v_t \in L_2(0, T; V') \} \subset C(0, T; H) \), e.g. [DL92]. By integration of (2.1) over \([0, T]\), we obtain the problem:
\[
(2.6) \quad \text{Find } u \in \mathcal{X} : \quad b(u, v) = f(v) \quad \forall v \in \mathcal{Y},
\]
with forms \( b(\cdot, \cdot) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}, f(\cdot) : \mathcal{Y} \to \mathbb{R} \) defined by [SS09] (5.6)-(5.7):
\[
(2.7) \quad b(u, v) := \int_0^T \langle v(t), u_t(t) + \mathcal{A}(t)[u] \rangle_{V \times V'} dt, \quad f(v) := \int_0^T \langle v(t), g(t) \rangle_{V \times V'} dt.
\]
We define the space-time operator \( B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}' \cap H^1(0, T)) \) by \( \langle v, B[u] \rangle := b(u, v) \) with \( \langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathcal{X} \times \mathcal{Y}} \), so that (2.6) is a variational formulation of the operator equation:
\[
(2.8) \quad \text{Find } u \in \mathcal{X} : \quad B[u] = f, \quad f \in \mathcal{Y}'.
\]

2.2. Well-posedness. The well-posedness of a space-time formulation of (non-periodic) initial value problems has been discussed in [SS09]. In Appendix A we verify the Babuška-Aziz conditions:

(i) Continuity: \( \gamma_B := \sup_{0 \neq v \in \mathcal{Y}} \sup_{0 \neq u \in \mathcal{X}} \frac{b(u, v)}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} < \infty \).

(ii) Inf-sup condition: \( \beta_B := \inf_{0 \neq u \in \mathcal{X}} \sup_{0 \neq v \in \mathcal{Y}} \frac{b(u, v)}{\|v\|_{\mathcal{Y}}} > 0 \).

(iii) Surjectivity: \( \sup_{0 \neq u \in \mathcal{X}} |b(u, v)| > 0 \) for all \( 0 \neq v \in \mathcal{Y} \).

**Proposition 2.1.** Problem (2.6) is well-posed. In particular, \( B \) from (2.8) is boundedly invertible with \( \|B\| = \gamma_B = \sqrt{T} \max\{1, \gamma\} \), \( \|B^{-1}\| = \frac{1}{\beta_B} = \frac{\sqrt{2} \max\{1, \alpha^{-1}\}}{\alpha \min\{1, \gamma^{-1}\}} \).

3. Equivalent bi-infinite matrix-vector problem

We consider the reformulation of (2.8) as an equivalent \( \ell_2 \)-problem, i.e., a discrete problem posed on the sequence space \( \ell_2 \). This was first introduced in [CDD01] [CDD02] for stationary problems and extended to parabolic problems in [SS09].

3.1. Riesz bases. We recall that for a separable Hilbert space \( \mathcal{H} \) of infinite dimension, a dense collection \( \mathcal{Y} := \{ \gamma_i : i \in \mathbb{N} \} \subset \mathcal{H} \) is called a Riesz basis for \( \mathcal{H} \) if there exist constants \( c, C > 0 \) such that for \( v = \sum_{i=1}^{\infty} v_i \gamma_i \), it holds that
\[
(3.1) \quad c \|v\|_{\ell_2(\mathbb{N})}^2 \leq \|v\|_{\mathcal{H}}^2 \leq C \|v\|_{\ell_2(\mathbb{N})}^2 \quad \forall v = (v_i)_{i \in \mathbb{N}} \in \ell_2(\mathbb{N}).
\]
The largest \( c \) and the smallest \( C \) for which (3.1) holds are referred to as lower and upper Riesz constants and are denoted by \( c_\gamma(\mathcal{H}) \) and \( C_\gamma(\mathcal{H}) \), respectively.\(^1\)

\(^1\) Sometimes a different definition of Riesz constants is used, namely \( c_\gamma(\mathcal{H}) \) and \( C_\gamma(\mathcal{H}) \) being the largest and the smallest constant such that \( c_\gamma(\mathcal{H}) \|v\|_{\ell_2(\mathbb{N})} \leq \|v\|_{\mathcal{H}} \leq C_\gamma(\mathcal{H}) \|v\|_{\ell_2(\mathbb{N})} \).
3.2. Wavelet discretization of the parabolic operator problem. Let us now consider two different Riesz bases:

\[ \hat{\Psi}^X := \{ \hat{\psi}^X_\lambda : \lambda \in \hat{J} \} \subset X, \quad \hat{\Psi}^Y := \{ \hat{\psi}^Y_\lambda : \lambda \in \hat{J} \} \subset Y, \]

labeled w.r.t. two (possibly) different countable index sets \( \hat{J} \) and \( \hat{J} \). More precisely, we consider a trial basis \( \hat{\Psi}^X \) for the ansatz space \( X \) and a test basis \( \hat{\Psi}^Y \) for the test space \( Y \) with associated Riesz constants \( c_X(\hat{\Psi}^X) \), \( C_X(\hat{\Psi}^X) \) and \( c_Y(\hat{\Psi}^Y) \), \( C_Y(\hat{\Psi}^Y) \). It is important to note that \( \hat{\Psi}^X \), \( \hat{\Psi}^Y \) arise from normalizing different Riesz bases \( \hat{\Psi}, \hat{\psi} \), for \( L_2((0,T) \times \Omega) \) w.r.t. \( \| \cdot \|_X \) and \( \| \cdot \|_Y \) (which is also the reason for our notation; see Section 3.2). Then there exists a unique expansion \( u = u^\top \hat{\Psi}^X \) of the solution \( u \) of (2.8) where we formally interpret both \( u \in \ell_2(\hat{J}) \) and \( \hat{\Psi}^X \) as column vectors. Now, the equivalent formulation of (2.8) reads as follows:

\[ \text{Find } u \in \ell_2(\hat{J}) : \quad Bu = f, \quad f \in \ell_2(\hat{J}), \]

where \( B := \left[ \langle \psi^X_\lambda, B[\psi^X_\mu] \rangle \right]_{\lambda,\mu \in \hat{J}} \) and \( f = \left[ \langle \psi^Y_\lambda, f \rangle \right]_{\lambda \in \hat{J}} \) is the bi-infinite stiffness matrix and \( f \) is the infinite right-hand side. It is easy to see that (3.3) is well-posed. Since, \( f \in \ell_2(\hat{J}) \) and \( B \in L(X,Y') \) are boundedly invertible, also \( B \in L(\ell(\hat{J}),\ell(\hat{J})) \) is boundedly invertible. In particular, with \( \| \cdot \| := \| \cdot \|_{\ell_2 \to \ell_2} \) (compare [SS99] (2.2) and (2.3)),

\[ \| B \| \leq \| B \|_{X \to Y'} C_X(\hat{\Psi}^X)^{1/2} C_Y(\hat{\Psi}^Y)^{1/2}, \quad \| B^{-1} \| \leq \frac{\| B^{-1} \|_{Y' \to X}}{C_X(\hat{\Psi}^X)^{1/2} C_Y(\hat{\Psi}^Y)^{1/2}}. \]

3.3. Further notation. We need to restrict the bi-infinite matrices \( B \) and \( B^\top \) in both rows and columns. For a pair \( (\Lambda, \tilde{J}) \) with \( \tilde{J} \subseteq \hat{J} \) and \( \Lambda \subseteq \tilde{J} \), set

\[ E_{\Lambda} : \ell_2(\Lambda) \to \ell_2(\tilde{J}) \quad \text{and} \quad R_{\Lambda} := E_{\Lambda}^\top : \ell_2(\tilde{J}) \to \ell_2(\Lambda), \]

where \( E_{\Lambda} \) is the trivial embedding, i.e., the extension of \( v_{\Lambda} \in \ell_2(\Lambda) \) by zeros to \( \ell_2(\tilde{J}) \). Consequently, its adjoint \( R_{\Lambda} \) is the restriction of \( v \in \ell_2(\tilde{J}) \) to \( v_{|\Lambda} \in \ell_2(\Lambda) \). For \( \tilde{\Lambda} \subseteq \tilde{J} \) and \( \Lambda \subseteq \tilde{J} \), we define the following restriction of \( B \) and \( B^\top \):

\[ \tilde{\Lambda} B_{\tilde{\Lambda}} := R_{\tilde{\Lambda}} B E_{\tilde{\Lambda}}, \quad B_{\Lambda} := \tilde{J} B_{\tilde{\Lambda}}, \quad \Lambda B_{\tilde{\Lambda}} := R_{\Lambda} B^\top E_{\tilde{\Lambda}}, \quad \tilde{\Lambda} B^\top := \tilde{J} B_{\tilde{\Lambda}}. \]

Finally, \( C \subseteq D \) means that \( C \) can be bounded by a constant times \( D \) and \( C \geq D \) is defined as \( D \leq C \). In this setting, \( C \approx D \) is defined as \( C \lesssim D \) and \( C \gtrsim D \).

4. Quasi-optimal algorithms
for bi-infinite matrix-vector problems

We may now focus on the approximate solution of (3.3). To this end, we first discuss what can be expected in terms of convergence rate and complexity.

4.1. Best \( N \)-term approximation. For a given number of degrees of freedom (d.o.f.) \( N \in \mathbb{N} \), the best approximation \( v_N \) of a function \( v = v^{\top} \hat{\Psi}^X \in X \) in the basis \( \hat{\Psi}^X \) with \( N \) d.o.f. is a nonlinear, best \( N \)-term approximation (e.g. [DeV98]), i.e., \( v_N = \arg \sigma_N(v) \), where the best \( N \)-term approximation error is defined as

\[ \sigma_N(v) := \inf_{\{ \Lambda \in \tilde{J} : \# \Lambda = N \}} \inf_{v_N \in \text{span}(\hat{\Psi}^X_{\Lambda} : \Lambda \in \tilde{\Lambda})} \| v - v_N \|_X. \]

Since \( \hat{\Psi}^X \) is a Riesz basis, it holds that \( \| v - v_N \|_{\ell_2} \approx \sigma_N(v) \) where \( v_N \) always denotes an \( N \)-term approximation of the vector \( v \) (i.e., the \( N \) largest coefficients
in modulus of \( \mathbf{v} \). As described in [DeV98], it is meaningful to collect all vectors \( \mathbf{v} \in \ell_2(\mathcal{F}) \) that permit an approximation rate \( s > 0 \) in the sense that \( \| \mathbf{v} - \mathbf{v}_N \|_{\ell_2} \lesssim N^{-s} \) within the nonlinear approximation class (compare [Ste09] (2)):

\[
A^s := \left\{ \mathbf{v} \in \ell_2(\mathcal{F}) : \| \mathbf{v} \|_{A^s} := \sup_{\varepsilon > 0} \varepsilon \left[ \min\{ N \in \mathbb{N}_0 : \| \mathbf{v} - \mathbf{v}_N \|_{\ell_2(\mathcal{F})} \leq \varepsilon \} \right]^s < \infty \right\}.
\]

For a given \( \mathbf{v} \in A^s \) and \( \varepsilon > 0 \), the required number of degrees of freedom \( N_\varepsilon \) in order to obtain \( \| \mathbf{v} - \mathbf{v}_N \|_{\ell_2} \leq \varepsilon \) is bounded by \( \mathcal{N}_\varepsilon \leq \varepsilon^{-1/s} \| \mathbf{v} \|_{A^s}^{1/s} \). It is important to remark that this bound on \( \mathcal{N}_\varepsilon \) is usually sharp (see [Ste09] (3)).

4.2. Quasi-optimal algorithms. Let us now assume that the solution is \( \mathbf{u} \in A^s \) for some \( s > 0 \) and that we want to approximate it with a target tolerance \( \varepsilon > 0 \). The benchmark is given by a best \( \mathcal{N}_\varepsilon \)-term approximation \( \mathbf{u}_{\mathcal{N}_\varepsilon} \) satisfying \( \sigma_{\mathcal{N}_\varepsilon}(\mathbf{u}) = \| \mathbf{u} - \mathbf{u}_{\mathcal{N}_\varepsilon} \|_{\ell_2(\mathcal{F})} \leq \varepsilon \), which is, however, in general not computable. So, we need to focus on the computation of a quasi-optimal approximation \( \mathbf{u}_\varepsilon \):

- (O1) Convergence rate: \( \| \mathbf{u} - \mathbf{u}_\varepsilon \|_{\ell_2(\mathcal{F})} \leq \varepsilon \) and \( \# \supp \mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s} \| \mathbf{u} \|_{A^s}^{1/s} \).
- (O2) Computational work: The number of operations required for the computation of \( \mathbf{u}_\varepsilon \) is of order \( \mathcal{O}(\varepsilon^{-1/s} \| \mathbf{u} \|_{A^s}^{1/s}) \); i.e., for any \( \varepsilon > 0 \), \( \mathbf{u}_\varepsilon \) can be computed within linear complexity, recalling that \( \mathcal{N}_\varepsilon \lesssim \varepsilon^{-1/s} \| \mathbf{u} \|_{A^s}^{1/s} \).

In order to realize (O2) we require the wavelet bases \( \hat{\Psi}^\mathcal{X} \) and \( \hat{\Psi}^\mathcal{Y} \) to be of tensor product type, which will be the topic of the next section.

5. Tensor product wavelet bases

Recall that \( \mathcal{X} \) and \( \mathcal{Y} \) can be characterized as follows (see [GO95]):

\[
\mathcal{X} \simeq [L_2(0, T) \otimes V] \cap [H^1_{\text{per}}(0, T) \otimes V'], \quad \mathcal{Y} \simeq L_2(0, T) \otimes V.
\]

Furthermore, by the definition of \( V \) in (2.3), the construction of \( \hat{\Psi}^\mathcal{X} \) and \( \hat{\Psi}^\mathcal{Y} \) can be obtained by tensorization of univariate wavelet bases.

5.1. Uniformly local, piecewise polynomial wavelet bases. Let us consider a univariate Sobolev space \( \mathcal{H} \in \{ H^1_{\text{per}}(0, T), V^{(1)}, \ldots, V^{(n)} \} \) with \( V^{(i)} \subset L_2(\Omega_i) \) and a univariate wavelet basis \( \Psi \) for \( L_2(\Omega) \) where \( \Omega \subset \mathbb{R} \) is either \( (0, T) \) (if \( \mathcal{H} = H^1_{\text{per}}(0, T) \)) or \( \Omega_i \) (if \( \mathcal{H} = V^{(i)} \), recall \( \Omega_i \subset \mathbb{R} \), i.e., w.l.o.g. \( \Omega_i = (0, 1) \)),

\[
\Psi = \bigcup_{j \in \mathbb{N}_0} \Psi_j = \{ \psi_\lambda : \lambda = (j, k) \in \mathcal{J} \} \subset \mathcal{H},
\]

as well as \( \Psi_j := \{ \psi_\lambda : \lambda \in \mathcal{J}_j \} \) and \( \mathcal{J}_j := \{ \lambda \in \mathcal{J} : |\lambda| = j \} \). Here, \( |\lambda| := j \geq 0 \) denotes the level (steering the diameter of the support of \( \psi_{j,k} \) in the sense that \( \text{diam}(\supp \psi_{j,k}) \approx 2^{-j} \)) and \( k \) is a translation index indicating the position of \( \supp \psi_{j,k} \). Note that the elements of \( \Psi_0 \) are not wavelets but scaling functions. For details on wavelets on the interval, we refer e.g. to [Urb09]. By the Wavelet Characterization Theorem [Dah97], if the elements of \( \Psi \) (and also those of the unique dual wavelet basis) are sufficiently smooth, the properly normalized collections \( \{ \psi_\lambda / \| \psi_\lambda \|_{\mathcal{H}} : \lambda \in \mathcal{J} \}, \{ \psi_\lambda / \| \psi_\lambda \|_{\mathcal{H}'} : \lambda \in \mathcal{J} \} \) are Riesz bases for the Sobolev spaces \( \mathcal{H} \) and \( \mathcal{H}' \), respectively. Besides that, we shall assume that \( \Psi \) is a uniformly local, piecewise polynomial wavelet basis of order \( d \in \mathbb{N} \), i.e.:

- (W1) Local supports: \( \text{diam}(\supp \psi_\lambda) \approx 2^{-|\lambda|} \) for all \( \lambda \in \mathcal{J} \).
(W2) Level-wise finite number of overlaps: There exists $C \in \mathbb{N}$ independent of $j \in \mathbb{N}_0$ such that $\sup_{\lambda \in J_j} \#\{ \lambda' \in J_j : |\text{supp } \psi_{\lambda} \cap \text{supp } \psi_{\lambda'}| > 0 \} \leq C$.

(W3) Piecewise polynomials: For all $\lambda \in J$, $\psi_{\lambda}$ is a piecewise polynomial of maximum degree $d - 1$ and has $d$ vanishing moments (except for scaling functions and few boundary adapted wavelets).

Furthermore, we assume that the projection $Q_j[v] := \sum_{\lambda \in J} v_{\lambda} \psi_{\lambda}$ for $v = \sum_{\lambda \in J} v_{\lambda} \psi_{\lambda}$ satisfies the Jackson estimates $\|\text{Id} - Q_j\|_{H^{d}(\Omega) \cap H \rightarrow L_2(\Omega)} \lesssim 2^{-d_j}$, $\|\text{Id} - Q_j\|_{H^{d}(\Omega) \cap H \rightarrow H^s} \lesssim 2^{-(d-m)j}$, where $m = 1$ if $H = H^1_{\text{per}}(0,T)$.

5.2. Temporal discretization. In order to ensure the periodic boundary conditions in time (see (2.1)) in $\mathcal{X}$ we need a (univariate) periodic wavelet basis

$$\Theta_{\text{per}} := \{ \theta_{\chi}^{\text{per}} : \chi \in J_{t}^{\text{per}} \} \subset H^1_{\text{per}}(0,T)$$

being a uniformly local, piecewise polynomial wavelet basis of order $d_t \in \mathbb{N}$ (the index $t$ stands for ‘time’) for $L_2(0,T)$ with associated Riesz constants $c_{L_2}(\Theta_{\text{per}})$, $C_{L_2}(\Theta_{\text{per}})$. We assume that the elements of $\Theta_{\text{per}}$ are sufficiently smooth so that the properly normalized collection $\{ \theta_{\chi}^{\text{per}}/\|\theta_{\chi}^{\text{per}}\|_{H^1} : \chi \in J_{t}^{\text{per}} \}$ is a Riesz basis for $H^1_{\text{per}}(0,T)$ with constants $c_{H^1_{\text{per}}}(\Theta_{\text{per}})$, $C_{H^1_{\text{per}}}(\Theta_{\text{per}})$. Recall that the construction of periodic wavelet bases is particularly easy [Urb09]. For the temporal part of the test space $\mathcal{Y}$ (involving also non-periodic functions), we consider a uniformly local, piecewise polynomial wavelet basis for $L_2(0,T)$,

$$\Theta := \{ \theta_{\lambda} : \lambda \in J_{t} \},$$

with Riesz constants $c_{L_2}(\Theta)$, $C_{L_2}(\Theta)$ and wavelets being not necessarily periodic.

5.3. Spatial discretization. For the spatial discretization, we use the fact that $\Omega = \Omega_1 \times \cdots \times \Omega_n$ is a product domain. Here, we shall use that $V$ is the (intersection of) tensor products of univariate Sobolev spaces (see (2.3)) with $L_2(\Omega) \subseteq V$ and $L_2(\Omega) \cong L_2(\Omega_1) \otimes \cdots \otimes L_2(\Omega_n)$ (see e.g. [GO95]). We assume that for $i \in \{1, \ldots, n\}$ we are given univariate uniformly local, piecewise polynomial wavelet bases of order $d_x \in \mathbb{N}$ (the index $x$ indicating the spatial variable) for $L_2(\Omega_i)$, $\Sigma^{(i)} := \{ \sigma^{(i)}_{\lambda} : \lambda \in J^{(i)} \} \subset V^{(i)}$. We require that these functions are sufficiently smooth so that $\{ \sigma^{(i)}_{\lambda}/\|\sigma^{(i)}_{\lambda}\|_{V^{(i)}} : \lambda \in J^{(i)} \}$, $\{ \sigma^{(i)}_{\lambda}/\|\sigma^{(i)}_{\lambda}\|_{V^{(i)}} : \lambda \in J^{(i)} \}$ are Riesz bases for $V^{(i)}$, $V^{(i)\bot}$ with constants $c_{V^{(i)}}(\Sigma^{(i)})$, $c_{V^{(i)}}(\Sigma^{(i)})$ and $c_{V^{(i)\bot}}(\Sigma^{(i)})$. Now,

$$\Sigma := \{ \sigma_{\lambda} : \lambda \in J_{x} \} := \Sigma^{(1)} \otimes \cdots \otimes \Sigma^{(n)}$$

is a Riesz basis for $L_2(\Omega)$ where $\sigma_{\lambda} := \sigma_{\lambda_1}^{(1)} \otimes \cdots \otimes \sigma_{\lambda_n}^{(n)}$ is a tensor product wavelet and $J_{x} := J^{(1)} \times \cdots \times J^{(n)}$ [Dij09] Lemma 3.1.7. Moreover,

$$\Sigma^V := \{ \sigma_{\lambda}/\|\sigma_{\lambda}\|_V : \lambda \in J_{x} \}, \quad \Sigma^{V\bot} := \{ \sigma_{\lambda}/\|\sigma_{\lambda}\|_{V\bot} : \lambda \in J_{x} \}$$

are Riesz bases for $V$, $V^{\bot}$ [Dij09] Lemma 3.1.8. The associated Riesz constants will be denoted by $c_{V}(\Sigma)$, $c_{V}(\Sigma)$, $c_{V}(\Sigma)$ and $c_{V}(\Sigma)$. 

A SPACE-TIME WAVELET METHOD FOR TIME-PERIODIC PDES 1315
5.4. **Space-time discretization.** We are now in a position to define the Riesz wavelet bases $\hat{\Psi}^X$ and $\hat{\Psi}^Y$ from (5.2). With $L_2(0, T; L_2(\Omega)) \approx L_2(0, T) \otimes L_2(\Omega)$, (5.7) $\hat{\Psi} := \{ \hat{\psi}_\lambda := \theta^\per_{\lambda_t} \otimes \sigma_{\lambda_x} : \lambda := (\lambda_t, \lambda_x) \in \tilde{\mathcal{J}} := \mathcal{J}^\per_t \times \mathcal{J}_x \} = \Theta^\per \otimes \Sigma$, (5.8) $\tilde{\Psi} := \{ \tilde{\psi}_\lambda := \vartheta_{\lambda_t} \otimes \sigma_{\lambda_x} : \lambda := (\lambda_t, \lambda_x) \in \tilde{\mathcal{J}} := \mathcal{J}_t \times \mathcal{J}_x \} = \Theta \otimes \Sigma$

are both Riesz bases for $L_2(0, T; L_2(\Omega))$. At this point, we only need to normalize the above Riesz bases appropriately (see [GO95] Propositions 1 and 2) so that (5.9) $\hat{\Psi}^X := \{ \hat{\psi}_\lambda / \| \hat{\psi}_\lambda \|_X : \lambda \in \tilde{\mathcal{J}} \} = D^X \hat{\Psi}$, $D^X := \text{diag} \left( \{ \| \hat{\psi}_\lambda \|^{-1}_X \}_{\lambda \in \tilde{\mathcal{J}}} \right)$, (5.10) $\tilde{\Psi}^Y := \{ \tilde{\psi}_\lambda / \| \tilde{\psi}_\lambda \|_Y : \lambda \in \tilde{\mathcal{J}} \} = D^Y \tilde{\Psi}$, $D^Y := \text{diag} \left( \{ \| \tilde{\psi}_\lambda \|^{-1}_Y \}_{\lambda \in \tilde{\mathcal{J}}} \right)$

are Riesz bases for $X$, respectively $Y$ (compare [SS09] Section 6).

**Remark 5.1.** We shall denote a tensor product wavelet basis $\Psi \in \{ \hat{\Psi}, \tilde{\Psi} \}$ as follows:

$\Psi = \Psi^{(0)} \otimes \Psi^{(1)} \otimes \cdots \otimes \Psi^{(n)} = \{ \psi_{\lambda} := \psi^{(0)}_{\lambda_0} \otimes \psi^{(1)}_{\lambda_1} \otimes \cdots \otimes \psi^{(n)}_{\lambda_n} : \lambda \in \mathcal{J} \}$,

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_n)$ and $\mathcal{J} := \mathcal{J}^{(0)} \times \mathcal{J}^{(1)} \times \cdots \times \mathcal{J}^{(n)}$. In this setting, it is clear that $\Psi^{(0)} \in \{ \Theta^\per, \Theta \}$, $\mathcal{J}^{(i)} \subset \mathcal{J}^{(i)}_t \times \mathcal{J}_x$ and $\Psi^{(i)} = \Sigma^{(i)}$ for $i \in \{ 1, \ldots, n \}$.

5.5. **Riesz constants for test and trial bases.** For the implementation of an AWGM, we need estimates for the Riesz constants $c_X(\hat{\Psi}), C_X(\hat{\Psi}), c_Y(\tilde{\Psi}), C_Y(\tilde{\Psi})$ in (3.2). Again, we use that $X$ and $Y$ are (intersections of) tensor products of Hilbert spaces. As in [SS09] §6, we have the following estimates for $\hat{\Psi}^X$ and $\tilde{\Psi}^Y$:

(5.11) $c_X(\hat{\Psi}) \geq \min \left\{ c_{L_2}(\Theta^\per \cdot C_\Sigma), c_{H^\per_{\per}}(\Theta^\per) \cdot c_{V^\per}(\Sigma) \right\}$, (5.12) $C_X(\hat{\Psi}) \leq \min \left\{ C_{L_2}(\Theta^\per \cdot C_\Sigma), C_{H^\per_{\per}}(\Theta^\per) \cdot C_{V^\per}(\Sigma) \right\}$, (5.13) $c_Y(\tilde{\Psi}) \geq c_{L_2}(\Theta) \cdot c_{V^\per}(\Sigma), C_Y(\tilde{\Psi}) \leq C_{L_2}(\Theta) \cdot C_{V^\per}(\Sigma)$.

The Riesz constants $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$ can also be bounded by those of the 1D bases $\Sigma^{(i)}, i \in \{ 1, \ldots, n \}$. Using (2.3), it can be shown as in [DSS09] §2, that (5.14) $c_{V^\per}(\Sigma) \geq \min_{m \in \{ 1, \ldots, n \}} \min \left\{ c_{L_2}(\Sigma^{(m)}), c_{V^\per}(\Sigma^{(m)}) \right\} \prod_{k \neq m} c_{L_2}(\Sigma^{(k)})$, (5.15) $C_{V^\per}(\Sigma) \leq \max_{m \in \{ 1, \ldots, n \}} \max \left\{ C_{L_2}(\Sigma^{(m)}), C_{V^\per}(\Sigma^{(m)}) \right\} \prod_{k \neq m} C_{L_2}(\Sigma^{(k)})$.

Unfortunately, the same approach does not apply to the (dual) Riesz constants $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$ of $\Sigma^V$ in (5.6). However, one may consider $\tilde{\Sigma}^V$ being the unique Riesz basis for $V$ that is dual to $\Sigma^V$, i.e., $(\Sigma^V, \Sigma^V)^* \Sigma^V = \text{Id}$. Denoting by $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$ the associated Riesz constants, it can be shown that $C_{V^\per}(\Sigma)^{-1} \leq c_{V^\per}(\Sigma)$ and $C_{V^\per}(\Sigma) \leq C_{V^\per}(\Sigma)^{-1}$. Observe that for computing bounds for $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$, we may proceed as for bounding $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$. We conclude that for the computation of the bounds in (5.11), (5.12) and (5.13), it is sufficient to compute bounds for univariate Riesz constants which can be easily approximated (e.g. [DI09] §2).

**Remark 5.2.** Recalling the construction of wavelets, note that the numerical approximation of $c_{V^\per}(\Sigma), C_{V^\per}(\Sigma)$ may be difficult since the the dual basis $\Sigma^V$ (and their derivatives) may not be available in a closed form. If sharp bounds are needed, one may use an $L_2(\Omega)$-orthonormal basis $\Sigma$ so that $\tilde{\Sigma}^V = \Sigma^V$, e.g. multiwavelets.
5.6. Best approximation rates. We need to know for which values of $s$ the solution $u$ of (6.3) is in $\mathcal{A}^s$. More precisely, for a fixed trial basis $\hat{\Psi}^\mathcal{X}$, the question is what is the largest value $s_{\text{max}}$ of $s$ for which $u \in \mathcal{A}^s$ can be expected and that cannot be increased by imposing higher smoothness conditions on $u$ (excluding special cases where $u$ is (close to) a finite vector). This value $s_{\text{max}}$ is referred to as the best possible approximation rate. For our setting, we may apply the results from [SS09, §7.2]. With $u = u^T \hat{\Psi}^\mathcal{X} \in \mathcal{X} \cap H^{d_t}(0,T) \otimes \mathcal{H}^{d_z}(\Omega)$ and the Sobolev space

$$\mathcal{H}^{d_z}(\Omega) := \bigcap_{i=1}^n \bigotimes_{j=1}^n Z_{ij},$$

where $Z_{ij} := \{ L_2(\Omega_i), \quad i \neq j, \quad H^{d_z}(\Omega_i), \quad i = j \}$

of dominating mixed derivatives, the best possible rate is given by

(5.16)

$$s_{\text{max}} = \min\{d_t - 1, d_z - m\}.$$

We recall that $d_t$ denotes the polynomial order of $\Theta^{\text{per}}$ and $d_z$ those of $\Sigma^{(1)}, \ldots, \Sigma^{(n)}$. This rate does not depend on the spatial dimension $n$. Moreover, we remark that $u \in \mathcal{H}^{d_z}(\Omega)$ is sufficient but not necessary for obtaining the above rate. In fact, the Sobolev space $H^{d_t}(0,T) \otimes \mathcal{H}^{d_z}(\Omega)$ can be replaced by a (weaker) Besov space of dominating mixed derivatives [Nit06, SU09]. Note that the order of the wavelet bases for the test space $\mathcal{Y}$ does not enter the best approximation rate.

6. Adaptive wavelet Galerkin methods

An infinite $\ell_2$-problem (3.3) arising from a wavelet discretization of (2.8) can be solved by an AWGM, e.g. [CDD01, GHS07]. We now first present the main idea of an AWGM for the solution of an (for convenience) elliptic operator problem. Secondly, we highlight the additional challenges related to parabolic problems and indicate a possible way-out using normal equations.

6.1. Elliptic operator problems. Solely for explanation purposes, we consider elliptic operator problems of the following type. For a linear, self-adjoint operator $\mathcal{C} \in \mathcal{L}(\mathcal{X}, \mathcal{X'})$ induced by a continuous and coercive bilinear form (i.e., $\langle v, \mathcal{C}[w] \rangle_{\mathcal{X} \times \mathcal{X'}} \lesssim \|v\|_{\mathcal{X}} \|w\|_{\mathcal{X'}}$, $\langle v, \mathcal{C}[v] \rangle_{\mathcal{X} \times \mathcal{X'}} \gtrsim \|v\|_{\mathcal{X'}}^2$ for all $v, w \in \mathcal{X}$), we consider:

(6.1) Find $u \in \mathcal{X}$: $\mathcal{C}[u] = g$, $g \in \mathcal{X'}$.

Analogously to (3.3), the equivalent $\ell_2$-problem to this problem reads

(6.2) Find $u \in \ell_2(\hat{\mathcal{J}})$: $\mathcal{C}u = g$, $g \in \ell_2(\hat{\mathcal{J}})$,

where $\mathcal{C} = \langle \hat{\Psi}^\mathcal{X}, \mathcal{C}[\hat{\Psi}^\mathcal{X}] \rangle_{\mathcal{X} \times \mathcal{X'}}$ and $g = \langle \hat{\Psi}^\mathcal{X}, g \rangle_{\mathcal{X} \times \mathcal{X'}}$ with $\hat{\Psi}^\mathcal{X}$ from (5.9). In the elliptic case, i.e., $\mathcal{X} = \mathcal{Y}$ we may use $\hat{\Psi}^\mathcal{X}$ as trial and test basis. Furthermore, $\mathcal{C}$ is s.p.d. and $\| \cdot \|^2 := \langle \cdot, \cdot \rangle_{\ell_2(\hat{\mathcal{J}}) \times \ell_2(\hat{\mathcal{J}})}$ defines an equivalent norm [Ste09, p. 565]

(6.3)

$$\| \mathcal{C}^{-1/2}v \|_{\ell_2} \leq \|v\| \leq \| \mathcal{C}^{1/2}v \|_{\ell_2}, \quad \forall v \in \ell_2(\hat{\mathcal{J}}).$$

The idea of an AWGM for (6.2) is outlined in an (idealized) Algorithm [Ste09, p. 567]. Within this algorithm, we make some non-realistic assumptions, which will be discussed below. Abandoning these assumptions will then give rise to the realizable AWGM variants introduced in later sections. Starting from an initial index set $\hat{\Lambda}_1 \subset \hat{\mathcal{J}}$, a sequence of nested finite index sets $(\hat{\Lambda}_k)_k$ is computed. On
each such \( \hat{A}_k \), a Galerkin problem is solved that yields the (finite) vector \( u_{\hat{A}_k} \). Due to the Riesz basis property, it holds that (see also (6.1))

\[
c_X(\hat{\Psi})^\frac{1}{2} \| u - u_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \leq \| u - u_{\hat{A}_k}^T \hat{\Psi}^X \|_{\mathcal{V}} \leq C_X(\hat{\Psi})^\frac{1}{2} \| u - u_{\hat{A}_k} \|_{\ell^2(\mathcal{J})}.
\]

Given \( u_{\hat{A}_k} \), the computation of the next \( \hat{A}_{k+1} \) is based on the infinitely supported residual \( g - Cu_{\hat{A}_k} \in \ell^2(\mathcal{J}) \) and the error estimator \( \| g - Cu_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \) which satisfies

\[
\| C \|^{-1} \| g - Cu_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \leq \| u - u_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \leq \| C \|^{-1} \| g - Cu_{\hat{A}_k} \|_{\ell^2(\mathcal{J})}.
\]

This also explains the stopping criterion in line 4 of Algorithm 1. Consequently, indices corresponding to the largest entries in the residual are added to \( A_k \). This so-called bulk-chasing process is steered by the parameter \( \delta \).

---

**Algorithm 1** \([u_\varepsilon] = \text{IDEALIZED–AWGM}[\varepsilon, A_1] \)

**Input:** Target tolerance \( \varepsilon \) and an index set \( \hat{A}_1 \neq \emptyset \).

**Parameter:** \( \delta \in (0, \kappa(C)^{-\frac{1}{2}}) \).

1. for \( k = 1, 2, \ldots \) do
2. Solve the Galerkin problem:

\[
(6.5) \quad \text{Find } u_{\hat{A}_k} \in \ell^2(\hat{A}_k) : \quad \hat{A}_k C_{\hat{A}_k} u_{\hat{A}_k} = g_{\hat{A}_k}, \quad g_{\hat{A}_k} := R_{\hat{A}_k} g \in \ell^2(\hat{A}_k).
\]
3. Residual computation: Compute \( g - Cu_{\hat{A}_k} \) and \( \nu_k := \| g - Cu_{\hat{A}_k} \|_{\ell^2} \).
4. if \( \nu_k \leq \varepsilon/\| C \|^{-1} \) then return \( u_\varepsilon := u_{\hat{A}_k} \).
5. Bulk chasing criterion: Find smallest index set \( \hat{A}_{k+1} \supset \hat{A}_k \) such that

\[
(6.6) \quad \| R_{\hat{A}_{k+1}} (g - Cu_{\hat{A}_k}) \|_{\ell^2(\hat{A}_{k+1})} \geq \delta \| g - Cu_{\hat{A}_k} \|_{\ell^2(\mathcal{J})}.
\]
6. end for

---

**Proposition 6.1** ([Ste09 Proposition 4.1]). The iterates \( u_{\hat{A}_k} \) produced by Algorithm 1 satisfy \( \| u - u_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \leq [1 - \delta^2 \kappa(C)^{-1}]^{k/2} \| u \| \). For the output \( u_\varepsilon \) it holds that \( \| u - u_\varepsilon \|_{\ell^2(\mathcal{J})} \leq \varepsilon \). If \( u \in A^s \) for some \( s > 0 \), it also holds for \( N_k := \# \hat{A}_k \) that

\[
(6.7) \quad \| u - u_{\hat{A}_k} \|_{\ell^2(\mathcal{J})} \lesssim \| u \|_{A^s}^{1/s} N_k^{-s}, \quad \# \text{supp } u_\varepsilon \lesssim \varepsilon^{-1/s} \| u \|_{A^s}^{1/s}.
\]

**Remark 6.2.** Algorithm 1 cannot be implemented as the residual cannot be computed exactly in general. Implementable versions are given in CDD01, GHS07. The algorithm in CDD01 requires an additional thresholding and thus can be expected to be less efficient than GHS07. The adaptive wavelet method in CDD02 relies on an inexact Richardson iteration that is applied directly to (6.2) without Galerkin projection. However, as shown in GHS07, this scheme can be expected to be less efficient than GHS07. Thus, we shall focus on GHS07 here.

---

**6.2. Parabolic problems.** One may try to analyze IDEALIZED–AWGM for \( Bu = f \) in (6.3). However, the generalization of the idealized scheme to (6.3) is not trivial: (1) **Symmetry and positive definiteness.** Recall that \( B \) from (6.3) is not s.p.d., so that \( \langle \cdot, B \cdot \rangle_{\ell^2(\mathcal{J}) \times \ell^2(\mathcal{J})} \) is not an equivalent norm on \( \ell^2(\mathcal{J}) \). However, the availability of an equivalent energy norm as in (6.3) is crucial for the convergence analysis of Algorithm 1 (see Ste09 Proposition 4.1). (2) **Bulk chasing and residual
computation. It is not clear how to construct $\hat{\Lambda}_{k+1}$ from $\hat{\Lambda}_k$. In analogy to (6.4), the residual $f - B\hat{u}_{\hat{\Lambda}_k} \in \ell_2(\tilde{\mathcal{J}})$ with error estimator $\|f - B\hat{u}_{\hat{\Lambda}_k}\|_{\ell_2(\mathcal{J})}$ satisfies

$$\|B\|^{-1}\|f - B\hat{u}_{\hat{\Lambda}_k}\|_{\ell_2(\mathcal{J})} \leq \|u - u_{\hat{\Lambda}_k}\|_{\ell_2(\mathcal{J})} \leq \|B\|^{-1}\|f - B\hat{u}_{\hat{\Lambda}_k}\|_{\ell_2(\mathcal{J})}.$$  

But the residual is an element of $\ell_2(\tilde{\mathcal{J}})$, $\tilde{\mathcal{J}} \neq \mathcal{J}$. Thus, we cannot compute $\hat{\Lambda}_k$ by selecting some contributions from the residual as in (6.6). (3) Petrov-Galerkin problems. Since $\hat{\Psi}^\top \neq \hat{\Psi}^\dagger$, the (well-posed) Galerkin problem in line 5 of Algorithm 1 here becomes a Petrov-Galerkin problem. Thus the uniform well-posedness of the finite-dimensional problems is no longer inherited from the infinite-dimensional problem (3.3) and has to be ensured for all $\hat{\Lambda}_k$.

Hence, we focus on the associated normal equations, as proposed in [CDD02]:

(6.9) Find $u \in \ell_2(\tilde{\mathcal{J}})$: $B^\top Bu = B^\top f$, $B^\top f \in \ell_2(\tilde{\mathcal{J}})$.

Since $B$ is boundedly invertible, the unique solution of (6.9) is also the unique solution of (3.3) (see [CDD02 Thm. 7.1]). Indeed, (6.9) are the (infinite) normal equations associated to the least squares problem (compare [CDD02 §7]) of finding $u \in \ell_2(\tilde{\mathcal{J}})$ such that $u = \arg\min_{v \in \ell_2(\tilde{\mathcal{J}})} \|Bv - f\|_{\ell_2(\mathcal{J})}^2$ for given $f \in \ell_2(\tilde{\mathcal{J}})$. We anticipate that one does not expect the usually dramatic effect of a squared condition number for $B^\top B$ since $B$ is wavelet-preconditioned; see below.

6.3. AWGMs for normal equations. Now we investigate if the reformulation of (3.3) in terms of (6.9) addresses the issues mentioned in Section 6.2.

(1) Symmetry and positive definiteness. Obviously, $B^\top B$ is symmetric. Moreover, by (3.4), it is also positive definite and it holds that

$$\|B^\top B\| \leq \|B\|^2, \quad \|(B^\top B)^{-1}\| \leq \|B^{-1}\|^2,$$

hence $\kappa(B^\top B) \leq \|B\|^2 \|B^{-1}\|^2$. Thus, we consider $Cu = g$ with $C = B^\top B$, $g = B^\top f$ and $\|\cdot\|^2 := (\cdot, B^\top B \cdot)$ and use Algorithm 1.

(2) Bulk chasing and residual computation. Instead of considering the residual in $\ell_2(\tilde{\mathcal{J}})$, we now obtain $B^\top (f - B\hat{u}_{\hat{\Lambda}_k}) \in \ell_2(\tilde{\mathcal{J}})$ with error estimator $\rho_k := \|B^\top (f - B\hat{u}_{\hat{\Lambda}_k})\|_{\ell_2(\mathcal{J})}$. In analogy to (6.4) and (6.8), we infer that

$$\|B\|^{-2}\rho_k \leq \|u - u_{\hat{\Lambda}_k}\|_{\ell_2(\tilde{\mathcal{J}})} \leq \|B^{-1}\|^2 \rho_k.$$  

In this setting, the residual $f - B\hat{u}_{\hat{\Lambda}_k}$ from (6.8) is also referred to as primal residual, whereas $B^\top (f - B\hat{u}_{\hat{\Lambda}_k})$ is called dual residual. Observe that this kind of residual allows for a bulk chasing strategy as used in line 5 of IDEALIZED−AWGM.

(3) Well-posedness. With $\hat{\Lambda} B^\top$ and $\hat{\Lambda}$ defined in (6.6), we get $(B^\top B)|_{\hat{\Lambda}\times\hat{\Lambda}} = \hat{\Lambda} B^\top B \hat{\Lambda}$ so that (6.5) for general $\hat{\Lambda} \subset \tilde{\mathcal{J}}$ with $C = B^\top B$ now reads as follows:

(6.12) Find $u_{\hat{\Lambda}} \in \ell_2(\hat{\Lambda})$: $\hat{\Lambda} B^\top B \hat{\Lambda} u_{\hat{\Lambda}} = \hat{\Lambda} B^\top f$, $\hat{\Lambda} B^\top f \in \ell_2(\hat{\Lambda})$.

Observe that the unique solution $u_{\hat{\Lambda}} = \arg\min_{v \in \ell_2(\hat{\Lambda})} \|B\hat{\Lambda}^\top B \hat{\Lambda} f - f\|_{\ell_2(\mathcal{J})}^2$ to (6.12) can also be characterized as the solution of a least-squares problem. Moreover, the Galerkin problem (6.12) is uniformly well-posed. Since $\hat{\Lambda} B^\top B$ is s.p.d., we infer from (6.10) that $\|\hat{\Lambda} B^\top B \hat{\Lambda}\| \leq \|B\|^2$ as well as $\|\hat{\Lambda} B^\top B \hat{\Lambda}^{-1}\| \leq \|B^{-1}\|^2$ for all $\hat{\Lambda} \subset \tilde{\mathcal{J}}$. In particular, the condition number $\kappa(\hat{\Lambda} B^\top B \hat{\Lambda})$ is bounded independently of $\hat{\Lambda}$.
Remark 6.3. Obviously, neither the residual in (6.11) nor the solution \( u_{\hat{A}_k} \) of (6.12) can be computed exactly since the involved matrices are of infinite dimension. In order to obtain an implementable scheme, we work with an approximation \( w_{\hat{A}_k} \) to \( u_{\hat{A}_k} \) and an approximate residual \( \tilde{r}_k \) to \( B^\top (f - Bu_{\hat{A}_k}) \). This will be discussed next.

7. An implementable space-time adaptive wavelet Galerkin method

Now we describe the quasi-optimal (in terms of (O1) and (O2)) AWGM for the numerical solution of (6.9) and call it LS–AWGM (least squares adaptive wavelet Galerkin method); see Algorithm 2. We first describe the required subroutines. We assume that \( u \in A^* \) and denote by \( w_{\hat{A}} \) an approximate solution to (6.12).

\[
(7.1) \quad \| B^\top (f - Bw_{\hat{A}}) - \tilde{r} \|_{\ell_2(\hat{J})} \leq \omega_{1s} \cdot \nu, \quad \nu := \| \tilde{r} \|_{\ell_2(\hat{J})},
\]

and the associated computational cost is of order \( O(\#\hat{A} + \nu^{-1/s}\| u \|_{A^*}^{1/s}) \).

\[
(7.2) \quad \| B^\top (f - Bw_{\hat{A}}) \|_{\ell_2(\hat{A})} \leq \gamma_{1s} \cdot \nu,
\]

where \( \nu \) is defined in (7.1) and the associated computational cost is of order \( O(\#\hat{A} + \nu^{-1/s}\| u \|_{A^*}^{1/s}) \). Moreover, we assume that we are given an initial value \( \hat{w}_{\hat{A}} \) satisfying \( \| B^\top (f - B\hat{w}_{\hat{A}}) \|_{\ell_2(\hat{A})} \leq (1 + \gamma_{1s}) \cdot \nu \).

\[
(7.3) \quad \hat{A} \subset \hat{A}, \quad \| R_{\hat{A}} \hat{r} \|_{\ell_2(\hat{A})} \geq \delta \| \hat{r} \|_{\ell_2(\hat{J})},
\]

and, up to some absolute multiple, \( \hat{A} \) is minimal among all sets that satisfy (7.3). The computational cost of this routine is of order \( O(\#\hat{A} + \# \text{supp} \hat{r}) \).

Algorithm 2 [\( u_\varepsilon = \text{LS–AWGM}[\varepsilon, \hat{A}_1, \nu_0] \)]

**Input:** Target tolerance \( \varepsilon \), finite index set \( \hat{A}_1 \subset \hat{J} \) and tolerance \( \nu_0 \approx \| B^\top f \|_{\ell_2(\hat{J})} \).

**Parameters:** \( \delta, \omega_{ls}, \gamma_{ls} \) with \( \omega_{ls} \in (0, \delta), \frac{\delta + \omega_{ls}}{1 - \omega_{ls}} < \kappa(B^\top B)^{-1/2} \),

\[ \gamma_{ls} \in (0, \frac{1 - \omega_{ls}(\delta - \omega_{ls})}{1 + \omega_{ls}} \kappa(B^\top B)^{-1}) \].

1: Set \( w_{\hat{A}_0} := 0 \).
2: for \( k = 1, 2, \ldots \) do
3: \( \hat{w}_{\hat{A}_k} := \text{GALSOLVE}[\hat{A}_k, w_{\hat{A}_{k-1}}, \gamma_{ls} \cdot \nu_{k-1}] \).
4: \( \hat{r}_k := \text{RESIDUAL}[\hat{w}_{\hat{A}_k}, \omega_{ls}] \) and set \( \nu_k := \| \hat{r}_k \|_{\ell_2} \).
5: if \( \nu_k \leq \varepsilon / \| B^{-1} \|^2 \), then return \( u_\varepsilon := w_{\hat{A}_k} \).
6: \( \hat{A}_{k+1} := \text{EXPAND}[\hat{A}_k, \hat{r}_k, \delta] \).
7: end for

In analogy to Proposition 6.1, we have the following result for LS–AWGM, which is a direct consequence of [Ste09, Proposition 4.2 and Theorem 4.1]: 
Theorem 7.1 (GHS07, Ste09). Let the assumptions on [RES], [GAL] and [EXP] and the requirements on $\delta, w_k, \gamma_k$ from Algorithm 2 hold. Then, the iterates $w_{\hat{A}_k}$ produced by LS–AWGM satisfy $\|u - w_{\hat{A}_k}\| \leq \rho^{k/2}\|u\|$ where $\rho := 1 - \left(\frac{\delta - w_k}{1 + w_k}\right)\kappa(B^\top B)^{-1} + \frac{\gamma_k^2}{(1 + w_k)^2}\kappa(B^\top B) < 1$ and the output $u_e$ satisfies $\|u - u_e\|_{\ell_2(\mathcal{J})} \leq \varepsilon$. If, moreover, $u \in A^s$ for some $s > 0$, it holds for $N_k := \#\hat{A}_k$ that

$$\|u - w_{\hat{A}_k}\|_{\ell_2(\mathcal{J})} \lesssim \|u\|^s A^s N_k^{-s}, \quad \#\supp u_e \lesssim \varepsilon^{-1/s}\|u\|_{A^s}^s. \tag{7.4}$$

Note that a realization of EXPAND can easily be obtained by an approximate sorting of $\hat{r}$ and a subsequent thresholding (e.g., Ste09, p. 569). Possible realizations of the routines RESIDUAL and GAL that are based on so-called APPLY-routines (i.e., an adaptive, column-wise approximation of $B$ and $B^\top$) have been discussed in SS09, Ste09. We shall focus on a multitree approach which has been shown to outperform APPLY-based AWGMs in elliptic settings (see KS13).

8. A multitree implementation

8.1. Tree and multitree structured index sets. Let $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\}$ be a univariate uniformly local, piecewise polynomial wavelet basis as in (5.2).

Definition 8.1. A subset $\Lambda \subset \mathcal{J}$ is called a tree if for any $\lambda \in \Lambda$ with $|\lambda| > 0$ it holds that $\supp \psi_\lambda \subset \bigcup_{\mu \in \Lambda, |\mu| = |\lambda| - 1} \supp \psi_\mu$.

It holds for all $\lambda, \mu \in \Lambda$ with $|\mu| = |\lambda| - 1$ and $\supp \psi_\lambda \cap \supp \psi_\mu > 0$ that

$$S_\mu \supset S_\lambda,$$

where

$$S_\mu := \{x \in \Omega : \text{dist}(x, \supp \psi_\mu) \leq D_\Psi 2^{-|\mu|}\}, \quad D_\Psi := \sup_{\lambda \in \mathcal{J}} 2^{|\lambda|} \text{diam}(\supp \psi_\lambda). \tag{8.1}$$

Let us now consider a tensor product wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\} \in \{\tilde{\Psi}, \tilde{\Psi}\}$ with $\tilde{\Psi}$ and $\tilde{\Psi}$ as in (5.7) and (5.8). The extension of Definition 8.1 then reads:

Definition 8.2 (KS14). An index set $\Lambda \in \mathcal{J}$ is called a multitree if for all $i \in \{0, \ldots, n\}$ and all indices $\mu_j \in \mathcal{J}^{(j)}$ for $j \neq i$, the index set

$$\Lambda^{(i)} := \{\lambda_i \in \mathcal{J}^{(i)} : (\mu_0, \ldots, \mu_{i-1}, \lambda_i, \mu_{i+1}, \ldots, \mu_n) \in \Lambda\} \subset \mathcal{J}^{(i)} \tag{8.2}$$

is either the empty set or a tree in the sense of Definition 8.1.

Loosely speaking, a multitree $\Lambda \in \mathcal{J}$ is “when frozen in any $n$ coordinate direction, a tree in the remaining coordinate” (see KS13, §3.1).

Remark 8.3. Note that quasi-optimality of LS–AWGM is maintained if $\hat{\Lambda}_k$ are required to be multitrees (cf. KS13). The only modification is to replace the unconstrained non-linear approximation space $A^s$ (see (4.1)) by the constrained approximation space $A^s_{\text{mult}} := \{v \in \ell_2(\mathcal{J}) : \|v\|_{A^s_{\text{mult}}} < \infty\}$, where $\|v\|_{A^s_{\text{mult}}} := \sup_{\varepsilon > 0} \varepsilon \cdot \left(\min\{N \in \mathbb{N}_0 : \|v - v_N\|_{\ell_2(\mathcal{J})} \leq \varepsilon \wedge \supp v_N \text{ is a multitree}\}\right)^s$. This means we only allow those $v_N$ that are supported on a multitree.

The reason for using trees and multitrees for solving linear operator equations instead of arbitrary index sets lies in the much more efficient evaluation of system matrices which we explain next. Moreover, tree and multitree-structured index sets are crucial ingredients for the evaluation of non-linearities in both tensor product settings (e.g., SS11) and non-tensor product settings (e.g., CDD03).
8.2. Fast evaluation of tensor product system matrices. We assume that for some \( M \in \mathbb{N} \), there exist univariate bilinear forms \( b_{i}^{(j)} \) such that

\[
\mathbf{B} = \mathbf{D}^{Y} \left[ \sum_{m=1}^{M} \prod_{i=0}^{n} b_{i}^{(j)} \left( \bar{\Psi}^{(i)}, \tilde{\Psi}^{(i)} \right) \right] \mathbf{D}^{X} = \mathbf{D}^{Y} \left[ \sum_{m=1}^{M} \bigotimes_{i=0}^{n} \bar{B}_{i}^{(j)} \right] \mathbf{D}^{X},
\]

where \( \bar{B}_{i}^{(j)} := b_{i}^{(j)} \left( \bar{\Psi}^{(i)}, \tilde{\Psi}^{(i)} \right) \) for \( i = 0, \ldots, n \) and \( m = 1, \ldots, M \). This means that \( \mathbf{B} \) is a preconditioned sum of tensor product bilinear forms. As we shall see below, this form holds true for a large class of operators. Moreover, we shall always assume that \( b_{i}^{(j)} \) are \emph{local} in the sense that \( b_{i}^{(j)}(w, v) = 0 \) whenever \( \text{supp} v \cap \text{supp} w = 0 \). The special structure of \( \mathbf{B} \) can be used to efficiently realize the application of \( \mathbf{A} \mathbf{B} \) to a vector \( \mathbf{v} \in \ell_{2}(\Lambda) \) for finite multitrees \( \Lambda \subseteq \mathcal{F} \) and \( \mathbf{A} \subseteq \mathcal{F} \). As described in [KS14], this can be realized in linear complexity, i.e., \( \mathcal{O}(\#\Lambda + \#\bar{\Lambda}) \), by using a \emph{separation} of \( \mathbf{A} \mathbf{B} \) into unidirectional operations and an efficient tree-based application of unidirectional operations. These principles are also known from sparse grid algorithms (see e.g. [Zen91, BG04]).

We recall the Kronecker product of two general (possibly bi-infinite) matrices \( \bar{A}^{(1)}, \bar{A}^{(2)} \) and identity matrices \( \bar{I}^{d^{(1)}}, \bar{I}^{d^{(2)}} \) of appropriate dimension:

\[
\bar{A}^{(1)} \otimes \bar{A}^{(2)} = [\bar{A}^{(1)} \otimes \bar{I}^{d^{(2)}}] \circ [\bar{I}^{d^{(1)}} \otimes \bar{A}^{(2)}] = [\bar{I}^{d^{(1)} \otimes \bar{A}^{(2)}}] \circ [\bar{A}^{(1)} \otimes \bar{I}^{d^{(2)}}].
\]

Then we split \( \bar{B}_{m}^{(j)} = \bar{L}_{m}^{(j)} + \bar{U}_{m}^{(j)} \) into a (strictly) lower \( \bar{L}_{m}^{(j)} := \begin{bmatrix} (\bar{B}_{m}^{(j)})_{\lambda,\mu} | \lambda > | \mu | \end{bmatrix} \) and an upper triangular matrix \( \bar{U}_{m}^{(j)} := \begin{bmatrix} (\bar{B}_{m}^{(j)})_{\lambda,\mu} | \lambda \leq | \mu | \end{bmatrix} \). With (8.4.), it can then be shown that there exist multitrees \( \bar{\Xi} \) and \( \Xi \) such that we have the following equivalent representation of \( \mathbf{A} \mathbf{B} \) :

\[
\mathbf{D}^{Y} \left[ \sum_{m=1}^{M} \mathbf{R}_{\Lambda} \left[ \bar{I}^{d^{(0)}} \otimes \bar{B}_{m}^{(1)} \otimes \cdots \otimes \bar{B}_{m}^{(n)} \right] \mathbf{E}_{\Pi} \circ \mathbf{R}_{\Pi} \left[ \bar{U}_{m}^{(0)} \otimes \bar{I}^{d^{(1)}} \otimes \cdots \otimes \bar{I}^{d^{(n)}} \right] \mathbf{E}_{\bar{\Lambda}} \right] =: (I)
\]

\[
+ \sum_{m=1}^{M} \mathbf{R}_{\Lambda} \left[ \bar{L}_{m}^{(0)} \otimes \bar{I}^{d^{(1)}} \otimes \cdots \otimes \bar{I}^{d^{(n)}} \right] \mathbf{E}_{\Pi} \circ \mathbf{R}_{\Pi} \left[ \bar{U}_{m}^{(0)} \otimes \bar{B}_{m}^{(1)} \otimes \cdots \otimes \bar{B}_{m}^{(n)} \right] \mathbf{E}_{\bar{\Lambda}} \right] \mathbf{D}^{X}. =: (II)
\]

It holds that \( \#\bar{\Xi} + \#\Xi \leq \#\Lambda + \#\bar{\Lambda} \). The application of (II), (III) (and (I), (IV) for \( n = 1 \)) is referred to as \emph{unidirectional operation} as only the application of the univariate matrices \( \bar{L}_{m}^{(0)} | \Lambda_{(0)} \times \Lambda_{(0)} \), \( \bar{U}_{m}^{(0)} | \Lambda_{(0)} \times \Lambda_{(0)} \) and \( \bar{B}_{m}^{(1)} | \Lambda_{(1)} \times \Lambda_{(1)} \) (\( n = 1 \)) is required. Due to the tree structure, these tasks can be realized in linear complexity despite the fact that not one of the matrices \( \bar{L}_{m}^{(0)}, \bar{U}_{m}^{(0)} \) or \( \bar{B}_{m}^{(1)} \) is sparse in general (see [KS14, §2]). For \( n > 2 \), the remaining parts (I) and (IV) can be treated recursively by applying the same procedure to \( \bar{B}_{m}^{(1)} \otimes \cdots \otimes \bar{B}_{m}^{(n)} \).

**Theorem 8.4** ([KS13, Theorem 3.1]). Let \( \mathbf{A} \) be a linear differential operator with polynomial coefficients and let \( \Lambda \subseteq \mathcal{F} \), \( \bar{\Lambda} \subseteq \mathcal{F} \) be multitrees. Then, for any \( \mathbf{v} \in \ell_{2}(\Lambda) \), the product \( \mathbf{A} \mathbf{B} \mathbf{v} \) can be computed in \( \mathcal{O}(\#\Lambda + \#\bar{\Lambda}) \) operations.

**Remark 8.5.** If \( \mathbf{A} \) is a linear differential operator with polynomial coefficients, \( \mathbf{B} \) has the form (8.3). Furthermore, all matrices can be applied in linear complexity if \( \Lambda^{(i)} \) and \( \bar{\Lambda}^{(i)} \) are trees (cf. [KS14, §2]).
8.3. RESIDUAL: Multitree residual approximation. We need to approximate the residual \( B^T (f - Bw_\lambda) \) by a residual of type \( \tilde{\xi} B_{\tilde{\xi}} (f_{\tilde{\xi}} - \tilde{\xi} B_A w_\lambda) \).

8.3.1. Primal residual. We first recall the approximation of the primal residual.

**Theorem 8.6** ([KSI13]). Let \( 0 < \omega < 1 \), let \( A \) be a differential operator with polynomial coefficients and let \( u \in A_{\text{mtree}}^s \) for some \( s > 0 \). Then, for all finite multitrees \( \hat{\Lambda} \subset \mathcal{J} \) and all \( w_\Lambda \in \ell_2(\hat{\Lambda}) \), there exists a multitree \( \tilde{\xi} = \tilde{\xi}(\hat{\Lambda}, \omega) \subset \mathcal{J} \) such that \( \# \tilde{\xi} \leq \# \hat{\Lambda} + \nu^{-1/s} \) with \( \nu := \|\tilde{r}\|_{\ell_2(\mathcal{J})}, \tilde{r} := \tilde{\xi} B_{\tilde{\xi}} f \) and

\[
\|(f - Bw_\Lambda) - \tilde{r}\|_{\ell_2(\mathcal{J})} \leq \omega \|\tilde{r}\|_{\ell_2(\mathcal{J})},
\]

(8.5)

**Remark 8.7.** Due to the multitree structure of \( \hat{\Lambda} \) and \( \tilde{\xi} \), the computational cost for computing \( \tilde{r} \) is \( O(\# \hat{\Lambda} + \nu^{-1/s}) \) if an entry \( f_\Lambda \) of \( f = (f_\Lambda)_{\Lambda \in \mathcal{J}} \) can be computed exactly at unit cost, which is e.g. the case if \( f \) is a (piecewise) polynomial. If this assumption is not met, replace \( f \) by some \( f_\epsilon \) with \( \|f - f_\epsilon\|_{\ell_2(\mathcal{J})} \leq \epsilon \) and \( \# \text{supp} f_\epsilon \leq \epsilon^{-1/s} \), which is possible if \( f \) is sufficiently (piecewise) smooth (see [KSI13 §3.4]).

8.3.2. Dual residual. We may now follow [KSI13 §1.1] using a wavelet compression of \( B \) and \( B^T \). If \( A \) is a linear differential operator with polynomial coefficients, it can be shown that for any \( 0 < \eta < 1 \), there exists \( B_\eta : \ell_2(\mathcal{J}) \to \ell_2(\mathcal{J}) \) such that

\[
\|B - B_\eta\| \leq \eta, \quad \|B^T - B_\eta^T\| \leq \eta,
\]

(8.6)

where the number of non-zeros in each row and each column of \( B_\eta \) is of order \( O(\eta^{-1/s^*}) \) for some \( s^* > s_{\max}(5.16) \). This means that \( B \) is \( s^* \)-admissible (see [SS09]). Assuming that \( \eta \) is chosen sufficiently small so that \( B_\eta \) and \( B_\eta^T \) are boundedly invertible, we obtain the estimate (see Proposition B.2)

\[
\|B^T (f - Bw_\Lambda) - B_\eta^T \tilde{r}\|_{\ell_2(\mathcal{J})} \leq \omega_{ls} \|B_\eta^T \tilde{r}\|_{\ell_2(\mathcal{J})},
\]

(8.7)

for \( \omega_{ls} = (\eta^{1/2} + (\|B\| + \eta)\omega)\|B_\eta^{-1}\| \) so that \( \omega_{ls} \to 0 \) as \( \omega \to 0 \) and \( \eta \to 0 \). Even though \( B_\eta \) and \( B_\eta^T \) are sparse (for fixed \( \eta \)), the application of these matrices to finite vectors can be computationally expensive since the product structure of \( B \) in (8.3) cannot be exploited. Unfortunately, the approximate residual \( B_\eta^T \tilde{r} \) is *not* necessarily supported on a multitree. Hence, we define the multitree-based residual

\[
\tilde{r} := \tilde{\xi} B_{\tilde{\xi}} (f_{\tilde{\xi}} - \tilde{\xi} B_A w_\lambda) = \tilde{\xi} B_{\tilde{\xi}} \tilde{r}
\]

(8.8)

such that \( \|B^T (f - Bw_\Lambda) - \tilde{r}\|_{\ell_2(\mathcal{J})} \leq \omega_{ls} \|\tilde{r}\|_{\ell_2(\mathcal{J})} \) where \( \tilde{\xi} \) is the smallest multitree containing \( \text{supp} B_\eta^T \tilde{r} \). The residual computation requires \( O(\# \tilde{\xi} + \# \tilde{\xi}) \) operations.

**Remark 8.8.** Theorem 8.6 only ensures the existence of an appropriate multitree \( \tilde{\xi} \) but does not give any information on its explicit construction. The same holds true for \( \tilde{\xi} \). In Section 8.4 we will discuss how we can construct the multitrees \( \tilde{\xi} \) and \( \tilde{\xi} \) without setting up the compressed matrix \( B_\eta^T \) so that \( \tilde{r} \) from (8.8) satisfies \( (\text{RES}) \). Furthermore, numerical experiments in Section 9 indicate appropriate choices of \( \tilde{\xi} \) and \( \tilde{\xi} \) with *preferably small* cardinalities and *optimal balancing* of the error arising from the approximations of the primal (see (8.5)) and dual residual (see (8.7)).
8.4. GALSOLVE: Multitree solution of finite-dimensional least squares problems. Concerning the numerical solution of the least squares problem (8.9), the approach proposed in [Ste09, SS09] consists of replacing $\hat{A}B^T\hat{B}A$ by a sparse approximation $\hat{A}[B^T_B]_A := R_{\hat{A}}[B^T_B]_A$ satisfying $\|\hat{A}[B^T_B]_A \| \lesssim \eta$. In analogy to (8.9), we consider:

\begin{equation}
(8.9) \quad \text{Find } u_{\eta,\hat{A}} \in \ell_2(\hat{A}) : \quad \hat{A}[B^T_B]_A u_{\eta,\hat{A}} = R_{\hat{A}}B^T_{\eta} f_{\hat{A}}.
\end{equation}

Indeed, under the assumption that $\eta$ is sufficiently small, $\kappa(\hat{A}[B^T_B]_A)$ is bounded independently of $\hat{A}$ (see Appendix B). In particular, there exist algorithms based on linear iterative solvers like the conjugate gradient (cg) method that approximate the conjugate gradient (cg) method that approximate (8.9) such that $\|u_{\hat{A}} - u_{\eta,\hat{A}}\|_{\ell_2(\hat{A})} \lesssim \eta$ and $\text{[GAL]}$ is satisfied. Similarly to the residual approximation, the disadvantage of this approach is that we cannot use the fast matrix-vector multiplication w.r.t. multitrees. To this end, we intend to compute $w_{\hat{A}}$ as an approximate solution of the problem:

\begin{equation}
(8.10) \quad \text{Find } x_{\hat{A}} \in \ell_2(\hat{A}) : \quad \hat{A}[B^T_B]_A x_{\hat{A}} = \hat{A}B^T_{\eta} f_{\hat{A}}.
\end{equation}

We could choose $\hat{A}$ as the smallest multitree that contains supp$B_{\eta}$ for all $\forall_{\hat{A}} \in \ell_2(\hat{A})$. However, this is not an implementable approach. Hence, we are concerned with the question how the multitree $\hat{A}$ can be constructed in dependency of $\hat{A}$ such that (1) the condition number of $\hat{A}B^T_{\eta}$ is uniformly bounded and (2) an approximate solution $w_{\hat{A}}$ to (8.10) satisfies $\text{[GAL]}$. This will be discussed in Sections 8.5 and 8.6.

For fixed multitrees, the solution of (8.10) can be computed e.g. with cg.

8.5. Choice of index sets. The expansion $\hat{A}_k \to \hat{A}_{k+1}$ of the trial sets in Algorithm 2 is based upon the residual $\hat{f}_k$, but it is not clear how to construct appropriate test sets $\hat{A}_k = \hat{A}_k(\hat{A}_k)$. Similarly for the auxiliary sets $\hat{E}_k$ and $\hat{\hat{E}}_k$ required for (8.8): While the construction of the test sets $\hat{E}_k$ for the primal residual in a Petrov-Galerkin setting has been investigated in [KS13], there are so far no results for good choices of $\hat{E}_k$ and $\hat{\hat{E}}_k$ within a Petrov-Galerkin framework.

Choice of test sets $\hat{A}_k$. For a given index set $\hat{A}_k \in \hat{J}$, we have to ensure that the finite-dimensional test set $\hat{A}_k \in \hat{J}$ is large enough to ensure well-posedness. At the same time, for efficiency we would like to choose $\hat{A}_k \in \hat{J}$ as small as possible. We describe a corresponding iteration. As initial sets $\hat{A}_0$, $\hat{A}_0$, we follow [And13] §6.2

\begin{equation}
(8.11) \quad \hat{A}_0 = \hat{A}_{SG,J} := \{\lambda \in \hat{J} : |\lambda| \leq J\},
\end{equation}

\begin{equation}
\hat{A}_0 = \hat{A}_{SG,J} := \{\lambda \in \hat{J} : |\lambda| \leq J \text{ or } |\lambda_i| = J + 1, |\lambda_i| = 0, 1 \leq i \leq n\},
\end{equation}

where $|\lambda| := \sum_{i=0}^{n} |\lambda_i|^2$. Such bases are provably stable; however, this only holds true for uniform (full or sparse) discretizations. In later iterations, i.e., for adaptively constructed trial sets $\tilde{A}_k$, $k \geq 0$, we propose the following (heuristic) choices:

(i) $\hat{A}_{Full} = \text{FullStableExpansion}(\hat{A}, \ell)$ is defined as

\begin{equation}
(8.13) \quad \hat{A}_{Full} := \{\lambda \in \hat{A} : \exists \mu \in \hat{A} \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \ell \text{ and } \text{dist}(\text{supp } \psi_{\lambda_j}^{(j)}, \text{supp } \psi_{\mu_j}^{(j)}) \leq D\psi_{\ell} 2^{-|\lambda_j|}\}.
\end{equation}

\text{We will also use } \hat{A}_{SG,J} \text{ and } \hat{A}_{SG,J} \text{ within a uniform sparse grid (SG) discretization.
(ii) $\hat{\Lambda}_{\text{Red}} = \text{ReducedStableExpansion}(\hat{\Lambda}, \ell)$ is a subset of $\Lambda_{\text{Full}}$ defined as

$$
\hat{\Lambda}_{\text{Red}} := \bigcup_{i=0}^{n} \{ \lambda \in \mathcal{J} : \exists \mu \in \hat{\Lambda} \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{i,j} \ell \\
\text{and dist}(\text{supp } \hat{\psi}_{\lambda_j}^{(j)}, \text{supp } \hat{\psi}_{\mu_j}^{(j)}) \leq D_{\hat{\psi}}(j) 2^{-|\lambda_j|} \}.
$$

(iii) $\hat{\Lambda}_{\text{Temp}} = \text{TemporalStableExpansion}(\hat{\Lambda}, \ell)$: consists of only temporal higher level extensions, i.e.,

$$
\hat{\Lambda}_{\text{Temp}} := \{ \lambda \in \mathcal{J} : \exists \mu \in \hat{\Lambda} \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{0,j} \ell \\
\text{and dist}(\text{supp } \hat{\psi}_{\lambda_j}^{(j)}, \text{supp } \hat{\psi}_{\mu_j}^{(j)}) \leq D_{\hat{\psi}}(j) 2^{-|\lambda_j|} \}.
$$

We refer to [KS13, Prop. 2] for a proof that the above index sets are indeed multitrees. An algorithmic realization is shown in Algorithm 3.

**Algorithm 3** $[\hat{\Lambda}] = \text{FullStableExpansion}[\hat{\Lambda}, \ell]$

**Input:** Finite index set $\hat{\Lambda} \subset \mathcal{J}$, expansion level $\ell \in \mathbb{N}$.

1. $\hat{\Lambda} := \emptyset \subset \mathcal{J}$.
2. for $\lambda = (\lambda_0, \ldots, \lambda_n) \in \hat{\Lambda}$ do
   3. Find all “neighbours” $\mu = (\mu_0, \ldots, \mu_n) \in \mathcal{J}$ on the same level:
      $\tilde{\Lambda} \leftarrow \hat{\Lambda} \cup \{ \mu \in \mathcal{J} : |\mu_i| = |\lambda_i|, \text{supp } \hat{\psi}_{\mu_i} \cap \text{supp } \hat{\psi}_{\lambda_i} \neq 0 \forall i = 0, \ldots, n \}$.
   4. Find all “neighbours” $\tilde{\mu} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_n) \in \mathcal{J}$ on the $\ell$ higher levels:
      $\tilde{\Lambda} \leftarrow \tilde{\Lambda} \cup \{ \tilde{\mu} \in \mathcal{J} : |\tilde{\mu}_i| = |\lambda_i| + j, 1 \leq j \leq \ell, \text{supp } \hat{\psi}_{\tilde{\mu}_i} \cap \text{supp } \hat{\psi}_{\lambda_i} \neq 0 \forall i = 0, \ldots, n \}$.
   5. Complete $\tilde{\Lambda}$ to form a multitree in the sense of Definition 8.2
   6. end for

**Choice of sets $\tilde{\Xi}_k$, $\hat{\Xi}_k$.** The proposed index set reads

$$
\hat{\Xi}_k = \text{ReducedMultiTreeCone}(\hat{\Lambda}, \ell)
$$

$$
\tilde{\Xi}_k := \bigcup_{i=0}^{n} \{ \lambda \in \mathcal{J} : \exists \mu \in \hat{\Lambda}_k \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{i,j} \ell \\
\text{and dist}(\text{supp } \hat{\psi}_{\lambda_j}^{(j)}, \text{supp } \hat{\psi}_{\mu_j}^{(j)}) \leq D_{\hat{\psi}}(j) 2^{-|\lambda_j|} \}.
$$

It was shown in [KS13, KS14] that this index set for $\ell = 1$ and the analogously defined $\text{FullMultiTreeCone}(\hat{\Lambda}, 1)$ are adequate choices for an accurate approximation of the primal residual in the Galerkin setting, where $\hat{\Psi}^X = \hat{\Psi}^Y, \hat{\Xi} = \hat{\Xi}$.

In our Petrov-Galerkin setting, we combine the multitree cone extension with the expansions (8.13)-(8.15). More precisely, we consider the two variants $\text{FullResConstruction}$ and $\text{OptimResConstruction}$; see Figure 1. For the primal residual (i.e., in $Y$), we expand $\hat{\Lambda}_k$ to $\hat{\Xi}_k^{\text{tmp}} = \text{ReducedMultiTreeCone}(\hat{\Lambda}_k, \ell)$ and obtain the desired $\hat{\Xi}_k$ by one of the expansion variants in (8.13)-(8.15). For the dual residual (in $X$), we consider two approaches. In the first one, shown in Figure 1(a) we take the set $\hat{\Xi}_k$ as above and set $\hat{\Xi}_k = \text{FullStableExpansion}(\hat{\Xi}_k, \ell)$ (with obvious inverted roles of primal and dual basis). Then, $\hat{\Xi}_k$ is the smallest multitree containing $\text{supp } B_i^T \tilde{r}_k$ for sufficiently small $\eta$. The second approach uses the by far smaller set $\hat{\Xi}_k = \hat{\Xi}_k^{\text{tmp}}$ as indicated in Figure 1(b) [KS13].
9. Numerical experiments

We report on numerical experiments for time-periodic problems \((2.1)\). We focus on the stability of the arising normal equations \((8.10)\) in view of different choices for \(\hat{\Lambda}_k\). Moreover, we numerically investigate the quantitative behavior of approximate primal and dual residuals in view of Remark 8.8. It is sufficient to consider the case \(n = 1\) (so that \(\Omega = (0, 1)\)), since we employ an \(L_2(0, 1)\)-orthonormal (multi-)wavelet basis \(\Sigma = \Sigma\) (see \((5.5)\)) as in [Rup13], with \(d_x = 2\) and homogeneous boundary conditions. In this case, the Riesz constants in \((5.4)\), \((5.5)\) are independent of \(n\). In particular, the condition numbers of \(B^\top B\) and of \(\hat{\Lambda}_k B \hat{\Lambda}_k^\top\) do not depend on \(n\) so that the 1D case gives all relevant information. In [KS13], it was shown numerically that the asymptotic behavior of the multitree-based residual only differs by a constant depending on \(n\) from the unconstrained case.

We choose \(\Theta_{\text{per}}\) (see \((5.3)\)) as a collection of bi-orthogonal B-spline wavelets of order \(d_t = \tilde{d}_t = 2\) on the real line, periodized onto \([0, T]\), [Urb09]. For \(\Theta\) (see \((5.4)\)), we choose bi-orthogonal B-spline wavelets from [Dij09] with \(d_t = \tilde{d}_t = 2\). As further parameters for the LS-AWGM we choose \(\delta = 0.7\), \(\gamma_{\text{ls}} = 0.01\) and, if not indicated differently, \(\ell = 1\) for the stable extensions from Section 8.5. We obtain qualitatively similar results for choosing \(\Sigma\) as in [Dij09] for \(d_x = \tilde{d}_x = 2\) even though they do not satisfy our assumptions.

We also compare the LS–AWGM to a (uniform) sparse grid approach (SG), i.e., to computing the solutions on a sequence of uniform finite-dimensional sets \(\hat{\Lambda}_{SG,J}, \hat{\Lambda}_{SG,J}, J = 0, 1, \ldots\), as in \((8.11)\), \((8.12)\), e.g. [Zen91],[BG04].

9.1. Heat equation. We consider the 1D-inhomogeneous heat equation

\[
\begin{aligned}
&u_t - u_{xx} = f(t, x) &\text{on } \Omega = (0, 1), \\
&u(t, 0) = u(t, 1) = 0 &\text{for all } t \in [0, T], \\
&u(0, x) = u(T, x) &\text{on } \Omega,
\end{aligned}
\]

with a discontinuous source function \(f(t, x) \equiv f(t) := K (\frac{Nt}{T} - \lfloor \frac{Nt}{T} \rfloor)\), \(N \in \mathbb{N}\), \(K \in \mathbb{R}_+\). Our figures correspond to the choice \(N = 3, K = 1\).

Starting with the optimized residual (as in Figure 1(b)) and the full stable expansions as in \((8.13)\), we investigate the convergence of the adaptive algorithm and the stability of the finite-dimensional systems \((8.10)\). The norms of primal and dual

\[\text{Figure 1. Constructions of index sets } \hat{\Xi}_k, \tilde{\Xi}_k \text{ for residual approximation.}\]
residuals are shown in Figure 2(a) for AWGM and SG. As expected, LS–AWGM reaches the optimal rate \( s_{\text{max}} = d - 1 = 1 \), whereas uniform SG suffers from the lack of smoothness of the solution. We observe in Figure 2(b) that the iteration numbers for the least squares cg method in each LS–AWGM-iteration stabilize at about 150 iterations in both approaches. This indicates that the choice of test sets \( \tilde{\Lambda}_k = \tilde{\Lambda}_{\text{Full}} \) yields stability. Figure 2(c) shows the cardinalities of the test sets. They grow only linearly with \( \#\tilde{\Lambda}_k \), so that both \( w_{\tilde{\Lambda}_k} \) and \( \hat{r}_k \) can be computed within linear complexity in each iteration (cf. ([GAL],[RES])). These results are based on OptimResConstruction for \( \hat{\Xi}_k \). In Figure 3 FullResConstruction is used. As \( \hat{\Xi}_k \) hardly impacts \( \tilde{r}_k \), we monitor only the dual residual. Since using a larger index set, \( \|\tilde{r}_k\|_{\ell_2(\mathcal{P})} \) is slightly increased (as expected), but it exhibits the same behaviour as OptimResConstruction (Figure 3(a)). This marginal improvement comes at a high cost: \( \#\hat{\Xi}_k \) is 40–50 times larger; see Figure 3(b).

Finally, in Figure 4 we compare the stable expansion types (Full, Reduced, Temporal). We find no discernible differences in the residual (Figure 4(a)) and only a very slight increase in the iteration numbers in GALSOLVE (Figure 4(b)). It seems that choosing \( \hat{\Xi}_k = \text{TemporalStableExpansion}(\hat{\Xi}_k^{\text{tmp}}, 1) \) yields results that are comparable to the other extensions, which could not be deduced from [KS13],[KS14]. All three methods seem stable, and we can reduce the size of the test sets by a factor of about 3.4 for \( \tilde{\Lambda}_k \) (and likewise by 2.5 for \( \hat{\Xi}_k \)).
9.2. Convection-diffusion-reaction equation. As a second example, we consider the convection-diffusion-reaction (CDR) equation

\[
\begin{align*}
    u_t - u_{xx} + u_x + u &= f(t, x) & \text{on } \Omega = (0, 1), \\
    u(t, 0) &= u(t, 1) = 0 & \text{for all } t \in [0, T], \\
    u(0, x) &= u(T, x) & \text{on } \overline{\Omega},
\end{align*}
\]

for an \( f(t, x) \) that yields \( u(t, x) = e^{-1000(x-(0.5+0.25 \sin(2\pi t))^2)} \); see Figure 5(a). Note that \( u \) is infinitely smooth but exhibits large gradients in non-axis-aligned directions.

The support centers (i.e., the centers of \( \text{supp} \hat{\psi}_\lambda, \lambda \in \hat{A}_k \)) in Figure 5(b) indicate that the AWGM benefits from its ability to refine not only independently in each dimension, but in particular \textit{locally in the full space-time domain}. This is also mirrored in Figure 6(a), where we observe the optimal \( s_{\text{max}} = 1 \) for the LS–AWGM, and a stable number of inner iterations (Figure 6(b)), employing the optimized construction of \( \hat{\Xi}_k \) and only temporal stable expansions for \( \hat{\Lambda}_k, \hat{\Xi}_k \). The smoothness of the solution allows for a convergence rate close to 1 for the sparse grid approach; however, the asymptotic regime and comparable residual norms are only reached for index sets that are over a magnitude larger.

Finally, we compare the above AWGM results with those obtained for larger sets, i.e., using full stable expansions and the \textit{FullResConstruction} for \( \hat{\Xi}_k \). As before, we see in Figures 7(a), 7(b) that we can reduce the size of the test sets \( \hat{A}_k, \hat{\Xi}_k \) by...
factors 2 to 3 without losing accuracy. Likewise, the full construction of $p - \|v\|$ then yields for $v$ (1) Schwarz’s, Hölder’s and Young’s inequalities.

By definition of $\langle \cdot, \cdot \rangle$, with only a slight improvement in the residual approximation. (10) A

\[ A(t)^{-1}w(t) \text{ for the adjoint } A(t)^* \text{ of } A(t). \] The bound $\|(A(t)^*)^{-1}\| \leq \alpha^{-1}$ then yields for $v_w(t) := z_w(t) + w(t)$ that $\|v_w\|_Y \leq \sqrt{2} \max\{1, \alpha^{-1}\} \|w\|_X < \infty$. By definition of $z_w$ and (2.2), $\langle z_w(t), \dot{w}(t) \rangle_{V \times V'} = \langle z_w(t), A(t)[z_w(t)] \rangle_{V \times V'} \geq \alpha \|z_w(t)\|_V^2 \geq \frac{\alpha}{\gamma} \|\dot{w}(t)\|_{V'}^2$. Since $w \in X$ is periodic, we have $\int_0^T \langle w, \dot{w} \rangle_{V \times V'} + \langle z_w, A(t)[w] \rangle_{V \times V'} dt = \int_0^T \langle w, \dot{w} \rangle_{V \times V'} + \int_0^T \langle \dot{w}, w \rangle_{V \times V'} = \int_0^T \frac{d}{dt} \|w(t)\|^2_H dt = \|w(T)\|^2_H - \|w(0)\|^2_H = 0$, so that we finally get $\min\{1, \alpha^{-1}\} \|w\|_X^2 \geq \frac{\alpha \min\{1, \gamma^{-2}\}}{\sqrt{2} \max\{1, \alpha^{-1}\}} \|w\|_X \|w\|_Y > 0$.

(3) Surjectivity. Let $0 \neq w \in Y$. We aim to construct $z \in X$ with $\langle w(t), \dot{z}(t) \rangle_{V \times V'} + \langle w(t), A(t)[z(t)] \rangle_{V \times V'} = \langle w(t), A(t)[v(t)] \rangle_{V \times V'}$ for all $w \in Y$, and $t$ a.e. on $(0, T)$,
as then \( b(z, v) = \int_{0}^{T} \langle v(t), A(t)[v(t)] \rangle_{V' \times V} \geq \alpha \|v\|_{V}^{2} > 0 \), so that the surjectivity condition is fulfilled.

(i) **Faedo-Galerkin approximation of an initial value problem.** Let \( \{\phi_{i} : i \in \mathbb{N}\} \) be a basis for \( V, \ V_{n} := \text{span}\{\phi_{i}, i = 1, \ldots, n\} \), \( z_{n}(t) := \sum_{i=1}^{n} \frac{\alpha_{i}(t)}{\|z\|_{V}} \phi_{i} \). Then the linear system of ODEs \( \langle w_{n}, \dot{z}_{n}(t) \rangle_{V' \times V} + \langle w_{n}, A(t)[z_{n}(t)] \rangle_{V' \times V} = \langle w_{n}, A(t)[v(t)] \rangle_{V' \times V} \), \( z_{n}(0) = z_{0, n} \), has a solution \( z_{n} \in C(0, T; V_{n}) \) with \( \dot{z}_{n} \in L_{2}(0, T; V_{n}) \) for all \( w_{n} \in V_{n} \) a.e. on \( I \) and for (arbitrary) \( z_{0, n} \in H \) and its orthogonal projection \( z_{0, n} \) onto \( V_{n} \).

(ii) **A priori estimates.** (i), (A.1) and Young’s inequality with some \( \varepsilon < \frac{\gamma}{2} \) yield

\[
\frac{1}{2} \frac{d}{dt} \|z_{n}(t)\|_{H}^{2} + \alpha \|z_{n}(t)\|_{V}^{2} \leq \varepsilon \|z_{n}(t)\|_{V}^{2} + \frac{\gamma}{4} \|v(t)\|_{V}^{2},
\]

and hence by integration over \([0, s], s \in [0, T]\), using \((\alpha - \varepsilon) > 0\) that \( \|z_{n}(s)\|_{H}^{2} - \|z_{n}(0)\|_{H}^{2} = \int_{0}^{s} \frac{\alpha}{\varepsilon} \|z_{n}(t)\|_{V}^{2} dt \leq \frac{\gamma}{4} \int_{0}^{s} \|v(t)\|_{V}^{2} dt \), so that sup\(t \in [0, T]\) \( \|z_{n}(t)\|_{H} \leq \infty \) and \( \{z_{n}\}_{n \in \mathbb{N}} \) is uniformly bounded in \( L_{\infty}(0, T; H) \). Similarly, we can conclude that \( 2(\alpha - \gamma) \|z_{n}\|_{Y} \leq \|z_{n}(0)\|_{H} - \|z_{n}(T)\|_{H} + \frac{\gamma}{4} \|v\|_{V}^{2} < \infty \), so that \( \{z_{n}\}_{n \in \mathbb{N}} \) is also uniformly bounded in \( Y \).

(iii) **Periodicity.** Abbreviate \( \bar{c} := \frac{\gamma}{4}, \bar{\alpha} := \frac{2(\alpha - \gamma)}{c_{1}} > 0 \) with \( c_{1} := \sup_{\phi \in V} \|\phi\|_{Y} \) and multiply (A.1) by \( e^{\bar{\alpha} t} \). Then

\[
\frac{d}{dt} \left( e^{\bar{\alpha} t} \|z_{n}(t)\|_{H}^{2} \right) = e^{\bar{\alpha} t} \frac{\alpha}{\varepsilon} \|z_{n}(t)\|_{V}^{2} \leq e^{\bar{\alpha} t} \|v(t)\|_{V}^{2},
\]

and by integration over \([0, T], T \in [0, T]\), we obtain

\[
\|z_{n}(T)\|_{H}^{2} \leq \|z_{n}(0)\|_{H}^{2} + \bar{c} e^{-\bar{\alpha} T} \int_{0}^{T} e^{\bar{\alpha} t} \|v(t)\|_{V}^{2} dt.
\]

Set \( M := \{z \in V_{n} : \|z\|_{H} \leq R := K^{2} (1 - e^{-\bar{\alpha} T})^{-\frac{1}{2}} \}, \ K := \bar{c} e^{-\bar{\alpha} T} \int_{0}^{T} e^{\bar{\alpha} t} \|v(t)\|_{V}^{2} dt. \) The set \( M \) is convex and compact in \( V_{N} \). If \( z_{n}(0) \in M \), (A.2) implies that \( \|z_{n}(T)\|_{H}^{2} \leq e^{-\bar{\alpha} T} R^{2} + K \leq R \), i.e., \( z_{n}(T) \in M \). Since by Gronwall’s lemma the mapping \( S : M \to M, \ z_{n}(0) \to z_{n}(T) \), is continuous, the existence of a fixed-point \( S(z_{n}) = z_{n} \in M \) follows from Brouwer’s fixed-point theorem. By the a priori estimates, the sequence \( \{z_{n}\}_{n \in \mathbb{N}} \) is bounded in \( H \), so that there exists a subsequence (also denoted by \( \{\bar{z}_{n}\} \)) converging weakly to some \( \bar{z} \in H \).

(iv) **Convergence.** Consider the periodic solution \( z_{n}(t) \) from (iii), i.e., the solution of the ODE system with initial value \( z_{0, n} = \bar{z}_{n} \). From the a priori estimates, we have that \( \{z_{n}\} \) is uniformly bounded in the separable space \( Y \), so that there exists a subsequence (also denoted \( \{z_{n}\} \)) converging weakly to some \( z \in Y \). For \( w_{n} := \theta(t) \phi_{j}, \ \theta(t) \in C^{1}(0, T) \), we then have by integration over \([0, T] \) and integration by parts of the first term that for all \( j = 1, \ldots, n \) \( \langle \dot{\theta} \phi_{j}, z_{n} \rangle = \langle \theta(0) \phi_{j} - \theta(T) \phi_{j}, z_{n} \rangle_{H} + \langle \theta(t) \phi_{j}, A(t)[v - z_{n}] \rangle_{H} \).

As \( z_{n} \to z \) in \( Y \) and \( \bar{z}_{n} \to \bar{z} \) in \( H \), we can pass to the limit \( n \to \infty \) and obtain

\[
\langle \dot{\theta} \phi_{j}, z \rangle = \langle \theta(0) \phi_{j} - \theta(T) \phi_{j}, \bar{z} \rangle_{H} + \langle \theta \phi_{j}, A(t)[v - z] \rangle.
\]

This particularly holds true for all \( \theta \in D(I) \), so that \( \bar{z} = A(\cdot)(v - z) \) in the distributional sense and hence \( \bar{z} \in L_{2}(0, T; V') \). Moreover, (A.3) implies that for \( w \in C^{1}(0, T; V) \), we have \( \langle \dot{w}, z \rangle - \langle w(0) - w(T), z \rangle = \langle w, A(t)[v - z] \rangle = -\langle \dot{w}, z \rangle + \langle w(T), z(T) \rangle_{H} - \langle w(0), z(0) \rangle_{H}, \) so that indeed \( \bar{z} = z(0) = z(T) \) in \( H \) and hence \( z \in X \). With this \( z \), the surjectivity condition is fulfilled.
A SPACE-TIME WAVELET METHOD FOR TIME-PERIODIC PDES 1331

APPENDIX B. AUXILIARY WAVELET COMPRESSION RESULTS

Here, we report two facts for \( B \) defined in (3.3) which are required in Section 8. We shall always assume that (8.6) holds. For further details, we refer to [Kes13].

**Lemma B.1** ([KS13]). For sufficiently small \( \eta < 1 \), \( B_\eta \in L(\ell_2(\mathcal{F}), \ell_2(\mathcal{F})) \) and \( B_\eta^\top B_\eta \in L(\ell_2(\mathcal{F}), \ell_2(\mathcal{F})) \) are boundedly invertible with bounds depending on \( \eta \).

**Proposition B.2** ([KS13]). Let the assumptions of Theorem 8.6 hold. Then, there exists a constant \( \omega_{ls} \) such that
\[
\|B^\top(f - Bw_p \Lambda) - p_r\|_{\ell_2(\mathcal{F})} \leq \omega_{ls} \|p_r\|_{\ell_2(\mathcal{F})},
\]
where \( p_r := B_\eta^\top q_r \).

REFERENCES

[AKV06] J. M. Alam, N. K.-R. Kevlahan, and O. V. Vasilyev, *Simultaneous space-time adaptive wavelet solution of nonlinear parabolic differential equations*, J. Comput. Phys. **214** (2006), no. 2, 829–857, DOI 10.1016/j.jcp.2005.10.009. MR2216610 (2006k:65242)

[And13] R. Andreev, *Stability of sparse space-time finite element discretizations of linear parabolic evolution equations*, IMA J. Numer. Anal. **33** (2013), no. 1, 242–260, DOI 10.1093/imanum/drs014. MR3020957

[BG04] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer. **13** (2004), 147–269, DOI 10.1017/S0962492904000182. MR2249147 (2007e:65102)

[CDD01] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations: convergence rates*, Math. Comp. **70** (2001), no. 233, 27–75, DOI 10.1090/S0025-5718-00-01252-7. MR1803124 (2002h:65021)

[CDD02] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods. II. Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245, DOI 10.1007/s10208-000-0018-z. MR1907380 (2003f:65212)

[CDD03] A. Cohen, W. Dahmen, and R. Devore, *Sparse evaluation of compositions of functions using multiscale expansions*, SIAM J. Math. Anal. **35** (2003), no. 2, 279–303, DOI 10.1137/S0036141002412070. MR2001102 (2004f:41036)

[CS11] N. Chegini and R. Stevenson, *Adaptive wavelet schemes for parabolic problems: sparse matrices and numerical results*, IMA J. Numer. Anal. **49** (2011), no. 1, 182–212, DOI 10.1137/10080010027. MR2783222 (2012e:65205)

[CS12] N. Chegini and R. Stevenson, *The adaptive tensor product wavelet scheme: sparse matrices and the application to singularly perturbed problems*, IMA J. Numer. Anal. **32** (2012), no. 1, 75–104, DOI 10.1093/imanum/drr013. MR2857244

[Dah97] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta numerica, 1997, Acta Numer., vol. 6, Cambridge Univ. Press, Cambridge, 1997, pp. 55–228, DOI 10.1017/S0962492900002713. MR1489256 (98m:65102)

[DeV98] R. A. DeVore, *Nonlinear approximation*, Acta numerica, 1998, Acta Numer., vol. 7, Cambridge Univ. Press, Cambridge, 1998, pp. 51–150, DOI 10.1017/S0962492900002816. MR1689432 (2001a:41034)

[Dij09] T. Dijkema, *Adaptive tensor product wavelet methods for solving PDEs*, PhD thesis, Universiteit Utrecht, 2009.

[DL92] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Vol. 5*, Evolution Problems. I, with the collaboration of Michel Artola, Michel Cessenat and Hélène Lanchon, translated from the French by Alan Craig, Springer-Verlag, Berlin, 1992. MR1156075 (92k:00006)

[DSS09] T. J. Dijkema, C. Schwab, and R. Stevenson, *An adaptive wavelet method for solving high-dimensional elliptic PDEs*, Constr. Approx. **30** (2009), no. 3, 423–455, DOI 10.1007/s00365-009-9064-0. MR2558688 (2010m:65270)

[GHS07] T. Gantumur, H. Harbrecht, and R. Stevenson, *An optimal adaptive wavelet method without coarsening of the iterands*, Math. Comp. **76** (2007), no. 258, 615–629, DOI 10.1090/S0025-5718-06-01917-X. MR2291830 (2008h:65310)

[GO95] M. Griebel and P. Oswald, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Adv. Comput. Math. **4** (1995), no. 1-2, 171–206, DOI 10.1007/BF02123478. MR1338900 (96e:65069)
[GO07] M. Griebel and D. Oeltz, A sparse grid space-time discretization scheme for parabolic problems, Computing 81 (2007), no. 1, 1–34, DOI 10.1007/s00607-007-0241-3. MR2369419
[HV95] G. Horton and S. Vandewalle, A space-time multigrid method for parabolic partial differential equations, SIAM J. Sci. Comput. 16 (1995), no. 4, 848–864, DOI 10.1137/0916050. MR1335894
[JSU07] D. Jürgens, M. Palm, S. Singer, and K. Urban, Numerical optimization of the Voith-Schneider® Propeller, ZAMM 87 (2007), no. 10, 698–710.
[KB06] Y. Kawajiri and L. T. Biegler, Optimization strategies for simulated moving bed and PowerFeed processes, AIChE Journal 52 (2006), no. 4, 1343–1350.
[Kes13] S. Kestler, On the adaptive tensor product wavelet Galerkin method with applications in finance, PhD thesis, University of Ulm, 2013.
[KS13] S. Kestler and R. Stevenson, An efficient approximate residual evaluation in the adaptive tensor product wavelet method, J. Sci. Comput. 57 (2013), no. 3, 439–463, DOI 10.1007/s10915-013-9712-1. MR3123552
[KS14] S. Kestler and R. Stevenson, Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions, J. Comput. Appl. Math. 260 (2014), 103–116, DOI 10.1016/j.cam.2013.09.015. MR3133333
[MV07] D. Meidner and B. Vexler, Adaptive space-time finite element methods for parabolic optimization problems, SIAM J. Control Optim. 46 (2007), no. 1, 116–142 (electronic), DOI 10.1137/060648994. MR2299622
[Nit06] P.-A. Nitsche, Best $N$ term approximation spaces for tensor product wavelet bases, Constr. Approx. 24 (2006), no. 1, 49–70, DOI 10.1007/s00365-005-0609-6. MR2217525
[Raa07] T. Raasch, Adaptive wavelet and frame schemes for elliptic and parabolic equations, PhD thesis, University Marburg, 2007.
[Rup13] A. Rupp, High dimensional wavelet methods for structures financial products, PhD thesis, University of Ulm, 2013.
[SS09] C. Schwab and R. Stevenson, Space-time adaptive wavelet methods for parabolic evolution problems, Math. Comp. 78 (2009), no. 267, 1293–1318, DOI 10.1090/S0025-5718-08-02205-9. MR2501051
[SS11] C. Schwab and R. Stevenson, Fast evaluation of nonlinear functionals of tensor product wavelet expansions, Numer. Math. 119 (2011), no. 4, 765–786, DOI 10.1007/s00211-011-0397-9. MR2854127
[Ste09] R. Stevenson, Adaptive wavelet methods for solving operator equations: an overview, Multiscale, nonlinear and adaptive approximation, Springer, Berlin, 2009, pp. 543–597, DOI 10.1007/978-3-642-04313-8_13. MR2648381
[SU09] W. Sickel and T. Ullrich, Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross, J. Approx. Theory 161 (2009), no. 2, 748–786, DOI 10.1016/j.jat.2009.01.001. MR2563079
[SZ99] H. L. Smith and X.-Q. Zhao, Dynamics of a periodically pulsed bio-reactor model, J. Differential Equations 155 (1999), no. 2, 368–404, DOI 10.1006/jdeq.1998.3587. MR1698559
[UP12] K. Urban and A. T. Patera, A new error bound for reduced basis approximation of parabolic partial differential equations (English, with English and French summaries), C. R. Math. Acad. Sci. Paris 350 (2012), no. 3-4, 203–207, DOI 10.1016/j.crma.2012.01.026. MR2891112
[UP14] K. Urban and A. T. Patera, An improved error bound for reduced basis approximation of linear parabolic problems, Math. Comp. 83 (2014), no. 288, 1599–1615, DOI 10.1090/S0025-5718-2013-02782-2. MR3194123
[Urb09] K. Urban, Wavelet Methods for Elliptic Partial Differential Equations, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2009. MR2473650
[Zen91] C. Zenger, Sparse grids, Parallel algorithms for partial differential equations (Kiel, 1990), Notes Numer. Fluid Mech., vol. 31, Vieweg, Braunschweig, 1991, pp. 241–251. MR1167882
