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 algorithm 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 parallelization. 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 problems, i.e., when searching for solutions \( u \) with \( u(0) = u(T) \). Such problems arise...
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 $C : \ell_2 \to \ell_2$, 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 $u_{A_k}$ satisfying $C_{\Lambda_k \times \Lambda_k} u_{A_k} = g_{\Lambda_k}$ is computed. The (infinite-dimensional) residual $r_{A_k} := g - Cu_{A_k}$ 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 $r_{A_k}$).

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 $\mathbf{B}$. In the present work, we use another approach based on \textit{multitree-structured} index sets as introduced in [KS13, KS14]. This allows for the \textit{exact} application of $\mathbf{B}$ for wavelet discretizations of linear differential operators with polynomial coefficients within linear complexity when the underlying wavelet basis is of \textit{tensor product} type. It basically consists of the so-called \textit{unidirectional principle} first introduced in sparse grid algorithms (e.g. [BG04, Zen91]) where coordinate directions may be treated \textit{separately}. The evaluation is then based on multitree-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, \textit{no} quantitative results on the numerical solution of parabolic operator problems using the multitree concept within an AWGM are available. So far, only numerical results for \textit{elliptic} operator 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 \textit{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, multitree-based AWGM is then presented in Sections 7 and 8. We underline our theoretical findings by numerical experiments in Section 9.

2. \textbf{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) \quad \text{in } V' \text{ for a.e. } t \in [0, T], \quad u(0) = u(T) \quad \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}
(2.2) \quad \langle v, A(t)[w] \rangle_{V \times V'} \leq \gamma \|v\|_V \|w\|_{V'}, \quad \langle v, A(t)[v] \rangle_{V \times V'} \geq \alpha \|v\|^2_V \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} W_{ij}, \text{ where } W_{ij} := \begin{cases} L_2(\Omega), & 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 $V = H^1_0(\Omega)$ and $V^{(i)} = H^1_0(\Omega_i)$ (see [GO95]).

2.1. \textbf{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 \textit{space-time variational formulation}: Defining

\begin{equation}
(2.4) \quad H^1_{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.,

\[
\mathcal{X} = \{ v \in L_2(0, T; V) : v_t \in L_2(0, T; V'), \ v(0) = v(T) \text{ in } H \},
\]

where \( \mathcal{X} \) is equipped with the norm \( \| v \|_{\mathcal{X}} := \| v \|_{L_2(0, T; V)}^2 + \| v_t \|_{L_2(0, T; V')}^2, \ v \in \mathcal{X} \). 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):

\[
(2.7) \quad b(u, v) := \int_0^T \langle v(t), u_t(t) + A(t)[u] \rangle_{V \times V'} dt, \ f(v) := \int_0^T \langle v(t), g(t) \rangle_{V \times V'} dt.
\]

We define the space-time operator \( \mathcal{B} \in \mathcal{L}(\mathcal{X}, \mathcal{Y}') \) by \( \langle v, \mathcal{B}[u] \rangle := b(u, v) \) 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:

\[
(2.8) \quad \text{Find } u \in \mathcal{X} : \quad \mathcal{B}[u] = f, \quad f \in \mathcal{Y}'.
\]

### 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:

\begin{enumerate}[\text{(i)}]
\item \textbf{Continuity:} \( \gamma_B := \sup_{0 \neq u \in \mathcal{X}} \sup_{0 \neq v \in \mathcal{Y}} \frac{b(u, v)}{\| u \|_{\mathcal{X}} \| v \|_{\mathcal{Y}}} < \infty \).
\item \textbf{Inf-sup condition:} \( \beta_B := \inf_{0 \neq u \in \mathcal{X}} \sup_{0 \neq v \in \mathcal{Y}} \frac{b(u, v)}{\| u \|_{\mathcal{X}} \| v \|_{\mathcal{Y}}} > 0 \).
\item \textbf{Surjectivity:} \( \sup_{0 \neq u \in \mathcal{X}} | b(u, v) | > 0 \) for all \( 0 \neq v \in \mathcal{Y} \).
\end{enumerate}

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

### 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:

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

labeled w.r.t. two (possibly) different countable index sets \( \mathcal{J} \) and \( \mathcal{J}' \). More precisely, we consider a trial basis \( \Psi^X \) for the ansatz space \( X \) and a test basis \( \Psi^Y \) for the test space \( Y \) with associated Riesz constants \( c_X(\Psi^X), C_X(\Psi^X) \) and \( c_Y(\Psi^Y), C_Y(\Psi^Y) \). It is important to note that \( \Psi^X, \Psi^Y \) 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 5). Then there exists a unique expansion \( u = u^\top \Psi^X \) of the solution \( u \) of \eqref{2.8} where we formally interpret both \( u \in \ell_2(\mathcal{J}) \) and \( \Psi^X \) as column vectors. Now, the equivalent formulation of \eqref{2.8} reads as follows:

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

where \( B := [\langle \psi^X_\lambda, B[\psi^X_\mu] \rangle_{\lambda,\mu \in \mathcal{J}} = [b(\psi^X_\lambda, \psi^Y_\mu)]_{\lambda,\mu \in \mathcal{J}} = \langle \Psi^Y, B[\Psi^X] \rangle \) is the bi-infinite stiffness matrix and \( f = [\langle \psi^Y_\lambda, f \rangle]_{\lambda \in \mathcal{J}} = \langle \Psi^Y, f \rangle \) is the infinite right-hand side. It is easy to see that \eqref{3.3} is well-posed. Since, \( f \in \ell_2(\mathcal{J}) \) and \( B \in \mathcal{L}(X, Y') \) are boundedly invertible, also \( B \in \mathcal{L}(\ell_2(\mathcal{J}), \ell_2(\mathcal{J})) \) is boundedly invertible. In particular, with \( \| \cdot \| := \| \cdot \|_{\ell_2 \to \ell_2} \) (compare \cite{SS09} (2.2) and (2.3)),

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

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

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

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

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

Finally, \( C \subseteq D \) means that \( C \) can be bounded by a constant times \( D \) and \( C \gtrsim D \) is defined as \( D \lesssim C \). In this setting, \( C \gtrsim D \) 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 \eqref{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 \( v = v^\top \Psi^X \in X \) in the basis \( \Psi^X \) with \( N \) d.o.f. is a nonlinear, best \( N \)-term approximation (e.g. \cite{DeV98}), i.e., \( v_N = \arg \sigma_N(v) \), where the best \( N \)-term approximation error is defined as

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

Since \( \Psi^X \) is a Riesz basis, it holds that \( \| v - v_N \|_{\ell_2} \sim \sigma_N(v) \) 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 $\mathbf{v} \in \ell_2(\mathcal{J})$ 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)])

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

For a given $\mathbf{v} \in \mathcal{A}^{s}$ 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 $N_\varepsilon \leq \varepsilon^{-1/s}\|\mathbf{v}\|_{\mathcal{A}^{s}}^{1/s}$. It is important to remark that this bound on $N_\varepsilon$ is usually sharp (see [Ste09, (3)]).

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

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

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]):

\begin{equation}
\mathcal{X} \approx \left[ L_2(0, T) \otimes V \right] \cap \left[ H^1_{\text{per}}(0, T) \otimes V' \right], \quad \mathcal{Y} \approx L_2(0, T) \otimes V.
\end{equation}

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. $\Omega_i = (0, 1)$),

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

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

\begin{equation}
\Theta_{\text{per}} := \{ \theta_{\lambda}^\text{per} : \lambda \in J_{t}^\text{per} \} \subset H^{1}_{\text{per}}(0, T)
\end{equation}

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 $\{ \theta_{\lambda}^\text{per} / \| \theta_{\lambda}^\text{per} \|_{H^{1}} : \lambda \in J_{t}^\text{per} \}$ is a Riesz basis for $H^{1}_{\text{per}}(0, T)$ with constants $c_{H^{1}_{\text{per}}}^\text{per} (\Theta_{\text{per}})$, $C_{H^{1}_{\text{per}}}^\text{per} (\Theta_{\text{per}})$. Recall that the construction of periodic wavelet bases is particularly easy [Urb09]. For the temporal part of the test space $Y$ (involving also non-periodic functions), we consider a uniformly local, piecewise polynomial wavelet basis for $L_{2}(0, T)$,

\begin{equation}
\Theta := \{ \theta_{\lambda} : \lambda \in J_{t} \},
\end{equation}

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]) with $L_{2}(\Omega) \subseteq V$ and $L_{2}(\Omega_{i}) \simeq L_{2}(\Omega_{1}) \otimes \cdots \otimes L_{2}(\Omega_{n})$ (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 $L_{2}(\Omega_{i})$, $\Sigma^{(i)} := \{ \sigma_{\lambda}^{(i)} : \lambda \in J^{(i)} \} \subset V^{(i)}$. We require that these functions are sufficiently smooth so that $\{ \sigma_{\lambda}^{(i)} / \| \sigma_{\lambda}^{(i)} \|_{V^{(i)}} : \lambda \in J^{(i)} \}$, $\{ \sigma_{\lambda}^{(i)} / \| \sigma_{\lambda}^{(i)} \|_{V^{(i)^{r}}} : \lambda \in J^{(i)} \}$ are Riesz bases for $V^{(i)}$, $V^{(i)^{r}}$ with constants $c_{V^{(i)}}(\Sigma^{(i)})$, $C_{V^{(i)}}(\Sigma^{(i)})$ and $c_{V^{(i)^{r}}}(\Sigma^{(i)})$, $C_{V^{(i)^{r}}}(\Sigma^{(i)})$. Now,

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

is a Riesz basis for $L_{2}(\Omega)$ where $\sigma_{\lambda} := \sigma_{\lambda}^{(1)} \otimes \cdots \otimes \sigma_{\lambda}^{(n)}$ is a tensor product wavelet and $J_{x} := J^{(1)} \times \cdots \times J^{(n)}$ [Dij09] Lemma 3.1.7. Moreover,

\begin{equation}
\Sigma^{V} := \{ \sigma_{\lambda} / \| \sigma_{\lambda} \|_{V} : \lambda \in J_{x} \}, \quad \Sigma^{V^{r}} := \{ \sigma_{\lambda} / \| \sigma_{\lambda} \|_{V^{r}} : \lambda \in J_{x} \}
\end{equation}

are Riesz bases for $V$, $V^{r}$ [Dij09] Lemma 3.1.8. The associated Riesz constants will be denoted by $c_{V}(\Sigma)$, $C_{V}(\Sigma)$, $c_{V^{r}}(\Sigma)$ and $C_{V^{r}}(\Sigma)$.
5.4. Space-time discretization. We are now in a position to define the Riesz wavelet bases $\tilde{\Psi}^X$ and $\tilde{\Psi}^Y$ from (5.2). With $L_2(0, T; L_2(\Omega)) \simeq L_2(0, T) \otimes L_2(\Omega)$,

$$\tilde{\Psi} := \{ \tilde{\psi}_\lambda := \vartheta_{\lambda}^{per} \otimes \sigma_{\lambda} : \lambda := (\lambda_t, \lambda_x) \in \tilde{J} := J_t^{per} \times J_x \} = \Theta^{per} \otimes \Sigma,$$

(5.7) $\tilde{\Psi}$ 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

$$\tilde{\Psi}^X := \{ \tilde{\psi}_\lambda ||\tilde{\psi}_\lambda||_X : \lambda \in \tilde{J} \} = D^X \tilde{\Psi}, \quad D^X := \text{diag} \left( \|\tilde{\psi}_\lambda\|_X^{-1} \right)_{\lambda \in \tilde{J}},$$

(5.9) $\tilde{\Psi}^X$ are Riesz bases for $L_2(0, T; L_2(\Omega))$. We conclude that for the computation of the bounds in (5.11), (5.12) and (5.13), it is sufficient to compute bounds for Riesz constants which can be easily approximated (e.g. [Dij09, Propositions 1 and 2]).

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

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

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_n)$ and $J := J^{(0)} \times J^{(1)} \times \cdots \times J^{(n)}$. In this setting, it is clear that $\Psi^{(0)} \in \{ \Theta^{per}, \Theta \}$, $J^{(0)} \in \{ J_t^{per}, J_t \}$ 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 $c_X(\tilde{\Psi}), C_X(\tilde{\Psi}), c_Y(\tilde{\Psi}), C_Y(\tilde{\Psi})$ in (5.7).

Again, we use that $X$ and $Y$ are (intersections of) tensor products of Hilbert spaces. As in [SS09, Section 6], we have the following estimates for $\tilde{\Psi}^X$ and $\tilde{\Psi}^Y$:

$$c_X(\tilde{\Psi}) \geq \min \left\{ C_{L_2}(\Theta^{per}) \cdot C_V(\Sigma), \quad C_{H_1^{per}}(\Theta^{per}) \cdot C_{V'}(\Sigma) \right\},$$

(5.11) $C_X(\tilde{\Psi}) \leq \min \left\{ C_{L_2}(\Theta^{per}) \cdot C_V(\Sigma), \quad C_{H_1^{per}}(\Theta^{per}) \cdot C_{V'}(\Sigma) \right\},$

(5.12) $C_Y(\tilde{\Psi}) \geq c_{L_2}(\Theta) \cdot c_V(\Sigma), \quad C_Y(\tilde{\Psi}) \leq C_{L_2}(\Theta) \cdot C_V(\Sigma).$

The Riesz constants $C_V(\Sigma)$, $C_{V'}(\Sigma)$ 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, Section 2], that

$$C_V(\Sigma) \geq \min_{m \in \{1, \ldots, n\}} \min \left\{ C_{L_2}(\Sigma^{(m)}), \quad C_{V}(\Sigma^{(m)}) \right\} \prod_{k \neq m} C_{L_2}(\Sigma^{(k)}),$$

(5.14) $C_V(\Sigma) \leq \max_{m \in \{1, \ldots, n\}} \max \left\{ C_{L_2}(\Sigma^{(m)}), \quad C_{V}(\Sigma^{(m)}) \right\} \prod_{k \neq m} C_{L_2}(\Sigma^{(k)}).$

Unfortunately, the same approach does not apply to the (dual) Riesz constants $c_{V'}(\Sigma)$, $C_{V'}(\Sigma)$ 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., $(\tilde{\Sigma}^V, \Sigma^{V'})_{V \times V'} = \text{Id}$. Denoting by $c_{V'}(\Sigma)$, $C_{V'}(\Sigma)$ the associated Riesz constants, it can be shown that $C_{V'}(\Sigma)^{-1} \leq c_{V'}(\Sigma)$ and $C_{V'}(\Sigma) \leq c_{V'}(\Sigma)^{-1}$. Observe that for computing bounds for $c_{V'}(\Sigma)$, $C_{V'}(\Sigma)$, we may proceed as for bounding $c_{V'}(\Sigma)$, $C_{V'}(\Sigma)$. 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. [Dij09, Section 2]).

Remark 5.2. Recalling the construction of wavelets, note that the numerical approximation of $c_{V'}(\Sigma)$, $C_{V'}(\Sigma)$ 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 $\Sigma^V = \Sigma^{V'}$, e.g. multiwavelets.
5.6. **Best approximation rates.** We need to know for which values of $s$ the solution $u$ of (3.3) is in $A^s$. 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 $u \in A^s$ 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 $u = u^T \hat{\Psi}^\mathcal{X} \in \mathcal{X} \cap H^{d_0}(0,T) \otimes \mathcal{H}^{d_x}(\Omega)$ and the Sobolev space

$$\mathcal{H}^{d_x}(\Omega) := \bigcap_{i=1}^n \bigotimes_{j=1}^n Z_{ij}, \text{ where } Z_{ij} := \begin{cases} L_2(\Omega_i), & i \neq j, \\ H^{d_x}(\Omega_i), & i = j \end{cases}$$

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

$$(5.16) \quad s_{\text{max}} = \min\{d_x - 1, d_x - m\}.$$ 

We recall that $d_x$ denotes the polynomial order of $\Theta^{\text{per}}$ and $d_x$ those of $\Sigma^{(1)}, \ldots, \Sigma^{(n)}$. This rate does not depend on the spatial dimension $n$. Moreover, we remark that $u \in \mathcal{H}^{d_x}(\Omega)$ is sufficient but not necessary for obtaining the above rate. In fact, the Sobolev space $H^{d_0}(0,T) \otimes \mathcal{H}^{d_x}(\Omega)$ 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 $\mathcal{C} \in \mathcal{L}(\mathcal{X}, \mathcal{X}^\prime)$ induced by a continuous and coercive bilinear form (i.e., $\langle v, \mathcal{C}[w]\rangle_{\mathcal{X},\mathcal{X}^\prime} \leq \|v\|_{\mathcal{X}}\|w\|_{\mathcal{X}^\prime}$, $\langle v, \mathcal{C}[v]\rangle_{\mathcal{X},\mathcal{X}^\prime} \geq \|v\|^2_{\mathcal{X}}$, for all $v, w \in \mathcal{X}$), we consider:

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

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

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

where $\mathcal{C} = (\hat{\Psi}^\mathcal{X}, \mathcal{C}[\hat{\Psi}^\mathcal{X}])_{\mathcal{X},\mathcal{X}^\prime}$ and $g = (\hat{\Psi}^\mathcal{X}, g)_{\mathcal{X},\mathcal{X}^\prime}$ 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 $\|\cdot\|^2 := \langle \cdot, \cdot \rangle_{\ell_2(\mathcal{J}) \times \ell_2(\mathcal{J})}$ defines an equivalent norm [Ste09, p. 565]

$$(6.3) \quad \|C^{-1}\|^{-\frac{1}{2}}\|v\|_{\ell_2} \leq \|v\| \leq \|C\|^\frac{1}{2}\|v\|_{\ell_2}, \quad \forall v \in \ell_2(\mathcal{J}).$$

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 $\hat{\mathcal{A}}_1 \subset \mathcal{J}$, a sequence of nested finite index sets $(\hat{\mathcal{A}}_k)_k$ 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 (3.1))

$$cX(\Psi)^{2}\|u - u_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})} \leq \|u - u_{\hat{A}_{k}}^{T}\Psi^{X}\|_{NN} \leq C_{X}(\Psi)^{2}\|u - u_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})}.$$}

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

$$(6.4) \quad \|C^{-1}\|\|g - Cu_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})} \leq \|u - u_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})} \leq \|C^{-1}\|\|g - Cu_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})}.$$}

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 $\Lambda_{k}$. This so-called bulk-chasing process is steered by the parameter $\delta$.

**Algorithm 1** $[u_{\varepsilon}] = \text{IDEALIZED–AWGM}[\varepsilon, \Lambda_{1}]$

*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

$$(6.6) \quad \|R_{\hat{A}_{k+1}}(g - Cu_{\hat{A}_{k}})\|_{\ell_{2}(\hat{A}_{k+1})} \geq \delta\|g - Cu_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})}.$$}

6. end for

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

$$(6.7) \quad \|u - u_{\hat{A}_{k}}\|_{\ell_{2}(\hat{J})} \lesssim \|u\|_{A^{s}s}^{1/s}N_{k}^{-s}, \quad \#\text{supp }u_{\varepsilon} \lesssim \varepsilon^{-1/s}\|u\|_{A^{s}}^{1/s}.$$}

**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 (5.2) 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 (3.3). However, the generalization of the idealized scheme to (3.3) is not trivial: (1) Symmetry and positive definiteness. Recall that $B$ from (3.3) is not s.p.d., so that $\langle \cdot, B \cdot \rangle_{\ell_{2}(\hat{J})\times \ell_{2}(\hat{J})}$ is not an equivalent norm on $\ell_{2}(\hat{J})$. However, the availability of an equivalent energy norm as in (3.3) is crucial for the convergence analysis of Algorithm 1 (see [Ste09] Proposition 4.1). (2) Bulk chasing and residual
A SPACE-TIME WAVELET METHOD FOR TIME-PERIODIC PDES 1319

computations. It is not clear how to construct $\hat{\Lambda}_{k+1}$ from $\hat{\Lambda}_k$. In analogy to (3.4), the residual $f - B u_{\hat{\Lambda}_k} \in \ell_2(\hat{\mathcal{J}})$ with error estimator $\|f - B u_{\hat{\Lambda}_k}\|_{\ell_2(\hat{\mathcal{J}})}$ satisfies

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

But the residual is an element of $\ell_2(\hat{\mathcal{J}})$, $\mathcal{J} \neq \hat{\mathcal{J}}$. Thus, we cannot compute $\hat{\Lambda}_k$ by selecting some contributions from the residual as in (6.6). (3) Petrov-Galerkin problems. Since $\hat{\Psi}^\top \neq \hat{\Psi}^3$, the (well-posed) Galerkin problem in line 5 of Algorithm 1 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]:

\begin{equation}
\text{(6.9) Find } u \in \ell_2(\hat{\mathcal{J}}) : \quad B^\top Bu = B^\top f, \quad B^\top f \in \ell_2(\hat{\mathcal{J}}).
\end{equation}

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 $u \in \ell_2(\hat{\mathcal{J}})$ such that $u = \arg\min_{v \in \ell_2(\hat{\mathcal{J}})} \|Bv - f\|_{\ell_2(\hat{\mathcal{J}})}^2$ for given $f \in \ell_2(\hat{\mathcal{J}})$. 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

\begin{equation}
\|B^\top B\| \leq \|B\|^2, \quad \|(B^\top B)^{-1}\| \leq \|B^{-1}\|^2,
\end{equation}

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 $\ell_2(\hat{\mathcal{J}})$, we now obtain $B^\top(f - B u_{\hat{\Lambda}_k}) \in \ell_2(\hat{\mathcal{J}})$ with error estimator $\rho_k := \|B^\top(f - B u_{\hat{\Lambda}_k})\|_{\ell_2(\hat{\mathcal{J}})}$. In analogy to (6.4) and (6.8), we infer that

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

In this setting, the residual $f - B u_{\hat{\Lambda}_k}$ from (6.8) is also referred to as primal residual, whereas $B^\top(f - B u_{\hat{\Lambda}_k})$ 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 $B_{\hat{\Lambda}}$ defined in (3.6), we get $(B^\top B)|_{\hat{\Lambda} \times \hat{\Lambda}} = \hat{\Lambda} B^\top B_{\hat{\Lambda}}$ so that (6.5) for general $\hat{\Lambda} \subset \hat{\mathcal{J}}$ with $C = B^\top B$ now reads as follows:

\begin{equation}
\text{(6.12) Find } u_{\hat{\Lambda}} \in \ell_2(\hat{\Lambda}) : \quad \hat{\Lambda} B^\top B_{\hat{\Lambda}} u_{\hat{\Lambda}} = \hat{\Lambda} B^\top f, \quad \hat{\Lambda} B^\top f \in \ell_2(\hat{\Lambda}).
\end{equation}

Observe that the unique solution $u_{\hat{\Lambda}} = \arg\min_{\nu_{\hat{\Lambda}} \in \ell_2(\hat{\Lambda})} \|B_{\hat{\Lambda}} V_{\hat{\Lambda}} f - f\|_{\ell_2(\hat{\mathcal{J}})}^2$ 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 $B^\top B$ 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 $\hat{\Lambda} \subset \hat{\mathcal{J}}$.

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 \( w_{\tilde{A}} \) to \( u_{\hat{A}_k} \) and an approximate residual \( \tilde{r}_k \) to \( B^T(f - Bu_{\hat{A}_k}) \). 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 \( u \in A^s \) and denote by \( w_{\tilde{A}} \) an approximate solution to (6.12).

**(RES) Approximate residual:** For a given relative tolerance \( 0 < \omega_{ls} < 1 \), the output \( \tilde{r} \) of RESIDUAL\([w_{\tilde{A}}, \omega_{ls}]\) should satisfy

\[
\|B^T(f - Bw_{\tilde{A}}) - \tilde{r}\|_{\ell_2(\mathcal{J})} \leq \omega_{ls} \cdot \nu, \quad \nu := \|\tilde{r}\|_{\ell_2(\mathcal{J})},
\]

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

**(GAL) Approximate Galerkin problem:** For a given relative tolerance \( 0 < \gamma_{ls} < 1 \), the output \( w_{\tilde{A}} \) of GALSOLVE\([\hat{A}, w_{\tilde{A}}, \gamma_{ls} : \nu]\) should satisfy

\[
\|\hat{A}B^T(f - B\tilde{A}w_{\tilde{A}})\|_{\ell_2(\hat{A})} \leq \gamma_{ls} \cdot \nu,
\]

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

**(EXP) Approximate bulk chasing:** For a given parameter \( 0 < \delta < 1 \), the output \( \hat{A} \subset \tilde{A} \) of EXPAND\([\hat{A}, \tilde{r}, \delta]\) should satisfy

\[
\hat{A} \supset \hat{A}, \quad \|R_{\hat{A}}\tilde{r}\|_{\ell_2(\hat{A})} \geq \delta \|\tilde{r}\|_{\ell_2(\mathcal{J})},
\]

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 \( O(\#\hat{A} + \#\text{supp} \tilde{r}) \).

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

\[
\text{Input: Target tolerance } \varepsilon, \text{ finite index set } \hat{A}_1 \subset \mathcal{J} \text{ and tolerance } \nu_0 \approx \|B^Tf\|_{\ell_2(\mathcal{J})}.
\]

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

1. Set \( w_{\hat{A}_0} := 0 \).
2. for \( k = 1, 2, \ldots \) do
3. \hspace{1em} \( w_{\hat{A}_k} := \text{GALSOLVE}[\hat{A}_k, w_{\hat{A}_{k-1}}, \omega_{ls} : \nu_{k-1}] \).
4. \hspace{1em} \( \tilde{r}_k := \text{RESIDUAL}[w_{\hat{A}_k}, \omega_{ls}] \) and set \( \nu_k := \|\tilde{r}_k\|_{\ell_2} \).
5. \hspace{1em} if \( \nu_k \leq \varepsilon / \|B^{-1}\|_2 \), then return \( u_\varepsilon := w_{\hat{A}_k} \).
6. \hspace{1em} \( \hat{A}_{k+1} := \text{EXPAND}[\hat{A}_k, \tilde{r}_k, \delta] \).
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 $\delta, \omega_{ls}, \gamma_{ls}$ from Algorithm 9.1 hold. Then, the iterates $w_{\Lambda_k}$ produced by LS–AWGM satisfy $\|u - w_{\Lambda_k}\| \leq \rho^{k/2} \|u\|$ where $\rho := 1 - \left(\frac{\delta - \omega_{ls}}{\omega_{ls}}\right)\kappa(B^\top B)^{-1} + \left(\frac{\gamma_{ls}}{\omega_{ls}}\right)^2 \kappa(B^{-1}B) < 1$ and the output $u_\varepsilon$ satisfies $\|u - u_\varepsilon\|_{\ell_2(\mathcal{J})} \leq \varepsilon$. If, moreover, $u \in \mathcal{A}^s$ for some $s > 0$, it holds for $N_k := \# \hat{\Lambda}_k$ that

$$\|u - w_{\Lambda_k}\|_{\ell_2(\mathcal{J})} \lesssim \|u\|_{\mathcal{A}^s N_k^{-s}}, \quad \# \text{supp} u_\varepsilon \lesssim \varepsilon^{-1/s} \|u\|_{\mathcal{A}^s}.$$  

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 GALSOLVE 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 $\text{supp} \psi_\lambda \subset \bigcup_{\mu \in \Lambda : |\mu| = |\lambda| - 1} \text{supp} \psi_\mu$.

It holds for all $\lambda, \mu \in \Lambda$ with $|\mu| = |\lambda| - 1$ and $\text{supp} \psi_\lambda \cap \text{supp} \psi_\mu > 0$ that $S_\mu \supseteq S_\lambda$, where

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

Let us now consider a tensor product wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\} = \{\hat{\Psi}, \tilde{\Psi}\}$ with $\hat{\Psi}$ 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 $\mu_j \in \mathcal{J}^{(i)}$ for $j \neq i$, the index set

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

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 $\mathcal{A}^s$ (see (4.1)) by the constrained approximation space $\mathcal{A}^s_{\text{mtree}} := \{v \in \ell_2(\mathcal{J}) : \|v\|_{\mathcal{A}^s_{\text{mtree}}} < \infty\}$, where $\|v\|_{\mathcal{A}^s_{\text{mtree}}} := \sup_{\varepsilon > 0} \varepsilon \cdot \left\{\min\{N \in \mathbb{N}_0 : \|v - v_N\|_{\ell_2(\mathcal{J})} \leq \varepsilon \wedge \text{supp} v_N \text{ is a multitree}\} \right\}^s$. 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 $b^{(i)}_m$ such that

\[(8.3) \quad B = D^Y \left[ \sum_{m=1}^M \prod_{i=0}^n b^{(i)}_m (\tilde{\Psi}^{(i)}, \tilde{\Psi}^{(i)}) \right] D^X = D^Y \left[ \sum_{m=1}^M \bigotimes_{i=0}^n \tilde{B}^{(i)}_m \right] D^X,
\]

where $\tilde{B}^{(i)}_m := b^{(i)}_m (\tilde{\Psi}^{(i)}, \tilde{\Psi}^{(i)})$ for $i = 0, \ldots, n$ and $m = 1, \ldots, M$. This means that $B$ 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 $b^{(i)}_m$ are local in the sense that $b^{(i)}_m (w, v) = 0$ whenever $|\text{supp } v \cap \text{supp } w| = 0$.

The special structure of $B$ can be used to efficiently realize the application of $\tilde{A}B\tilde{A}$ to a vector $\tilde{v}_\Lambda \in \ell_2(\tilde{\Lambda})$ for finite multitrees $\tilde{\Lambda} \in \tilde{\mathcal{J}}$ and $\tilde{\Lambda} \in \tilde{\mathcal{J}}$. As described in [KS14], this can be realized in linear complexity, i.e., $O(\#\tilde{\Lambda} + \#\tilde{\Lambda})$, by using a separation of $\tilde{A}B\tilde{A}$ 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 $\tilde{A}^{(1)}, \tilde{A}^{(2)}$ and identity matrices $\tilde{I}_{d}^{(1)}, \tilde{I}_{d}^{(2)}$ of appropriate dimension:

\[(8.4) \quad \tilde{A}^{(1)} \otimes \tilde{A}^{(2)} = [\tilde{A}^{(1)} \otimes \tilde{I}_{d}^{(2)}] \circ [\tilde{I}_{d}^{(1)} \otimes \tilde{A}^{(2)}] = [\tilde{I}_{d}^{(1)} \otimes \tilde{A}^{(2)}] \circ [\tilde{A}^{(1)} \otimes \tilde{I}_{d}^{(2)}].
\]

Then we split $\tilde{B}^{(i)}_m = \tilde{L}^{(i)}_m + \tilde{U}^{(i)}_m$ into a (strictly) lower triangular matrix $\tilde{L}^{(i)}_m := [(\tilde{B}^{(i)}_m)_{\lambda,\mu}]_{|\lambda| > |\mu|}$ and an upper triangular matrix $\tilde{U}^{(i)}_m := [(\tilde{B}^{(i)}_m)_{\lambda,\mu}]_{|\lambda| \leq |\mu|}$. With (8.4), it can then be shown that there exist multitrees $\tilde{E}$ and $\tilde{E}$ such that we have the following equivalent representation of $\tilde{A}B\tilde{A}$:

\[
D^Y \left[ \sum_{m=1}^M R_\Lambda \left[ \tilde{I}_{d}^{(0)} \otimes \tilde{B}^{(1)}_m \otimes \cdots \otimes \tilde{B}^{(n)}_m \right] E_{\Pi} \circ R_\Pi \left[ \tilde{U}^{(0)}_m \otimes \tilde{I}_{d}^{(1)} \otimes \cdots \otimes \tilde{I}_{d}^{(n)} \right] E_{\tilde{\Lambda}} \right] =: (I)
\]

\[
+ \sum_{m=1}^M R_\Lambda \left[ \tilde{L}^{(0)}_m \otimes \tilde{I}_{d}^{(1)} \otimes \cdots \otimes \tilde{I}_{d}^{(n)} \right] E_{\Pi} \circ R_\Pi \left[ \tilde{I}_{d}^{(0)} \otimes \tilde{B}^{(1)}_m \otimes \cdots \otimes \tilde{B}^{(n)}_m \right] E_{\tilde{\Lambda}} \right] =: (II)
\]

\[
+ \sum_{m=1}^M R_\Lambda \left[ \tilde{L}^{(0)}_m \otimes \tilde{I}_{d}^{(1)} \otimes \cdots \otimes \tilde{I}_{d}^{(n)} \right] E_{\Pi} \circ R_\Pi \left[ \tilde{U}^{(0)}_m \otimes \tilde{I}_{d}^{(1)} \otimes \cdots \otimes \tilde{I}_{d}^{(n)} \right] E_{\tilde{\Lambda}} \right] =: (III)
\]

It holds that $\#\tilde{E} + \#\tilde{E} \leq \#\tilde{\Lambda} + \#\tilde{\Lambda}$. The application of (II), (III) (and (I), (IV) for $n = 1$) is referred to as unidirectional operation as only the application of the univariate matrices $\tilde{L}^{(0)}_m |_{\tilde{\Lambda}(0) \times \tilde{\Lambda}(0)}$, $\tilde{U}^{(0)}_m |_{\tilde{\Lambda}(0) \times \tilde{\Lambda}(0)}$ and $\tilde{B}^{(1)}_m |_{\tilde{\Lambda}(1) \times \tilde{\Lambda}(1)}$ $(n = 1)$ is required. Due to the tree structure, these tasks can be realized in linear complexity despite the fact that not one of the matrices $\tilde{L}^{(0)}_m$, $\tilde{U}^{(0)}_m$ or $\tilde{B}^{(1)}_m$ 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 $\tilde{B}^{(1)}_m \otimes \cdots \otimes \tilde{B}^{(n)}_m$.

**Theorem 8.4 ([KS14] Theorem 3.1).** Let $A$ be a linear differential operator with polynomial coefficients and let $\Lambda \subset \mathcal{J}$, $\tilde{\Lambda} \in \tilde{\mathcal{J}}$ be multitrees. Then, for any $\tilde{v}_\Lambda \in \ell_2(\tilde{\Lambda})$, the product $\tilde{A}B\tilde{A}v_\Lambda$ can be computed in $O(\#\tilde{\Lambda} + \#\tilde{\Lambda})$ operations.

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

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

**Theorem 8.6** ([KS13]). Let $0 < \omega < 1$, let $A$ be a differential operator with polynomial coefficients and let $u \in A_{\text{mtree}}$ for some $s > 0$. Then, for all finite multitrees $\hat{A} \subset \hat{J}$ and all $w_{\hat{A}} \in \ell_2(\hat{A})$, there exists a multitree $\Xi = \Xi(\hat{A}, \omega) \subset \hat{J}$ such that $\#\Xi \lesssim \#\hat{A} + \nu^{-1/s}$ with $\nu := \|\hat{f}\|_{\ell_2(\hat{J})}$, $f_{\hat{\Xi}} := R_{\hat{\Xi}}f$ and

$$||f - Bw_\hat{A}) - \hat{r}||_{\ell_2(\hat{J})} \leq \omega||\hat{r}||_{\ell_2(\hat{J})},$$

where $\hat{r} := f_{\hat{\Xi}} - \hat{\Xi}B_{\hat{\Xi}}w_{\hat{\Xi}}$.

**Remark 8.7.** Due to the multitree structure of $\hat{A}$ and $\Xi$, the computational cost for computing $\hat{r}$ is $O(\#\hat{A} + \nu^{-1/s})$ if an entry $f_\lambda$ of $f = (f_\lambda)_{\lambda \in \hat{J}}$ 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 $f_\varepsilon$ with $||f - f_\varepsilon||_{\ell_2(\hat{J})} \leq \varepsilon$ and $\#\text{supp} f_\varepsilon \lesssim \varepsilon^{-1/s}$, which is possible if $f$ is sufficiently (piecewise) smooth (see [KS13 §3.4]).

8.3.2. Dual residual. We may now follow [KS13 §1.1] using a wavelet compression of $B$ and $B^\top$. If $A$ is a linear differential operator with polynomial coefficients, it can be shown that for any $0 < \eta < 1$, there exists $B_\eta : \ell_2(\hat{J}) \rightarrow \ell_2(\hat{J})$ such that

$$||B - B_\eta|| \leq \eta, \quad ||B^\top - B_\eta^\top|| \leq \eta,$$

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 $B_\eta^\top$ are boundedly invertible, we obtain the estimate (see Proposition B.2)

$$||B^\top(f - Bw_\hat{A}) - B_\eta^\top\hat{r}||_{\ell_2(\hat{J})} \leq \omega_0||B_\eta^\top\hat{r}||_{\ell_2(\hat{J})},$$

for $\omega_0 = (\eta \frac{1}{1 - \omega} + (||B|| + \eta)\omega)||B_\eta^{-1}||$ so that $\omega_0 \rightarrow 0$ as $\omega \rightarrow 0$ and $\eta \rightarrow 0$. Even though $B_\eta$ and $B_\eta^\top$ 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. Fortunately, the approximate residual $B_\eta^\top\hat{r}$ is not necessarily supported on a multitree. Hence, we define the multitree-based residual

$$\hat{r} := \hat{\Xi}B^{\top}_{\hat{\Xi}}(f_{\hat{\Xi}} - \hat{\Xi}B_{\hat{\Xi}}w_{\hat{\Xi}}) = \hat{\Xi}B_{\hat{\Xi}}\hat{r},$$

such that $||B^\top(f - Bw_\hat{A}) - \hat{r}||_{\ell_2(\hat{J})} \leq \omega_0||\hat{r}||_{\ell_2(\hat{J})}$ where $\hat{\Xi}$ is the smallest multitree containing $\text{supp} B_\eta^\top\hat{r}$. The residual computation requires $O(\#\hat{\Xi} + \#\hat{\Xi})$ operations.

**Remark 8.8.** Theorem 8.6 only ensures the existence of an appropriate multitree $\hat{\Xi}$ but does not give any information on its explicit construction. The same holds true for $\hat{\Xi}$. In Section 8.5 we will discuss how we can construct the multitrees $\hat{\Xi}$ and $\hat{\Xi}$ without setting up the compressed matrix $B_\eta^\top$ so that $\hat{r}$ from (8.8) satisfies (RES).

Furthermore, numerical experiments in Section 8.8 indicate appropriate choices of $\hat{\Xi}$ and $\hat{\Xi}$ 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 (6.9), the approach proposed in [Ste09, SS09] consists of replacing \( \tilde{\Lambda} B^T B \tilde{\Lambda} \) by a sparse approximation \( \tilde{\Lambda} [B^T B] \tilde{\Lambda} := R_{\tilde{\Lambda}} [B^T B] R_{\tilde{\Lambda}} \) satisfying \( \| \tilde{\Lambda} B^T B \tilde{\Lambda} - \tilde{\Lambda} [B^T B] \tilde{\Lambda} \| \lesssim \eta \). In analogy to (6.9), we consider:

(8.9) Find \( u_{\eta,\tilde{\Lambda}} \in \ell_2(\tilde{\Lambda}) : \tilde{\Lambda} [B^T B] \tilde{\Lambda} u_{\eta,\tilde{\Lambda}} = R_{\tilde{\Lambda}} B^T f_{\tilde{\Lambda}} \).

Indeed, under the assumption that \( \eta \) is sufficiently small, \( \kappa(\tilde{\Lambda} [B^T B] \tilde{\Lambda}) \) is bounded independently of \( \tilde{\Lambda} \) (see Appendix B). In particular, there exist algorithms based on linear iterative solvers like the conjugate gradient (cg) method that approximate 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 \( w_{\tilde{\Lambda}} \) as an approximate solution of the problem:

(8.10) Find \( x_{\tilde{\Lambda}} \in \ell_2(\tilde{\Lambda}) : \tilde{\Lambda} B_{\tilde{\Lambda}}^T \tilde{\Lambda} B_{\tilde{\Lambda}} x_{\tilde{\Lambda}} = \tilde{\Lambda} B_{\tilde{\Lambda}}^T f_{\tilde{\Lambda}} \).

We could choose \( \tilde{\Lambda} \) as the smallest multitree that contains \( \text{supp} B_{\eta} \tilde{\Lambda} \) for all \( \tilde{\Lambda} \in \ell_2(\tilde{\Lambda}) \). However, this is not an implementable approach. Hence, we are concerned with the question how the multitree \( \tilde{\Lambda} \) can be constructed in dependency of \( \tilde{\Lambda} \) such that (1) the condition number of \( \tilde{\Lambda} B_{\eta} \tilde{\Lambda} \) is uniformly bounded and (2) an approximate solution \( w_{\tilde{\Lambda}} \) to (8.10) satisfies \( \text{GAL} \). This will be discussed in Sections 8.3 and 9. For fixed multitrees, the solution of (8.10) can be computed e.g. with cg.

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

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

(8.11) \( \tilde{\Lambda}_0 = \tilde{\Lambda}_{SG,J} := \{ \lambda \in \tilde{J} : |\lambda| \leq J \} \),

(8.12) \( \tilde{\Lambda}_0 = \tilde{\Lambda}_{SG,J} := \{ \lambda \in \tilde{J} : |\lambda| \leq J \) or \(|\lambda_0| = J + 1, |\lambda_i| = 0, 1 \leq i \leq n \} \),

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 \( \tilde{\Lambda}_k, k > 0 \), we propose the following (heuristic) choices:

(i) \( \tilde{\Lambda}_{\text{Full}} = \text{FullStableExpansion}(\tilde{\Lambda}, \ell) \) is defined as

(8.13) \( \tilde{\Lambda}_{\text{Full}} := \{ \lambda \in \tilde{J} : \exists \mu \in \tilde{\Lambda} \) s.t. for all \( j = 0, \ldots, n \) : \(|\lambda_j| \leq |\mu_j| + \ell \)

and \( \text{dist}(\text{supp} \tilde{\psi}_{\lambda_j}, \text{supp} \tilde{\psi}_{\mu_j}) \leq D \nu_j 2^{-|\lambda_j|} \).

\(^2\text{We will also use } \tilde{\Lambda}_{SG,J} \text{ and } \tilde{\Lambda}_{SG,J} \text{ within a uniform sparse grid (SG) discretization.}\)
Algorithm 3 \( \hat{\Lambda} = \text{FullStableExpansion}[\hat{\Lambda}, \ell] \)

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

1: \( \hat{\Lambda} := \emptyset \subset \hat{\mathcal{J}} \).
2: **for** \( \lambda = (\lambda_0, \ldots, \lambda_n) \in \hat{\Lambda} \) **do**
3: \( \text{Find all “neighbours” } \mu = (\mu_0, \ldots, \mu_n) \in \hat{\mathcal{J}} \text{ on the same level:} \)
4: \( \hat{\Lambda} \leftarrow \hat{\Lambda} \cup \{ \mu \in \hat{\mathcal{J}} : |\lambda_i| = |\mu_i|, \supp \hat{\psi}_{\mu_i} \cap \supp \hat{\psi}_{\lambda_i} = \emptyset \text{ for all } i = 0, \ldots, n \}. \)
5: **end for**
6: \( \text{Complete } \hat{\Lambda} \text{ to form a multitree in the sense of Definition [8.2]} \)

**Choice of sets** \( \hat{\mathcal{E}}_k, \hat{\mathcal{E}}_\ell \). The proposed index set reads

\[
\hat{\mathcal{E}}_k = \text{ReducedMultiTreeCone}(\hat{\Lambda}, \ell)
\]

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

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 \( \hat{\psi}^X = \hat{\psi}^Y \), \( \hat{\mathcal{E}} = \hat{\mathcal{E}} \).

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 \( \mathcal{Y} \)), we expand \( \hat{\Lambda}_k \) to \( \hat{\mathcal{E}}_{k}^{\text{tmp}} = \text{ReducedMultiTreeCone}(\hat{\Lambda}_k, \ell) \) and obtain the desired \( \hat{\mathcal{E}}_k \) by one of the expansion variants in [8.13]–[8.15]. For the dual residual (in \( \mathcal{X} \)), we consider two approaches. In the first one, shown in Figure 1(a), we take the set \( \hat{\mathcal{E}}_k \) as above and set \( \hat{\mathcal{E}}_k = \text{FullStableExpansion}(\hat{\mathcal{E}}_k, \ell) \) (with obvious inverted roles of primal and dual basis). Then, \( \hat{\mathcal{E}}_k \) is the smallest multitree containing \( \supp \hat{\mathcal{B}}_j \), \( \hat{\mathcal{R}}_k \) for sufficiently small \( \eta \). The second approach uses the by far smaller set \( \hat{\mathcal{E}}_k = \hat{\mathcal{E}}_{k}^{\text{tmp}} \) as indicated in Figure 1(b) [KS13].
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 $q_{\Lambda_k}$. 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 (5.14), (5.15) are independent of $n$. In particular, the condition numbers of $B^T B$ and of $\Lambda_k B_{\Lambda_k}^T B_{\Lambda_k} \Lambda_k$ 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 $\delta = 0.7$ and $\gamma_{ls} = 0.01$ 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 [Di] 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

$$
\left\{
\begin{array}{ll}
u_t - \nu_{xx} = f(t, x) & \text{on } \Omega = (0,1), \\
u(t, 0) = u(t, 1) = 0 & \text{for all } t \in [0,T], \\
u(0, x) = u(T, x) & \text{on } \overline{\Omega},
\end{array}
\right.
$$

with a discontinuous source function $f(t, x) \equiv f(t) := K \left( \frac{Nt}{T} - \lfloor \frac{Nt}{T} \rfloor \right)$, $N \in \mathbb{N}$, $K \in \mathbb{R}_+$. 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

\[3\]We have chosen a larger value for $\delta$ than required by Algorithm 2 for efficiency reasons.

\[4\]Note that these bases cannot be normalized to be a Riesz basis of $H^{-1}(\Omega)$.
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 \( \Lambda_k = \hat{\Lambda}_{\text{Full}} \) yields stability. Figure 2(c) shows the cardinalities of the test sets. They grow only linearly with \( |p| \), so that both \( w_{\hat{\Lambda}_k} \) 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 \( \hat{r}_k \), we monitor only the dual residual. Since using a larger index set, \( \| \hat{r}_k \|_{\ell_2(P)} \) 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 \( \hat{\Xi}_k = \text{TemporalStableExpansion}(\hat{\Xi}^{\text{tmp}}, 1) \) yields results that are comparable to the other extensions, which could not be deduced from [KS13,KST14]. All three methods seem stable, and we can reduce the size of the test sets by a factor of about 3.4 for \( \hat{\Lambda}_k \) (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

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

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 \( \text{supp } \psi_{\lambda}, \lambda \in \hat{A}_k \)) in Figure 5(b) indicate that the AWGM benefits from its ability to refine not only independently in each dimension, but in particular locally in the full space-time domain. This is also mirrored in Figure 6(a), where we observe the optimal \( s_{\max} = 1 \) 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 \( \hat{A}_k, \hat{\Xi}_k \). 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 FullResConstruction for \( \Xi_k \). 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 \( \hat{\Xi}_k \) yields index sets that are approximately 20 times as large as for the optimized version (cf. Figure 7(c)), with only a slight improvement in the residual approximation.

**Appendix A. Proof of Proposition 2.1**

We follow [SS09] to verify the Babuška-Aziz conditions in a time-periodic setting.

1. **Continuity.** Follows from (2.2), the definitions of \( \|\cdot\|_X, \|\cdot\|_Y \) as well as Cauchy-Schwarz’s, Hölder’s and Young’s inequalities.

2. **Inf-sup condition.** We consider an arbitrary \( 0 \neq w \in X \) and define \( z_w(t) := (A(t)^{-1} \dot{w}(t) \) for the adjoint \( A(t)^* \) 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
\[
\int_0^T \langle \dot{w}, \dot{w} \rangle_{V \times V'} dt = \int_0^T \langle \dot{w}, w \rangle_{V \times V'} dt + \int_0^T \langle \dot{w}, w \rangle_{V \times V'} = \int_0^T \frac{d}{dt} \|w(t)\|_H^2 dt = \|w(T)\|_H^2 - \|w(0)\|_H^2 = 0,
\]
so that we finally get
\[
b(w, v_w) \geq \alpha \min\{1, \gamma^{-2}\} \|w\|_X^2 \geq \frac{\alpha \min\{1, \gamma^{-2}\}}{\sqrt{2} \max\{1, \alpha^{-1}\}} \|w\|_X \|v_w\|_Y > 0.
\]

3. **Surjectivity.** Let \( \neq v \in Y \). 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 \( b(z, v) = \int_0^T \langle v(t), A(t)[v(t)] \rangle_{V \times V'} \geq \alpha \|v\|_{J}^2 > 0 \), so that the surjectivity condition is fulfilled.

(i) \textit{Faedo-Galerkin approximation of an initial value problem.} Let \( \{\phi_i : i \in \mathbb{N}\} \) be a basis for \( V \), \( V_n := \text{span}\{\phi_i, i = 1, \ldots, n\} \), \( z_n(t) := \sum_{i=1}^n z_i^{(n)}(t) \phi_i \). Then the linear system of ODEs \( \langle w_n, \dot{z}_n(t) \rangle_{V \times V'} + \langle w_n, A(t)[z_n(t)] \rangle_{V \times V'} = \langle w_n, A(t)[v(t)] \rangle_{V \times V'} \), \( z_n(0) = z_{n0} \), 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) \( z_0 \in H \) and its orthogonal projection \( z_{n0} \) onto \( V_n \).

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

\[
\frac{1}{2} \frac{d}{dt} \|z_n(t)\|_H^2 + \alpha \|z_n(t)\|_V^2 \leq \gamma \varepsilon \|z_n(t)\|_V + \frac{\gamma}{4 \varepsilon} \|v(t)\|_V^2,
\]

and hence by integration over \( [0, s], s \in [0, T] \), using \( (\alpha - \gamma \varepsilon) > 0 \) that \( \|z_n(s)\|_H^2 - \|z_n(0)\|_H^2 = \int_0^s \frac{d}{dt} \|z_n(t)\|_H^2 dt \leq \frac{\gamma}{2 \varepsilon} \int_0^s \|v(t)\|_V^2 dt \), so that sup \( s \in [0, T] \|z_n(s)\|_H < \infty \) and \( \{z_n\}_{n \in \mathbb{N}} \) is uniformly bounded in \( L^\infty(0, T; H) \). Similarly, we can conclude that

\[
2(\alpha - \gamma \varepsilon) \|z_n\|_V \leq \|z_n(0)\|_H - \|z_n(T)\|_H^2 + \frac{\gamma}{2 \varepsilon} \|v\|_V^2 < \infty,
\]

so that \( \{z_n\}_{n \in \mathbb{N}} \) is also uniformly bounded in \( \mathcal{V} \).

(iii) \textit{Periodicity.} Abbreviate \( \check{c} := \frac{\gamma}{4 \varepsilon}, \check{\alpha} := \frac{2(\alpha - \gamma \varepsilon)}{c_1} > 0 \) with \( c_1 := \sup_{\phi \in V} \|\phi\|_V \), and multiply (A.1) by \( e^{\check{\alpha} t} \). Then \( \frac{d}{dt} \left( e^{\check{\alpha} t} \|z_n(t)\|_H^2 \right) = e^{\check{\alpha} t} \frac{d}{dt} \|z_n(t)\|_H^2 + e^{\check{\alpha} t} \check{\alpha} \|z_n(t)\|_H^2 \leq e^{\check{\alpha} t} \|v(t)\|_V^2 \), and by integration over \( [0, T] \), we obtain

\[
\|z_n(T)\|_H^2 \leq e^{-\check{\alpha} T} \|z_n(0)\|_H^2 + \check{c} e^{-\check{\alpha} T} \int_0^T e^{\check{\alpha} t} \|v(t)\|_V^2 dt.
\]

Set \( M := \{z \in V_n : \|z\|_H \leq R := K\check{c} (1 - e^{-\check{\alpha} T})^{-\frac{1}{2}} \}, \ K := \check{c} e^{-\check{\alpha} T} \int_0^T e^{\check{\alpha} t} \|v(t)\|_V^2 dt \). The set \( M \) is convex and compact in \( V_n \). If \( z_n(0) \in M \), (A.2) implies that \( \|z_n(T)\|_H^2 \leq e^{-\check{\alpha} T} R^2 + K \leq R \), i.e., \( z_n(T) \in M \). Since by Gronwall’s lemma the mapping \( S : M \rightarrow M, z_n(0) \mapsto z_n(T) \), is continuous, the existence of a fixed-point \( S(\tilde{z}_n) = \tilde{z}_n \in M \) follows from Brouwer’s fixed-point theorem. By the a priori estimates, the sequence \( \{\tilde{z}_n\}_{n \in \mathbb{N}} \) is bounded in \( H \), so that there exists a subsequence (also denoted by \( \{\tilde{z}_n\} \)) converging weakly to some \( \tilde{z} \in H \).

(iv) \textit{Convergence.} Consider the periodic solution \( z_n(t) \) from (iii), i.e., the solution of the ODE system with initial value \( z_{n0} = \tilde{z}_n \). From the a priori estimates, we have that \( \{z_n\} \) is uniformly bounded in the separable space \( \mathcal{V} \), so that there exists a subsequence (also denoted \( \{z_n\} \)) converging weakly to some \( z \) in \( \mathcal{V} \). 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 \( j = 1, \ldots, n - \langle \dot{\phi}_j, z_n \rangle = \langle \theta(0) \phi_j, -\theta(T) \phi_j, \tilde{z}_n \rangle_H + \langle \theta \phi_j, A(t)[v - z_n] \rangle_H \). As \( z_n \rightharpoonup z \) in \( \mathcal{V} \) and \( \tilde{z}_n \rightharpoonup \tilde{z} \) in \( H \), we can pass to the limit \( n \rightarrow \infty \) and obtain

\[
\langle \dot{\phi}_j, z \rangle = \langle \theta(0) \phi_j - \theta(T) \phi_j, \tilde{z} \rangle_H + \langle \theta \phi_j, A(t)[v - z] \rangle_H.
\]

This particularly holds true for all \( \theta \in \mathcal{D}(I) \), so that \( \tilde{z} = \mathcal{A}(\cdot)(v - z) \) in the distributional sense and hence \( \tilde{z} \in L^2(0, T; V') \). Moreover, (A.3) implies that for \( w \in C^1(0, T; V) \), we have \( -\langle \dot{w}, z \rangle - \langle w(0) - w(T), \tilde{z} \rangle = \langle w, A(t)[v - z] \rangle \), \( -\langle \ddot{w}, z \rangle + \langle w(T), z(T) \rangle_H - \langle w(0), z(0) \rangle_H \), so that indeed \( \tilde{z} = z(0) = z(T) \) in \( H \) and hence \( z \in \mathcal{X} \). With this \( z \), the surjectivity condition is fulfilled.
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 $\eta < 1$, $B_\eta \in L(\ell_2(\mathcal{J}), \ell_2(\mathcal{J}))$ and $B_\eta^\top B_\eta \in L(\ell_2(\mathcal{J}), \ell_2(\mathcal{J}))$ 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
\[ \| B_\eta^\top (\mathbf{f} - Bw_\mathbf{A}) - \mathbf{r} \|_{\ell_2(\mathcal{J})} \leq \omega_{ls} \| \mathbf{r} \|_{\ell_2(\mathcal{J})}, \quad \mathbf{r} := B_\eta^\top \mathbf{r}. \]

REFERENCES


Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany
E-mail address: sebastian.kestler@gmail.com

Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany
E-mail address: kristina.steih@uni-ulm.de

Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany
E-mail address: karsten.urban@uni-ulm.de