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 (PDE) of the form

$$u_t + \mathcal{A}(t)[u] = g(t) \qquad \text{on } \Omega \subset \mathbb{R}^n, \text{ for } t \in [0, T],$$

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, UP13]. 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 $\mathbf{B}\mathbf{u} = \mathbf{f}$ 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 $\mathbf{C}\mathbf{u} = \mathbf{g}$ in ℓ_2 with a *symmetric* positive definite (s.p.d.) stiffness matrix $\mathbf{C} : \ell_2 \to \ell_2$, an infinite right-hand side $\mathbf{g} \in \ell_2$ and a unique solution $\mathbf{u} \in \ell_2$ 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 $\mathbf{\Lambda}_k$ steering the local refinement. This means that a finite vector $\mathbf{u}_{\mathbf{\Lambda}_k}$ satisfying $\mathbf{C}|_{\mathbf{\Lambda}_k \times \mathbf{\Lambda}_k} \mathbf{u}_{\mathbf{\Lambda}_k} = \mathbf{g}|_{\mathbf{\Lambda}_k}$ is computed. The (infinite-dimensional) residual $\mathbf{r}_{\mathbf{\Lambda}_k} := \mathbf{g} - \mathbf{C}\mathbf{u}_{\mathbf{\Lambda}_k}$ is then approximated to serve as an error estimator and to identify an update, i.e. an (usually refined) index set $\mathbf{\Lambda}_{k+1}$ (typically $\mathbf{\Lambda}_{k+1}$ corresponds to the significant coefficients of $\mathbf{r}_{\mathbf{\Lambda}_k}$).

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 $\mathbf{C} := \mathbf{B}^\top \mathbf{B}$ and right-hand side $\mathbf{g} = \mathbf{B}^\top \mathbf{f}$, is a natural approach for initial value problems (e.g. [CS11, SS09]) and, as well we will show in this article, also for time-periodic problems.

The treatment of normal equations by adaptive wavelet methods has first been discussed in [CDD02]. The main difficulty lies in the (approximate) evaluation of $\mathbf{B}^\top \mathbf{B}$ and $\mathbf{B}^\top \mathbf{f}$. 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 *multitree-structured* index sets as introduced in [KS12, KS13]. This allows for the *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 *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 multitree-structured index sets that permit a tree structure when all but one coordinate directions are frozen.

We stress that, to the best of our knowledge, *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 *elliptic* operator problems are presented in [KS12, KS13]. 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 ℓ_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, multitree-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 $\boldsymbol{\Omega} := \Omega_1 \times \cdots \times \Omega_n \subset \mathbb{R}^n$ be a product domain and V be a real separable Hilbert space with dual V' such that $V \hookrightarrow H := L_2(\boldsymbol{\Omega}) \hookrightarrow V'$ is a Gelfand triple. For $\mathcal{A}(t) \in \mathcal{L}(V, V')$ and $g \in L_2(0, T; V')$ we consider the time-periodic equation

$$(2.1) \qquad u_t + \mathcal{A}(t)[u(t)] = g(t) \text{ in } V' \text{ for a.e. } t \in [0, T], \qquad u(0) = u(T) \text{ in } H.$$

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

$$(2.2) \quad \langle v, \mathcal{A}(t)[w] \rangle_{V \times V'} \le \gamma \|w\|_V \|v\|_V, \quad \langle v, \mathcal{A}(t)[v] \rangle_{V \times V'} \ge \alpha \|v\|_V^2 \quad \forall v, w \in V.$$

Moreover, we assume that the space V is a Sobolev space of nonnegative order

$$(2.3) \qquad V := \bigcap_{i=1}^n \bigotimes_{j=1}^n W_{ij}, \text{ where } W_{ij} := \left\{ \begin{array}{ll} L_2(\Omega_i), & i \ne j, \\ V^{(i)}, & i = j, \end{array} \right.$$

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 $V = H_0^1(\boldsymbol{\Omega})$ and $V^{(i)} = H_0^1(\Omega_i)$ (see [GO95]).

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

$$(2.4) \qquad H_{\text{per}}^1(0, T) := \{v \in H^1(0, T) : v(0) = v(T)\},$$

we consider the spaces $\mathcal{Y} := L_2(0, T; V)$ and $\mathcal{X} := L_2(0, T; V) \cap H^1_{\mathrm{per}}(0, T; V')$, i.e.,

$$(2.5) \qquad \mathcal{X} = \{v \in L_2(0, T; V) : v_t \in L_2(0, T; V'), \ v(0) = v(T) \text{ in } H\},$$

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

$$(2.6) \qquad \qquad \text{Find } u \in \mathcal{X}: \qquad b(u,v) = f(v) \qquad \forall v \in \mathcal{Y},$$

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)]

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

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 $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathcal{Y} \times \mathcal{Y}'}$, so that (2.6) is a variational formulation of the operator equation:

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

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:* $\gamma_{\mathcal{B}} := \sup_{0 \neq u \in \mathcal{X}} \sup_{0 \neq v \in \mathcal{Y}} \frac{b(u,v)}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} < \infty.$

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

(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, $\mathcal{B}$ from* (2.8) *is boundedly invertible with $\|\mathcal{B}\| = \gamma_{\mathcal{B}} = \sqrt{2} \max\{1, \gamma\}$, $\|\mathcal{B}^{-1}\| = \frac{1}{\beta_{\mathcal{B}}} = \frac{\sqrt{2} \max\{1, \alpha^{-1}\}}{\alpha \min\{1, \gamma^{-2}\}}.$*

3. Equivalent bi-infinite matrix-vector problem

We consider the reformulation of (2.8) as an *equivalent ℓ_2-problem*, i.e., a *discrete* problem posed on the sequence space ℓ_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 $\Upsilon := \{\gamma_i : i \in \mathbb{N}\} \subset \mathcal{H}$ 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

$$(3.1) \qquad c\|\mathbf{v}\|_{\ell_2(\mathbb{N})}^2 \leq \|v\|_{\mathcal{H}}^2 \leq C\|\mathbf{v}\|_{\ell_2(\mathbb{N})}^2 \quad \forall \mathbf{v} = (v_i)_{i \in \mathbb{N}} \in \ell_2(\mathbb{N}).$$

The largest c and the smallest C for which (3.1) holds, are referred to as *lower* and *upper Riesz constant* and are denoted by $c_{\Upsilon}(\mathcal{H})$ and $C_{\Upsilon}(\mathcal{H})$, respectively.[1]

---

[1]Sometimes a different definition of Riesz constants is used, namely $c_{\Upsilon}(\mathcal{H})$ and $C_{\Upsilon}(\mathcal{H})$ being the largest and the smallest constant such that $c_{\Upsilon}(\mathcal{H})\|\mathbf{v}\|_{\ell_2(\mathbb{N})} \leq \|v\|_{\mathcal{H}} \leq C_{\Upsilon}(\mathcal{H})\|\mathbf{v}\|_{\ell_2(\mathbb{N})}.$

3.2. Wavelet discretization of the parabolic operator problem.

Let us now consider two *different* Riesz bases

$$(3.2) \qquad \widehat{\boldsymbol{\Psi}}^{\mathcal{X}} := \{\widehat{\psi}_{\boldsymbol{\lambda}}^{\mathcal{X}} : \boldsymbol{\lambda} \in \widehat{\mathcal{J}}\} \subset \mathcal{X}, \qquad \check{\boldsymbol{\Psi}}^{\mathcal{Y}} := \{\check{\psi}_{\boldsymbol{\lambda}}^{\mathcal{Y}} : \boldsymbol{\lambda} \in \check{\mathcal{J}}\} \subset \mathcal{Y},$$

labeled w.r.t. two (possibly) different countable index sets $\widehat{\mathcal{J}}$ and $\check{\mathcal{J}}$. More precisely, we consider a *trial* basis $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ for the ansatz space $\mathcal{X}$ and a *test* basis $\check{\boldsymbol{\Psi}}^{\mathcal{Y}}$ for the test space $\mathcal{Y}$ with associated Riesz constants $c_{\mathcal{X}}(\widehat{\boldsymbol{\Psi}}^{\mathcal{X}})$, $C_{\mathcal{X}}(\widehat{\boldsymbol{\Psi}}^{\mathcal{X}})$ and $c_{\mathcal{Y}}(\check{\boldsymbol{\Psi}}^{\mathcal{Y}})$, $C_{\mathcal{Y}}(\check{\boldsymbol{\Psi}}^{\mathcal{Y}})$. It is important to note that $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$, $\check{\boldsymbol{\Psi}}^{\mathcal{Y}}$ arise from normalizing *different* Riesz bases $\widehat{\boldsymbol{\Psi}}$, $\check{\boldsymbol{\Psi}}$, for $L_2((0,T) \times \boldsymbol{\Omega})$ w.r.t. $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$ (which is also the reason for our notation, see Section 5). Then there exists a *unique* expansion $u = \mathbf{u}^{\top} \widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ of the solution u of (2.8) where we formally interpret both $\mathbf{u} \in \ell_2(\widehat{\mathcal{J}})$ and $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ as column vectors. Now, the equivalent formulation of (2.8) reads as follows:

$$(3.3) \qquad \text{Find } \mathbf{u} \in \ell_2(\widehat{\mathcal{J}}) : \qquad \mathbf{B}\mathbf{u} = \mathbf{f}, \qquad \mathbf{f} \in \ell_2(\check{\mathcal{J}}),$$

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

$$(3.4) \quad \|\mathbf{B}\| \le \|\mathcal{B}\|_{\mathcal{X} \to \mathcal{Y}'} C_{\mathcal{X}}(\widehat{\boldsymbol{\Psi}}^{\mathcal{X}})^{\frac{1}{2}} C_{\mathcal{Y}}(\check{\boldsymbol{\Psi}}^{\mathcal{Y}})^{\frac{1}{2}}, \quad \|\mathbf{B}^{-1}\| \le \frac{\|\mathcal{B}^{-1}\|_{\mathcal{Y}' \to \mathcal{X}}}{c_{\mathcal{X}}(\widehat{\boldsymbol{\Psi}}^{\mathcal{X}})^{\frac{1}{2}} c_{\mathcal{Y}}(\check{\boldsymbol{\Psi}}^{\mathcal{Y}})^{\frac{1}{2}}}.$$

3.3. Further notations.

We need to restrict the bi-infinite matrices $\mathbf{B}$ and $\mathbf{B}^{\top}$ in both rows and columns. For a pair $(\boldsymbol{\Lambda}, \mathcal{J})$ with $\mathcal{J} \in \{\widehat{\mathcal{J}}, \check{\mathcal{J}}\}$ and $\boldsymbol{\Lambda} \subseteq \mathcal{J}$, set

$$(3.5) \qquad \mathbf{E}_{\boldsymbol{\Lambda}} : \ell_2(\boldsymbol{\Lambda}) \to \ell_2(\mathcal{J}), \quad \text{and} \quad \mathbf{R}_{\boldsymbol{\Lambda}} := \mathbf{E}_{\boldsymbol{\Lambda}}^{\top} : \ell_2(\mathcal{J}) \to \ell_2(\boldsymbol{\Lambda}),$$

where $\mathbf{E}_{\boldsymbol{\Lambda}}$ is the trivial embedding, i.e., the extension of $\mathbf{v}_{\boldsymbol{\Lambda}} \in \ell_2(\boldsymbol{\Lambda})$ by zeros to $\ell_2(\mathcal{J})$. Consequently, its adjoint $\mathbf{R}_{\boldsymbol{\Lambda}}$ is the restriction of $\mathbf{v} \in \ell_2(\mathcal{J})$ to $\mathbf{v}|_{\boldsymbol{\Lambda}} \in \ell_2(\boldsymbol{\Lambda})$. For $\widehat{\boldsymbol{\Lambda}} \subseteq \widehat{\mathcal{J}}$ and $\check{\boldsymbol{\Lambda}} \subseteq \check{\mathcal{J}}$, we define the following restriction of $\mathbf{B}$ and $\mathbf{B}^{\top}$:

$$(3.6) \quad {}_{\check{\boldsymbol{\Lambda}}}\mathbf{B}_{\widehat{\boldsymbol{\Lambda}}} := \mathbf{R}_{\check{\boldsymbol{\Lambda}}} \mathbf{B} \mathbf{E}_{\widehat{\boldsymbol{\Lambda}}}, \quad \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}} := {}_{\check{\mathcal{J}}}\mathbf{B}_{\widehat{\boldsymbol{\Lambda}}}, \quad {}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}_{\check{\boldsymbol{\Lambda}}}^{\top} := \mathbf{R}_{\widehat{\boldsymbol{\Lambda}}} \mathbf{B}^{\top} \mathbf{E}_{\check{\boldsymbol{\Lambda}}}, \quad {}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^{\top} := {}_{\widehat{\mathcal{J}}}\mathbf{B}_{\check{\boldsymbol{\Lambda}}}^{\top}.$$

Finally, $C \lesssim D$ means that C can be bounded by a constant times D and $C \gtrsim D$ is defined as $D \lesssim C$. In this setting, $C \eqsim D$ 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 $\mathcal{N}$-term approximation.

For a given number of degrees of freedom (d.o.f.) $\mathcal{N} \in \mathbb{N}$, the best approximation $v_{\mathcal{N}}$ of a function $v = \mathbf{v}^{\top} \widehat{\boldsymbol{\Psi}}^{\mathcal{X}} \in \mathcal{X}$ in the basis $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ with $\mathcal{N}$ d.o.f. is a *nonlinear, best $\mathcal{N}$-term approximation* (e.g. [DeV98]), i.e., $v_{\mathcal{N}} = \arg \sigma_{\mathcal{N}}(v)$, where the best $\mathcal{N}$-term approximation error is defined as

$$\sigma_{\mathcal{N}}(v) := \inf_{\{\widehat{\boldsymbol{\Lambda}} \in \widehat{\mathcal{J}} : \#\widehat{\boldsymbol{\Lambda}} = \mathcal{N}\}} \inf_{\{v_{\mathcal{N}} \in \text{span}\{\widehat{\psi}_{\boldsymbol{\lambda}}^{\mathcal{X}} : \boldsymbol{\lambda} \in \widehat{\boldsymbol{\Lambda}}\}\}} \|v - v_{\mathcal{N}}\|_{\mathcal{X}}.$$

Since $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ is a Riesz basis, it holds that $\|\mathbf{v} - \mathbf{v}_{\mathcal{N}}\|_{\ell_2} \eqsim \sigma_{\mathcal{N}}(v)$ where $\mathbf{v}_{\mathcal{N}}$ always denotes an $\mathcal{N}$-term approximation of the vector $\mathbf{v}$ (i.e., the $\mathcal{N}$ largest coefficients in modulus of $\mathbf{v}$). As described in [DeV98], it is meaningful to collect all vectors

$\mathbf{v} \in \ell_2(\widehat{\mathcal{J}})$ that permit an *approximation rate* $s > 0$ in the sense that $\|\mathbf{v} - \mathbf{v}_{\mathcal{N}}\|_{\ell_2} \lesssim \mathcal{N}^{-s}$ within the *nonlinear approximation class* (compare [Ste09, (2)]):

$$(4.1) \quad \mathcal{A}^s := \Big\{ \mathbf{v} \in \ell_2(\widehat{\mathcal{J}}) : \|\mathbf{v}\|_{\mathcal{A}^s} := \sup_{\varepsilon > 0} \varepsilon \big[\min\{\mathcal{N} \in \mathbb{N}_0 : \|\mathbf{v} - \mathbf{v}_{\mathcal{N}}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \varepsilon\} \big]^s < \infty \Big\}.$$

For a given $\mathbf{v} \in \mathcal{A}^s$ and $\varepsilon > 0$, the required number of degrees of freedom $\mathcal{N}_\varepsilon$ in order to obtain $\|\mathbf{v} - \mathbf{v}_{\mathcal{N}_\varepsilon}\|_{\ell_2} \leq \varepsilon$ is bounded by $\mathcal{N}_\varepsilon \leq \varepsilon^{-1/s} \|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}$. 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 $\mathbf{u} \in \mathcal{A}^s$ for some $s > 0$ and that we want to approximate it with a target tolerance $\varepsilon > 0$. The *benchmark* is given by a best $\mathcal{N}_\varepsilon$-term approximation $\mathbf{u}_{\mathcal{N}_\varepsilon}$ satisfying $\sigma_{\mathcal{N}_\varepsilon}(\mathbf{u}) = \|\mathbf{u} - \mathbf{u}_{\mathcal{N}_\varepsilon}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \varepsilon$ which is, however, in general *not* computable. So, we need to focus on the computation of a *quasi-optimal* approximation $\mathbf{u}_\varepsilon$:

(O1) *Convergence rate*: $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(\widehat{\mathcal{J}})} \leq \varepsilon$ and $\# \operatorname{supp} \mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s} \|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}$.

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

In order to realize (O2), we require the wavelet bases $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ and $\widecheck{\boldsymbol{\Psi}}^{\mathcal{Y}}$ 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]),

$$(5.1) \qquad \mathcal{X} \eqsim \big[L_2(0,T) \otimes V \big] \cap \big[H^1_{\mathrm{per}}(0,T) \otimes V' \big], \quad \mathcal{Y} \eqsim L_2(0,T) \otimes V.$$

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

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

$$(5.2) \qquad \Psi = \bigcup_{j \in \mathbb{N}_0} \Psi_j = \{\psi_\lambda : \lambda = (j,k) \in \mathcal{J}\} \subset \mathcal{H},$$

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 $\operatorname{diam}(\operatorname{supp} \psi_{j,k}) \eqsim 2^{-j}$) and k is a *translation index* indicating the *position* of $\operatorname{supp} \psi_{j,k}$. Note that the elements of Ψ_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 Ψ (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 Ψ is a *uniformly local, piecewise polynomial wavelet basis of order $d \in \mathbb{N}$*, i.e.:

(W1) Local supports: $\operatorname{diam}(\operatorname{supp} \psi_\lambda) \eqsim 2^{-|\lambda|}$ 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 $\sup_{\lambda \in \mathcal{J}_j} \#\{\lambda' \in \mathcal{J}_j : |\operatorname{supp} \psi_\lambda \cap \operatorname{supp} \psi_{\lambda'}| > 0\} \leq C$.

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

Furthermore, we assume that the projection $Q_j[v] := \sum_{\{\lambda \in \mathcal{J} : |\lambda| < j\}} v_\lambda \psi_\lambda$ for $v = \sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda$ satisfies the following Jackson estimates $\|\mathrm{Id} - Q_j\|_{H^d(\Omega) \cap \mathcal{H} \to L_2(\Omega)} \lesssim 2^{-dj}$, $\|\mathrm{Id} - Q_j\|_{H^d(\Omega) \cap \mathcal{H} \to \mathcal{H}} \lesssim 2^{-(d-m)j}$, $\|\mathrm{Id} - Q_j\|_{H^d(\Omega) \cap \mathcal{H} \to \mathcal{H}'} \lesssim 2^{-(d+m)j}$, where $m = 1$ if $\mathcal{H} = H^1_{\mathrm{per}}(0, T)$.

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

$$(5.3) \qquad \Theta^{\mathrm{per}} := \left\{ \theta_\lambda^{\mathrm{per}} : \lambda \in \mathcal{J}_t^{\mathrm{per}} \right\} \subset H^1_{\mathrm{per}}(0, T)$$

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 $c_{L_2}(\Theta^{\mathrm{per}})$, $C_{L_2}(\Theta^{\mathrm{per}})$. We assume that the elements of Θ^{per} are sufficiently smooth so that the *properly normalized* collection $\left\{ \theta_\lambda^{\mathrm{per}} / \|\theta_\lambda^{\mathrm{per}}\|_{H^1} : \lambda \in \mathcal{J}_t^{\mathrm{per}} \right\}$ is a Riesz basis for $H^1_{\mathrm{per}}(0, T)$ with constants $c_{H^1_{\mathrm{per}}}(\Theta^{\mathrm{per}})$, $C_{H^1_{\mathrm{per}}}(\Theta^{\mathrm{per}})$. 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 $L_2(0, T)$,

$$(5.4) \qquad \Theta := \left\{ \vartheta_\lambda : \lambda \in \mathcal{J}_t \right\},$$

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 $\boldsymbol{\Omega} = \Omega_1 \times \cdots \times \Omega_n$ is a product domain. Here, we shall use that V is the (intersection of) tensor products of univariate Sobolev spaces (see (2.3)) with $L_2(\boldsymbol{\Omega}) \subseteq V$ and $L_2(\boldsymbol{\Omega}) \approx L_2(\Omega_1) \otimes \cdots \otimes L_2(\Omega_n)$ (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 $L_2(\Omega_i)$, $\Sigma^{(i)} := \{ \sigma_\lambda^{(i)} : \lambda \in \mathcal{J}^{(i)} \} \subset V^{(i)}$. We require that these functions are sufficiently smooth so that $\{ \sigma_\lambda^{(i)} / \|\sigma_\lambda^{(i)}\|_{V^{(i)}} : \lambda \in \mathcal{J}^{(i)} \}$, $\{ \sigma_\lambda^{(i)} / \|\sigma_\lambda^{(i)}\|_{V^{(i)'}} : \lambda \in \mathcal{J}^{(i)} \}$ are Riesz bases for $V^{(i)}$, $V^{(i)'}$ with constants $c_{V^{(i)}}(\Sigma^{(i)})$, $C_{V^{(i)}}(\Sigma^{(i)})$ and $c_{V^{(i)'}}(\Sigma^{(i)})$, $C_{V^{(i)'}}(\Sigma^{(i)})$. Now,

$$(5.5) \qquad \boldsymbol{\Sigma} := \left\{ \boldsymbol{\sigma_\lambda} : \boldsymbol{\lambda} \in \boldsymbol{\mathcal{J}}_x \right\} := \Sigma^{(1)} \otimes \cdots \otimes \Sigma^{(n)}$$

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

$$(5.6) \qquad \boldsymbol{\Sigma}^V := \left\{ \boldsymbol{\sigma_\lambda} / \|\boldsymbol{\sigma_\lambda}\|_V : \boldsymbol{\lambda} \in \boldsymbol{\mathcal{J}}_x \right\}, \quad \boldsymbol{\Sigma}^{V'} := \left\{ \boldsymbol{\sigma_\lambda} / \|\boldsymbol{\sigma_\lambda}\|_{V'} : \boldsymbol{\lambda} \in \boldsymbol{\mathcal{J}}_x \right\}$$

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

5.4. Space-time discretization.
We are now in the position to define the Riesz wavelet bases $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}}$ and $\check{\boldsymbol{\Psi}}^{\mathcal{Y}}$ from (3.2). With $L_2(0, T; L_2(\boldsymbol{\Omega})) \approx L_2(0, T) \otimes L_2(\boldsymbol{\Omega})$,

$$(5.7) \qquad \widehat{\boldsymbol{\Psi}} := \left\{ \widehat{\psi}_{\boldsymbol{\lambda}} := \theta_{\lambda_t}^{\mathrm{per}} \otimes \boldsymbol{\sigma_{\lambda_x}} : \boldsymbol{\lambda} := (\lambda_t, \boldsymbol{\lambda}_x) \in \widehat{\boldsymbol{\mathcal{J}}} := \mathcal{J}_t^{\mathrm{per}} \times \boldsymbol{\mathcal{J}}_x \right\} = \Theta^{\mathrm{per}} \otimes \boldsymbol{\Sigma},$$

$$(5.8) \qquad \check{\boldsymbol{\Psi}} := \left\{ \check{\psi}_{\boldsymbol{\lambda}} := \vartheta_{\lambda_t} \otimes \boldsymbol{\sigma_{\lambda_x}} : \boldsymbol{\lambda} := (\lambda_t, \boldsymbol{\lambda}_x) \in \check{\boldsymbol{\mathcal{J}}} := \mathcal{J}_t \times \boldsymbol{\mathcal{J}}_x \right\} = \Theta \otimes \boldsymbol{\Sigma},$$

are both Riesz bases for $L_2(0, T; L_2(\mathbf{\Omega}))$. At this point, we only need to normalize the above Riesz bases appropriately (see [GO95, Propositions 1 & 2]) so that

$$(5.9) \qquad \widehat{\mathbf{\Psi}}^{\mathcal{X}} := \{\widehat{\psi}_{\boldsymbol{\lambda}}/\|\widehat{\psi}_{\boldsymbol{\lambda}}\|_{\mathcal{X}} : \boldsymbol{\lambda} \in \widehat{\boldsymbol{\mathcal{J}}}\} = \mathbf{D}^{\mathcal{X}}\widehat{\mathbf{\Psi}}, \quad \mathbf{D}^{\mathcal{X}} := \mathrm{diag}\left[(\|\widehat{\psi}_{\boldsymbol{\lambda}}\|_{\mathcal{X}}^{-1})_{\boldsymbol{\lambda} \in \widehat{\boldsymbol{\mathcal{J}}}}\right],$$

$$(5.10) \quad \widecheck{\mathbf{\Psi}}^{\mathcal{Y}} := \{\widecheck{\psi}_{\boldsymbol{\lambda}}/\|\widecheck{\psi}_{\boldsymbol{\lambda}}\|_{\mathcal{Y}} : \boldsymbol{\lambda} \in \widecheck{\boldsymbol{\mathcal{J}}}\} = \mathbf{D}^{\mathcal{Y}}\widecheck{\mathbf{\Psi}}, \quad \mathbf{D}^{\mathcal{Y}} := \mathrm{diag}\left[(\|\widecheck{\psi}_{\boldsymbol{\lambda}}\|_{\mathcal{Y}}^{-1})_{\boldsymbol{\lambda} \in \widecheck{\boldsymbol{\mathcal{J}}}}\right],$$

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

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

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

where $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\mathcal{J}} := \mathcal{J}^{(0)} \times \mathcal{J}^{(1)} \times \cdots \times \mathcal{J}^{(n)}$. In this setting, it is clear that $\Psi^{(0)} \in \{\Theta^{\mathrm{per}}, \Theta\}$, $\mathcal{J}^{(0)} \in \{\mathcal{J}_t^{\mathrm{per}}, \mathcal{J}_t\}$ 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 $\mathrm{c}_{\mathcal{X}}(\widehat{\mathbf{\Psi}})$, $\mathrm{C}_{\mathcal{X}}(\widehat{\mathbf{\Psi}})$, $\mathrm{c}_{\mathcal{Y}}(\widecheck{\mathbf{\Psi}})$, $\mathrm{C}_{\mathcal{Y}}(\widecheck{\mathbf{\Psi}})$ in (3.4). Again, we use that $\mathcal{X}$ and $\mathcal{Y}$ are (intersections of) tensor products of Hilbert spaces. As in [SS09, §6], we have the following estimates for $\widehat{\mathbf{\Psi}}^{\mathcal{X}}$ and $\widecheck{\mathbf{\Psi}}^{\mathcal{Y}}$

$$(5.11) \qquad \mathrm{c}_{\mathcal{X}}(\widehat{\mathbf{\Psi}}) \geq \min\left\{\mathrm{c}_{L_2}(\Theta^{\mathrm{per}}) \cdot \mathrm{c}_V(\mathbf{\Sigma}), \mathrm{c}_{H_{\mathrm{per}}^1}(\Theta^{\mathrm{per}}) \cdot \mathrm{c}_{V'}(\mathbf{\Sigma})\right\},$$

$$(5.12) \qquad \mathrm{C}_{\mathcal{X}}(\widehat{\mathbf{\Psi}}) \leq \min\left\{\mathrm{C}_{L_2}(\Theta^{\mathrm{per}}) \cdot \mathrm{C}_V(\mathbf{\Sigma}), \mathrm{C}_{H_{\mathrm{per}}^1}(\Theta^{\mathrm{per}}) \cdot \mathrm{C}_{V'}(\mathbf{\Sigma})\right\},$$

$$(5.13) \qquad \mathrm{c}_{\mathcal{Y}}(\widecheck{\mathbf{\Psi}}) \geq \mathrm{c}_{L_2}(\Theta) \cdot \mathrm{c}_V(\mathbf{\Sigma}), \quad \mathrm{C}_{\mathcal{Y}}(\widecheck{\mathbf{\Psi}}) \leq \mathrm{C}_{L_2}(\Theta) \cdot \mathrm{C}_V(\mathbf{\Sigma}).$$

The Riesz constants $\mathrm{c}_V(\mathbf{\Sigma})$, $\mathrm{C}_V(\mathbf{\Sigma})$ 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

$$(5.14) \qquad \mathrm{c}_V(\mathbf{\Sigma}) \geq \min_{m \in \{1, \ldots, n\}} \min\left\{\mathrm{c}_{L_2}(\Sigma^{(m)}), \mathrm{c}_{V^{(m)}}(\Sigma^{(m)})\right\} \prod_{k \neq m} \mathrm{c}_{L_2}(\Sigma^{(k)}),$$

$$(5.15) \qquad \mathrm{C}_V(\mathbf{\Sigma}) \leq \max_{m \in \{1, \ldots, n\}} \max\left\{\mathrm{C}_{L_2}(\Sigma^{(m)}), \mathrm{C}_{V^{(m)}}(\Sigma^{(m)})\right\} \prod_{k \neq m} \mathrm{C}_{L_2}(\Sigma^{(k)}).$$

Unfortunately, the same approach does not apply to the (dual) Riesz constants $\mathrm{c}_{V'}(\mathbf{\Sigma})$, $\mathrm{C}_{V'}(\mathbf{\Sigma})$ of $\mathbf{\Sigma}^{V'}$ in (5.6). However, one may consider $\widetilde{\mathbf{\Sigma}}^V$ being the unique Riesz basis for V that is *dual* to $\mathbf{\Sigma}^{V'}$, i.e., $\langle\widetilde{\mathbf{\Sigma}}^V, \mathbf{\Sigma}^{V'}\rangle_{V \times V'} = \mathrm{Id}$. Denoting by $\mathrm{c}_V(\widetilde{\mathbf{\Sigma}})$, $\mathrm{C}_V(\widetilde{\mathbf{\Sigma}})$ the associated Riesz constants, it can be shown that $\mathrm{C}_V(\widetilde{\mathbf{\Sigma}})^{-1} \leq \mathrm{c}_{V'}(\mathbf{\Sigma})$ and $\mathrm{C}_{V'}(\mathbf{\Sigma}) \leq \mathrm{c}_V(\widetilde{\mathbf{\Sigma}})^{-1}$. Observe that for computing bounds for $\mathrm{c}_V(\widetilde{\mathbf{\Sigma}})$, $\mathrm{C}_V(\widetilde{\mathbf{\Sigma}})$, we may proceed as for bounding $\mathrm{c}_V(\mathbf{\Sigma})$, $\mathrm{C}_V(\mathbf{\Sigma})$. 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, §2]).

Remark 5.2. Recalling the construction of wavelets, note that the *numerical approximation* of $\mathrm{c}_V(\widetilde{\mathbf{\Sigma}})$, $\mathrm{C}_V(\widetilde{\mathbf{\Sigma}})$ may be difficult since the the dual basis $\widetilde{\mathbf{\Sigma}}^V$ (and their derivatives) may not be available in a closed form. If sharp bounds are needed, one may use an $L_2(\mathbf{\Omega})$-orthonormal basis $\mathbf{\Sigma}$ so that $\widetilde{\mathbf{\Sigma}}^V = \mathbf{\Sigma}^V$, e.g. multiwavelets.

5.6. **Best approximation rates.** We need to know for which values of s the solution $\mathbf{u}$ of (3.3) is in $\mathcal{A}^s$. More precisely, for a fixed trial basis $\widehat{\mathbf{\Psi}}^{\mathcal{X}}$, the question is what is the *largest* value $s_{\max}$ of s for which $\mathbf{u} \in \mathcal{A}^s$ can be expected and that cannot be increased by imposing higher smoothness conditions on u (excluding special cases where $\mathbf{u}$ is (close to) a finite vector). This value $s_{\max}$ is referred to

as *best possible approximation rate*. For our setting, we may apply the results from [SS09, §7.2]. With $u = \mathbf{u}^\top \widehat{\mathbf{\Psi}}^{\mathcal{X}} \in \mathcal{X} \cap H^{d_t}(0, T) \otimes \mathscr{H}^{d_x}(\mathbf{\Omega})$ and the Sobolev space

$$\mathscr{H}^{d_x}(\mathbf{\Omega}) := \bigcap_{i=1}^{n} \bigotimes_{j=1}^{n} Z_{ij}, \text{ where } Z_{ij} := \left\{ \begin{array}{ll} L_2(\Omega_i), & i \neq j, \\ H^{d_x}(\Omega_i), & i = j, \end{array} \right.$$

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

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

We recall that d_t denotes the polynomial order of Θ^{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 $u \in \mathscr{H}^{d_x}(\mathbf{\Omega})$ is sufficient but not necessary for obtaining the above rate. In fact, the Sobolev space $H^{d_t}(0, T) \otimes \mathscr{H}^{d_x}(\mathbf{\Omega})$ 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 ℓ_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:

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

Analogously to (3.3), the equivalent ℓ_2-problem to this problem reads:

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

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

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

The idea of an AWGM for (6.2) is outlined in an (idealized) Algorithm 1, [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 $\widehat{\mathbf{\Lambda}}_1 \subset \widehat{\mathcal{J}}$, a sequence of *nested* finite index sets $(\widehat{\mathbf{\Lambda}}_k)_k$ is computed. On each such $\widehat{\mathbf{\Lambda}}_k$, a Galerkin problem is solved that yields the (finite) vector $\mathbf{u}_{\widehat{\mathbf{\Lambda}}_k}$. Due to the Riesz basis property, it holds that (see also (3.1))

$$c_{\mathcal{X}}(\widehat{\mathbf{\Psi}})^{\frac{1}{2}} \|\mathbf{u} - \mathbf{u}_{\widehat{\mathbf{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \|u - \mathbf{u}_{\widehat{\mathbf{\Lambda}}_k}^\top \widehat{\mathbf{\Psi}}^{\mathcal{X}}\|_{\mathcal{X}} \leq C_{\mathcal{X}}(\widehat{\mathbf{\Psi}})^{\frac{1}{2}} \|\mathbf{u} - \mathbf{u}_{\widehat{\mathbf{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})}.$$

Given $\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$, the computation of the next $\widehat{\boldsymbol{\Lambda}}_{k+1}$ is based on the *infinitely* supported *residual* $\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k} \in \ell_2(\widehat{\mathcal{J}})$ and the *error estimator* $\|\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})}$ which satisfies:

$$(6.4) \qquad \|\mathbf{C}\|^{-1}\|\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \le \|\mathbf{u} - \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \le \|\mathbf{C}^{-1}\|\|\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})}.$$

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 $\boldsymbol{\Lambda}_k$. This so-called *bulk-chasing* process is steered by the parameter δ.

Algorithm 1 $[\mathbf{u}_\varepsilon] = \mathbf{IDEALIZED\text{–}AWGM}[\varepsilon, \boldsymbol{\Lambda}_1]$

Input: Target tolerance ε and an index set $\widehat{\boldsymbol{\Lambda}}_1 \ne \emptyset$.
Parameter: $\delta \in (0, \kappa(\mathbf{C})^{-\frac{1}{2}})$.
 1: **for** $k = 1, 2, \dots$ **do**
 2:     *Solve the Galerkin problem:*

$$(6.5) \qquad \text{Find } \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k} \in \ell_2(\widehat{\boldsymbol{\Lambda}}_k): \qquad {}_{\widehat{\boldsymbol{\Lambda}}_k}\mathbf{C}_{\widehat{\boldsymbol{\Lambda}}_k}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k} = \mathbf{g}_{\widehat{\boldsymbol{\Lambda}}_k}, \quad \mathbf{g}_{\widehat{\boldsymbol{\Lambda}}_k} := \mathbf{R}_{\widehat{\boldsymbol{\Lambda}}_k}\mathbf{g} \in \ell_2(\widehat{\boldsymbol{\Lambda}}_k).$$

 3:     *Residual computation:* Compute $\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$ and $\nu_k := \|\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2}$.
 4:     **if** $\nu_k \le \varepsilon/\|\mathbf{C}^{-1}\|$ **then return** $\mathbf{u}_\varepsilon := \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$.
 5:     *Bulk chasing criterion:* Find smallest index set $\widehat{\boldsymbol{\Lambda}}_{k+1} \supset \widehat{\boldsymbol{\Lambda}}_k$ such that

$$(6.6) \qquad \|\mathbf{R}_{\widehat{\boldsymbol{\Lambda}}_{k+1}}(\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k})\|_{\ell_2(\widehat{\boldsymbol{\Lambda}}_{k+1})} \ge \delta\|\mathbf{g} - \mathbf{C}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})}.$$

 6: **end for**

Proposition 6.1 ([Ste09, Proposition 4.1]). *The iterates $\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$ produced by Algorithm 1 satisfy $\|\mathbf{u} - \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\| \le [1 - \delta^2\kappa(\mathbf{C})^{-1}]^{k/2}\|\mathbf{u}\|$. For the output $\mathbf{u}_\varepsilon$ it holds $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(\widehat{\mathcal{J}})} \le \varepsilon$. If $\mathbf{u} \in \mathcal{A}^s$ for some $s > 0$, it also holds for $\mathcal{N}_k := \#\widehat{\boldsymbol{\Lambda}}_k$ that*

$$(6.7) \qquad \|\mathbf{u} - \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s}\mathcal{N}_k^{-s}, \quad \#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s}^{1/s}.$$

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], also 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 $\mathbf{Bu} = \mathbf{f}$ in (3.3). However, the generalization of the idealized scheme to (3.3) is not trivial: *(1) Symmetry and positive definiteness.* Recall that $\mathbf{B}$ from (3.3) is *not* s.p.d., so that $\langle \cdot, \mathbf{B} \cdot \rangle_{\ell_2(\widehat{\mathcal{J}}) \times \ell_2(\check{\mathcal{J}})}$ is *not* an equivalent norm on $\ell_2(\widehat{\mathcal{J}})$. 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 $\widehat{\boldsymbol{\Lambda}}_{k+1}$ from $\widehat{\boldsymbol{\Lambda}}_k$. In analogy to (6.4), the residual $\mathbf{f} - \mathbf{Bu}_{\widehat{\boldsymbol{\Lambda}}_k} \in \ell_2(\check{\mathcal{J}})$ with error estimator $\|\mathbf{f} - \mathbf{Bu}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\check{\mathcal{J}})}$ satisfies

$$(6.8) \qquad \|\mathbf{B}\|^{-1}\|\mathbf{f} - \mathbf{Bu}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\check{\mathcal{J}})} \le \|\mathbf{u} - \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \le \|\mathbf{B}^{-1}\|\|\mathbf{f} - \mathbf{Bu}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\check{\mathcal{J}})}.$$

But the residual is an element of $\ell_2(\check{\mathcal{J}})$, $\check{\mathcal{J}} \neq \widehat{\mathcal{J}}$. Thus, we *cannot* compute $\widehat{\boldsymbol{\Lambda}}_k$ by selecting some contributions from the residual as in (6.6). *(3) Petrov-Galerkin problems.* Since $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}} \neq \check{\boldsymbol{\Psi}}^{\mathcal{Y}}$, the (well-posed) Galerkin problem in line 5 of Algorithm 1 here becomes a *Petrov*-Galerkin problem. Hence 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 $\widehat{\boldsymbol{\Lambda}}_k$.

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

$$(6.9) \qquad \text{Find } \mathbf{u} \in \ell_2(\widehat{\mathcal{J}}): \qquad \mathbf{B}^\top \mathbf{B} \mathbf{u} = \mathbf{B}^\top \mathbf{f}, \qquad \mathbf{B}^\top \mathbf{f} \in \ell_2(\widehat{\mathcal{J}}).$$

Since $\mathbf{B}$ 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 $\mathbf{u} \in \ell_2(\widehat{\mathcal{J}})$ such that $\mathbf{u} = \mathrm{argmin}_{\mathbf{v} \in \ell_2(\widehat{\mathcal{J}})} \|\mathbf{B}\mathbf{v} - \mathbf{f}\|^2_{\ell_2(\check{\mathcal{J}})}$ for given $\mathbf{f} \in \ell_2(\check{\mathcal{J}})$. We anticipate that one does *not* expect the usually dramatic effect of a squared condition number for $\mathbf{B}^\top \mathbf{B}$ since $\mathbf{B}$ 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, $\mathbf{B}^\top \mathbf{B}$ is symmetric. Moreover, by (3.4), it is also positive definite and it holds that

$$(6.10) \qquad \qquad \|\mathbf{B}^\top \mathbf{B}\| \leq \|\mathbf{B}\|^2, \quad \|(\mathbf{B}^\top \mathbf{B})^{-1}\| \leq \|\mathbf{B}^{-1}\|^2,$$

hence $\kappa(\mathbf{B}^\top \mathbf{B}) \leq \|\mathbf{B}\|^2 \|\mathbf{B}^{-1}\|^2$. Thus, we consider $\mathbf{C}\mathbf{u} = \mathbf{g}$ with $\mathbf{C} = \mathbf{B}^\top \mathbf{B}$, $\mathbf{g} = \mathbf{B}^\top \mathbf{f}$ and $\|\cdot\|^2 := \langle \cdot, \mathbf{B}^\top \mathbf{B} \cdot \rangle$ and use Algorithm 1.
(2) Bulk chasing and residual computation. Instead of considering the residual in $\ell_2(\check{\mathcal{J}})$, we now obtain $\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}) \in \ell_2(\widehat{\mathcal{J}})$ with error estimator $\rho_k := \|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k})\|_{\ell_2(\widehat{\mathcal{J}})}$. In analogy to (6.4) and (6.8), we infer that

$$(6.11) \qquad \qquad \|\mathbf{B}\|^{-2}\rho_k \leq \|\mathbf{u} - \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \|\mathbf{B}^{-1}\|^2 \rho_k.$$

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

$$(6.12) \qquad \text{Find } \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}} \in \ell_2(\widehat{\boldsymbol{\Lambda}}): \qquad {}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}} \mathbf{u}_{\widehat{\boldsymbol{\Lambda}}} = {}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{f}, \qquad {}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{f} \in \ell_2(\widehat{\boldsymbol{\Lambda}}).$$

Observe that the unique solution $\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}} = \mathrm{argmin}_{\mathbf{v}_{\widehat{\boldsymbol{\Lambda}}} \in \ell_2(\widehat{\boldsymbol{\Lambda}})} \|\mathbf{B}_{\widehat{\boldsymbol{\Lambda}}}\mathbf{v}_{\widehat{\boldsymbol{\Lambda}}} - \mathbf{f}\|^2_{\ell_2(\check{\mathcal{J}})}$ 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 $\mathbf{B}^\top \mathbf{B}$ is s.p.d., we infer from (6.10) that $\|_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}}\| \leq \|\mathbf{B}\|^2$ as well as $\|(_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}})^{-1}\| \leq \|\mathbf{B}^{-1}\|^2$ for all $\widehat{\boldsymbol{\Lambda}} \subseteq \widehat{\mathcal{J}}$. In particular, the condition number $\kappa(_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}^\top \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}})$ is bounded *independently* of $\widehat{\boldsymbol{\Lambda}}$.

Remark 6.3. Obviously, neither the residual in (6.11) nor the solution $\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$ 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 $\mathbf{w}_{\widehat{\boldsymbol{\Lambda}}_k}$ to $\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k}$ and an approximate residual $\check{\mathbf{r}}_k$ to $\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{u}_{\widehat{\boldsymbol{\Lambda}}_k})$. 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 $\mathbf{u} \in \mathcal{A}^s$ and denote by $\mathbf{w}_{\widehat{\Lambda}}$ an approximate solution to (6.12).

(**RES**) *Approximate residual:* For a given relative tolerance $0 < \omega_{\mathrm{ls}} < 1$, the output $\widehat{\mathbf{r}}$ of **RESIDUAL**$[\mathbf{w}_{\widehat{\Lambda}}, \omega_{\mathrm{ls}}]$ should satisfy

$$(7.1) \qquad \|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\Lambda}}) - \widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \omega_{\mathrm{ls}} \cdot \nu, \quad \nu := \|\widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})},$$

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

(**GAL**) *Approximate Galerkin problem:* For a given relative tolerance $0 < \gamma_{\mathrm{ls}} < 1$, the output $\mathbf{w}_{\widehat{\Lambda}}$ of **GALSOLVE**$[\widehat{\Lambda}, \overline{\mathbf{w}}_{\widehat{\Lambda}}, \gamma_{\mathrm{ls}} \cdot \nu]$ should satisfy

$$(7.2) \qquad \|_{\widehat{\Lambda}}\mathbf{B}^\top(\mathbf{f} - \mathbf{B}_{\widehat{\Lambda}}\mathbf{w}_{\widehat{\Lambda}})\|_{\ell_2(\widehat{\Lambda})} \leq \gamma_{\mathrm{ls}} \cdot \nu,$$

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

(**EXP**) *Approximate bulk chasing:* For a given parameter $0 < \delta < 1$, the output $\widehat{\underline{\Lambda}} \subset \widehat{\mathcal{J}}$ of **EXPAND**$[\widehat{\Lambda}, \widehat{\mathbf{r}}, \delta]$ should satisfy

$$(7.3) \qquad \widehat{\underline{\Lambda}} \supset \widehat{\Lambda}, \qquad \|\mathbf{R}_{\widehat{\underline{\Lambda}}}\widehat{\mathbf{r}}\|_{\ell_2(\widehat{\underline{\Lambda}})} \geq \delta\|\widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})},$$

and, up to some absolute multiple, $\widehat{\underline{\Lambda}}$ is minimal among all sets that satisfy (7.3). The computational cost of this routine is of order $\mathcal{O}(\#\widehat{\Lambda} + \# \operatorname{supp} \widehat{\mathbf{r}})$.

Algorithm 2 $[\mathbf{u}_\varepsilon] = \mathbf{LS\text{–}AWGM}[\varepsilon, \widehat{\Lambda}_1, \nu_0]$

Input: Target tolerance ε, finite index set $\widehat{\Lambda}_1 \subset \widehat{\mathcal{J}}$ and tolerance $\nu_0 \approx \|\mathbf{B}^\top\mathbf{f}\|_{\ell_2(\widehat{\mathcal{J}})}$.
Parameters: $\delta, \omega_{\mathrm{ls}}, \gamma_{\mathrm{ls}}$ with $\omega_{\mathrm{ls}} \in (0, \delta)$, $\frac{\delta + \omega_{\mathrm{ls}}}{1 - \omega_{\mathrm{ls}}} < \kappa(\mathbf{B}^\top\mathbf{B})^{-\frac{1}{2}}$,
    $\gamma_{\mathrm{ls}} \in (0, \frac{(1-\omega_{\mathrm{ls}})(\delta-\omega_{\mathrm{ls}})}{1+\omega_{\mathrm{ls}}}\kappa(\mathbf{B}^\top\mathbf{B})^{-1})$.
  1: Set $\mathbf{w}_{\widehat{\Lambda}_0} := 0$.
  2: **for** $k = 1, 2, \ldots$ **do**
  3:     $\mathbf{w}_{\widehat{\Lambda}_k} := \mathbf{GALSOLVE}[\widehat{\Lambda}_k, \mathbf{w}_{\widehat{\Lambda}_{k-1}}, \gamma_{\mathrm{ls}} \cdot \nu_{k-1}]$.
  4:     $\widehat{\mathbf{r}}_k := \mathbf{RESIDUAL}[\mathbf{w}_{\widehat{\Lambda}_k}, \omega_{\mathrm{ls}}]$ and set $\nu_k := \|\widehat{\mathbf{r}}_k\|_{\ell_2}$.
  5:     **if** $\nu_k \leq \varepsilon/\|\mathbf{B}^{-1}\|^2$ **then return** $\mathbf{u}_\varepsilon := \mathbf{w}_{\widehat{\Lambda}_k}$.
  6:     $\widehat{\Lambda}_{k+1} := \mathbf{EXPAND}[\widehat{\Lambda}_k, \widehat{\mathbf{r}}_k, \delta]$
  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 & Theorem 4.1]:

Theorem 7.1 ([GHS07, Ste09])**.** *Let the assumptions on* (**RES**)*,* (**GAL**) *and* (**EXP**) *and the requirements on* $\delta, \omega_{\mathrm{ls}}, \gamma_{\mathrm{ls}}$ *from Algorithm 2 hold. Then, the iterates* $\mathbf{w}_{\widehat{\Lambda}_k}$ *produced by* **LS–AWGM** *satisfy* $\|\mathbf{u} - \mathbf{w}_{\widehat{\Lambda}_k}\| \leq \rho^{k/2}\|\mathbf{u}\|$ *where* $\rho := 1 -$

$(\frac{\delta - \omega_{ls}}{1 + \omega_{ls}})\kappa(\mathbf{B}^\top \mathbf{B})^{-1} + \frac{\gamma_{ls}^2}{(1 - \omega_{ls})^2}\kappa(\mathbf{B}^\top \mathbf{B}) < 1$ *and the output* $\mathbf{u}_\varepsilon$ *satisifes* $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(\hat{\mathcal{J}})} \leq \varepsilon$. *If, moreover,* $\mathbf{u} \in \mathcal{A}^s$ *for some* $s > 0$, *it holds for* $\mathcal{N}_k := \#\hat{\mathbf{\Lambda}}_k$ *that*

$$(7.4) \qquad \|\mathbf{u} - \mathbf{w}_{\hat{\mathbf{\Lambda}}_k}\|_{\ell_2(\hat{\mathcal{J}})} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s} \mathcal{N}_k^{-s}, \quad \#\mathrm{supp}\, \mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s} \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s}.$$

Note that a realization of **EXPAND** can easily be obtained by an *approximate* sorting of $\hat{\mathbf{r}}$ 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 $\mathbf{B}$ and $\mathbf{B}^\top$) 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 $\mathrm{supp}\, \psi_\lambda \subset \bigcup_{\mu \in \Lambda; |\mu| = \lambda - 1} \mathrm{supp}\, \psi_\mu$.

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

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

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

Definition 8.2 ([KS12]). An index set $\mathbf{\Lambda} \in \boldsymbol{\mathcal{J}}$ 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

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

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

Loosely speaking, a multitree $\mathbf{\Lambda} \in \boldsymbol{\mathcal{J}}$ is "when frozen in any n coordinate directions, a tree in the remaining coordinate" (see [KS13, §3.1]).

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

$$(8.3) \qquad \mathbf{B} = \mathbf{D}^{\mathcal{Y}} \Big[\sum_{m=1}^{M} \prod_{i=0}^{n} b_m^{(i)}(\widehat{\Psi}^{(i)}, \widecheck{\Psi}^{(i)}) \Big] \mathbf{D}^{\mathcal{X}} = \mathbf{D}^{\mathcal{Y}} \Big[\sum_{m=1}^{M} \bigotimes_{i=0}^{n} \vec{B}_m^{(i)} \Big] \mathbf{D}^{\mathcal{X}},$$

where $\vec{B}_m^{(i)} := b_m^{(i)}(\widehat{\Psi}^{(i)}, \widecheck{\Psi}^{(i)})$ 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 $b_m^{(i)}$ are *local* in the sense that $b_m^{(i)}(w, v) = 0$ whenever $|\operatorname{supp} v \cap \operatorname{supp} w| = 0$. The special structure of $\mathbf{B}$ can be used to efficiently realize the application of $_{\widecheck{\Lambda}}\mathbf{B}_{\widehat{\Lambda}}$ to a vector $\mathbf{v}_{\widehat{\Lambda}} \in \ell_2(\widehat{\Lambda})$ for finite multitrees $\widehat{\Lambda} \in \widehat{\mathcal{J}}$ and $\widecheck{\Lambda} \in \widecheck{\mathcal{J}}$. As described in [KS12], this can be realized in linear complexity, i.e., $\mathcal{O}(\#\widehat{\Lambda} + \#\widecheck{\Lambda})$ by using a *separation* of $_{\widecheck{\Lambda}}\mathbf{B}_{\widehat{\Lambda}}$ 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 (possible bi-infinite) matrices $\vec{A}^{(1)}, \vec{A}^{(2)}$ and identity matrices $\vec{\mathrm{Id}}^{(1)}, \vec{\mathrm{Id}}^{(2)}$ of appropriate dimension:

$$(8.4) \quad \vec{A}^{(1)} \otimes \vec{A}^{(2)} = \big[\vec{A}^{(1)} \otimes \vec{\mathrm{Id}}^{(2)}\big] \circ \big[\mathrm{Id}^{(1)} \otimes \vec{A}^{(2)}\big] = \big[\vec{\mathrm{Id}}^{(1)} \otimes \vec{A}^{(2)}\big] \circ \big[\vec{A}^{(1)} \otimes \vec{\mathrm{Id}}^{(2)}\big].$$

Then we split $\vec{B}_m^{(i)} = \vec{L}_m^{(i)} + \vec{U}_m^{(i)}$ into a (stricly) lower $\vec{L}_m^{(i)} := \big[(\vec{B}_m^{(i)})_{\lambda,\mu}\big]_{|\lambda|>|\mu|}$ and an upper triangular matrix $\vec{U}_m^{(i)} := \big[(\vec{B}_m^{(i)})_{\lambda,\mu}\big]_{|\lambda|\leq|\mu|}$. With (8.4), it can then be shown that there exist multitrees $\overline{\Xi}$ and $\underline{\Xi}$ such that we have the following *equivalent* representation of $_{\widecheck{\Lambda}}\mathbf{B}_{\widehat{\Lambda}}$,

$$\mathbf{D}^{\mathcal{Y}} \Bigg[\sum_{m=1}^{M} \underbrace{\mathbf{R}_{\widecheck{\Lambda}} \big[\vec{\mathrm{Id}}^{(0)} \otimes \vec{B}_m^{(1)} \otimes \cdots \otimes \vec{B}_m^{(n)}\big] \mathbf{E}_{\overline{\Pi}}}_{=:(\mathrm{I})} \circ \underbrace{\mathbf{R}_{\overline{\Pi}} \big[\vec{U}_m^{(0)} \otimes \vec{\mathrm{Id}}^{(1)} \otimes \cdots \otimes \vec{\mathrm{Id}}^{(n)}\big] \mathbf{E}_{\widehat{\Lambda}}}_{=:(\mathrm{II})}$$

$$+ \sum_{m=1}^{M} \underbrace{\mathbf{R}_{\widecheck{\Lambda}} \big[\vec{L}_m^{(0)} \otimes \vec{\mathrm{Id}}^{(1)} \otimes \cdots \otimes \vec{\mathrm{Id}}^{(n)}\big] \mathbf{E}_{\underline{\Pi}}}_{=:(\mathrm{III})} \circ \underbrace{\mathbf{R}_{\underline{\Pi}} \big[\vec{\mathrm{Id}}^{(0)} \otimes \vec{B}_m^{(1)} \otimes \cdots \otimes \vec{B}_m^{(n)}\big] \mathbf{E}_{\widehat{\Lambda}}}_{=:(\mathrm{IV})} \Bigg] \mathbf{D}^{\mathcal{X}}.$$

It holds $\#\overline{\Xi} + \#\underline{\Xi} \lesssim \#\widecheck{\Lambda} + \#\widehat{\Lambda}$. 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 $\vec{L}_m^{(0)}|_{\widecheck{\Lambda}^{(0)} \times \widehat{\Lambda}^{(0)}}$, $\vec{U}_m^{(0)}|_{\widecheck{\Lambda}^{(0)} \times \widehat{\Lambda}^{(0)}}$ and $\vec{B}_m^{(1)}|_{\widecheck{\Lambda}^{(1)} \times \widehat{\Lambda}^{(1)}}$ ($n = 1$) is required. Due to the tree structure, these tasks can be realized in linear complexity despite the fact that neither of the matrices $\vec{L}_m^{(0)}$, $\vec{U}_m^{(0)}$ or $\vec{B}_m^{(1)}$ is sparse in general (see [KS12, §2]). For $n > 2$, the remaining parts (I) and (IV) can be treated recursively by applying the same procedure to $\vec{B}_m^{(1)} \otimes \cdots \otimes \vec{B}_m^{(n)}$.

Theorem 8.4 ([KS12, Theorem 3.1])**.** *Let $\mathcal{A}$ be a linear differential operator with polynomial coefficients and let $\widehat{\Lambda} \subset \widehat{\mathcal{J}}$, $\widecheck{\Lambda} \in \widecheck{\mathcal{J}}$ be multitrees. Then, for any $\mathbf{v}_{\widehat{\Lambda}} \in \ell_2(\widehat{\Lambda})$, the product $_{\widecheck{\Lambda}}\mathbf{B}_{\widehat{\Lambda}}\mathbf{v}_{\widehat{\Lambda}}$ can be computed in $\mathcal{O}(\#\widehat{\Lambda} + \#\widecheck{\Lambda})$ operations.*

Remark 8.5. If $\mathcal{A}$ 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 $\widehat{\Lambda}^{(i)}$ and $\widecheck{\Lambda}^{(i)}$ are trees (cf. [KS12, §2]).

8.3. RESIDUAL: Multitree residual approximation.
We need to approximate the residual $\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\mathbf{\Lambda}}})$ by a residual of type $_{\widehat{\Xi}}\mathbf{B}_{\widehat{\Xi}}^\top\big(\mathbf{f}_{\widecheck{\Xi}} - {}_{\widecheck{\Xi}}\mathbf{B}_{\widehat{\mathbf{\Lambda}}}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}\big)$.

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

Theorem 8.6 ([KS13]). *Let $0 < \omega < 1$, let $\mathcal{A}$ be a differential operator with polynomial coefficients and let $\mathbf{u} \in \mathcal{A}^s_{\mathrm{mtree}}$ for some $s > 0$. Then, for all finite multitrees $\widehat{\mathbf{\Lambda}} \subset \widehat{\mathcal{J}}$ and all $\mathbf{w}_{\widehat{\mathbf{\Lambda}}} \in \ell_2(\widehat{\mathbf{\Lambda}})$, there exists a multitree $\widecheck{\Xi} = \widecheck{\Xi}(\widehat{\mathbf{\Lambda}}, \omega) \subset \widecheck{\mathcal{J}}$ such that $\#\widecheck{\Xi} \lesssim \#\widehat{\mathbf{\Lambda}} + \nu^{-1/s}$ with $\nu := \|\widecheck{\mathbf{r}}\|_{\ell_2(\widecheck{\mathcal{J}})}$, $\mathbf{f}_{\widecheck{\Xi}} := \mathbf{R}_{\widecheck{\Xi}}\mathbf{f}$ and*

$$(8.5) \qquad \|(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}) - \widecheck{\mathbf{r}}\|_{\ell_2(\widecheck{\mathcal{J}})} \le \omega\|\widecheck{\mathbf{r}}\|_{\ell_2(\widecheck{\mathcal{J}})}, \quad \widecheck{\mathbf{r}} := \mathbf{f}_{\widecheck{\Xi}} - {}_{\widecheck{\Xi}}\mathbf{B}_{\widehat{\mathbf{\Lambda}}}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}.$$

Remark 8.7. Due to the multitree structure of $\widehat{\mathbf{\Lambda}}$ and $\widecheck{\Xi}$, the computational cost for *computing* $\widecheck{\mathbf{r}}$ is $\mathcal{O}(\#\widehat{\mathbf{\Lambda}} + \nu^{-1/s})$ if an entry $\mathbf{f}_{\boldsymbol{\lambda}}$ of $\mathbf{f} = (\mathbf{f}_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda} \in \widecheck{\mathcal{J}}}$ 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 $\mathbf{f}$ by some $\mathbf{f}_\varepsilon$ with $\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(\widecheck{\mathcal{J}})} \le \varepsilon$ and $\#\operatorname{supp}\mathbf{f}_\varepsilon \lesssim \varepsilon^{-1/s}$ 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 $\mathbf{B}$ and $\mathbf{B}^\top$. If $\mathcal{A}$ is a linear differential operator with polynomial coefficients, it can be shown that for any $0 < \eta < 1$, there exists $\mathbf{B}_\eta : \ell_2(\widehat{\mathcal{J}}) \to \ell_2(\widecheck{\mathcal{J}})$ such that

$$(8.6) \qquad \|\mathbf{B} - \mathbf{B}_\eta\| \le \eta, \quad \|\mathbf{B}^\top - \mathbf{B}_\eta^\top\| \le \eta,$$

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

$$(8.7) \qquad \|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}) - \mathbf{B}_\eta^\top\widecheck{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})} \le \omega_{\mathrm{ls}}\|\mathbf{B}_\eta^\top\widecheck{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})},$$

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

$$(8.8) \qquad \widehat{\mathbf{r}} := {}_{\widehat{\Xi}}\mathbf{B}_{\widehat{\Xi}}^\top\big(\mathbf{f}_{\widecheck{\Xi}} - {}_{\widecheck{\Xi}}\mathbf{B}_{\widehat{\mathbf{\Lambda}}}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}\big) = {}_{\widehat{\Xi}}\mathbf{B}_{\widehat{\Xi}}^\top\widecheck{\mathbf{r}}$$

such that $\|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\mathbf{\Lambda}}}) - \widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})} \le \omega_{\mathrm{ls}}\|\widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})}$ where $\widehat{\Xi}$ is the smallest multitree containing $\operatorname{supp}\mathbf{B}_\eta^\top\widecheck{\mathbf{r}}$. The residual computation requires $\mathcal{O}(\#\widehat{\Xi} + \#\widecheck{\Xi})$ operations.

Remark 8.8. Theorem 8.6 only ensures the existence of an appropriate multitree $\widecheck{\Xi}$ but does not give any information on its explicit construction. The same holds true for $\widehat{\Xi}$. In Section 8.5, we will discuss how we can construct the multitrees $\widecheck{\Xi}$ and $\widehat{\Xi}$ *without* setting up the compressed matrix $\mathbf{B}_\eta^\top$ so that $\widehat{\mathbf{r}}$ from (8.8) satisfies (**RES**). Furthermore, numerical experiments in Section 9 indicate appropriate choices of $\widecheck{\Xi}$ and $\widehat{\Xi}$ 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 $_{\widehat{\Lambda}}\mathbf{B}^\top\mathbf{B}_{\widehat{\Lambda}}$ by a *sparse* approximation $_{\widehat{\Lambda}}[\mathbf{B}_\eta^\top\mathbf{B}_\eta]_{\widehat{\Lambda}} := \mathbf{R}_{\widehat{\Lambda}}[\mathbf{B}_\eta^\top\mathbf{B}_\eta]\mathbf{E}_{\widehat{\Lambda}}$ satisfying $\|_{\widehat{\Lambda}}\mathbf{B}^\top\mathbf{B}_{\widehat{\Lambda}} - {}_{\widehat{\Lambda}}[\mathbf{B}_\eta^\top\mathbf{B}_\eta]_{\widehat{\Lambda}}\| \lesssim \eta$. In analogy to (6.9), we consider:

$$(8.9) \qquad \text{Find } \mathbf{u}_{\eta,\widehat{\Lambda}} \in \ell_2(\widehat{\Lambda}): \qquad {}_{\widehat{\Lambda}}[\mathbf{B}_\eta^\top\mathbf{B}_\eta]_{\widehat{\Lambda}}\mathbf{u}_{\eta,\widehat{\Lambda}} = \mathbf{R}_{\widehat{\Lambda}}\mathbf{B}_\eta^\top\mathbf{f}_{\check{\Lambda}}.$$

Indeed, under the assumption that η is sufficiently small, $\kappa(_{\widehat{\Lambda}}[\mathbf{B}_\eta^\top\mathbf{B}_\eta]_{\widehat{\Lambda}})$ is bounded independently of $\widehat{\Lambda}$ (see Appendix B). In particular, there exist algorithms based on linear iterative solvers like the conjugate gradient (cg) method that approximate (8.9) such that $\|\mathbf{u}_{\widehat{\Lambda}} - \mathbf{u}_{\eta,\widehat{\Lambda}}\|_{\ell_2(\widehat{\Lambda})}\| \lesssim \eta$ and (**GAL**) is satisfied. Similar 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 $\mathbf{w}_{\widehat{\Lambda}}$ as an approximate solution of the problem:

$$(8.10) \qquad \text{Find } \mathbf{x}_{\widehat{\Lambda}} \in \ell_2(\widehat{\Lambda}): \qquad {}_{\widehat{\Lambda}}\mathbf{B}_{\check{\Lambda}\widehat{\Lambda}}^\top\mathbf{B}_{\widehat{\Lambda}}\mathbf{x}_{\widehat{\Lambda}} = {}_{\widehat{\Lambda}}\mathbf{B}_{\check{\Lambda}}^\top\mathbf{f}_{\check{\Lambda}}.$$

We could choose $\check{\Lambda}$ as the smallest multitree that contains supp $\mathbf{B}_\eta\mathbf{v}_{\widehat{\Lambda}}$ for all $\mathbf{v}_{\widehat{\Lambda}} \in \ell_2(\widehat{\Lambda})$. However, this is *not* an implementable approach. Hence, we are concerned with the question how the multitree $\check{\Lambda}$ can be constructed in dependency of $\widehat{\Lambda}$ such that (1) the condition number of $_{\widehat{\Lambda}}\mathbf{B}_{\widehat{\Lambda}}^\top$ is uniformly bounded and (2) an approximate solution $\mathbf{w}_{\widehat{\Lambda}}$ to (8.10) satisfies (**GAL**). This will be discussed in Sections 8.5 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 $\widehat{\Lambda}_k \to \widehat{\Lambda}_{k+1}$ of the *trial* sets in Algorithm 2 is based upon the residual $\widehat{\mathbf{r}}_k$, but it is so not clear how to construct appropriate *test* sets $\check{\Lambda}_k = \check{\Lambda}_k(\widehat{\Lambda}_k)$. Similarly for the auxiliary sets $\widehat{\Xi}_k$ and $\check{\Xi}_k$ required for (8.8): While the construction of the test sets $\check{\Xi}_k$ for the primal residual in a Galerkin setting has been investigated in [KS13], there are so far no results for good choices of $\check{\Xi}_k$ and $\widehat{\Xi}_k$ within a Petrov-Galerkin framework.

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

$$(8.11) \qquad \widehat{\Lambda}_0 = \widehat{\Lambda}_{SG,J} := \{\boldsymbol{\lambda} \in \widehat{\mathcal{J}}: |\boldsymbol{\lambda}| \le J\},$$

$$(8.12) \qquad \check{\Lambda}_0 = \check{\Lambda}_{SG,J} := \{\boldsymbol{\lambda} \in \check{\mathcal{J}}: |\boldsymbol{\lambda}| \le J \text{ or } |\lambda_0| = J+1, |\lambda_i| = 0, 1 \le i \le n\},$$

where $|\boldsymbol{\lambda}| := \sum_{i=0}^n |\lambda_i|$.[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 $\widehat{\Lambda}_k$, $k > 0$, we propose the following (heuristic) choices:

(i) $\check{\Lambda}_{\text{Full}} = \textbf{FullStableExpansion}(\widehat{\Lambda}, \ell)$ is defined as

$$(8.13) \qquad \check{\Lambda}_{\text{Full}} := \big\{\boldsymbol{\lambda} \in \check{\mathcal{J}}: \exists\, \boldsymbol{\mu} \in \widehat{\Lambda} \text{ s.t. for all } j = 0, \ldots, n: |\lambda_j| \le |\mu_j| + \ell$$
$$\text{and } \operatorname{dist}\big(\operatorname{supp}\check{\psi}_{\lambda_j}^{(j)}, \operatorname{supp}\widehat{\psi}_{\mu_j}^{(j)}\big) \le \mathrm{D}_{\check{\Psi}^{(j)}}2^{-|\lambda_j|}\big\}.$$

---

[2] We will also use $\widehat{\Lambda}_{SG,J}$ and $\check{\Lambda}_{SG,J}$ within a uniform sparse grid (SG) discretization.

(ii) $\check{\boldsymbol{\Lambda}}_{\mathrm{Red}} = \mathbf{ReducedStableExpansion}(\widehat{\boldsymbol{\Lambda}}, \ell)$ is a subset of $\check{\boldsymbol{\Lambda}}_{\mathrm{Full}}$ defined as

$$(8.14) \quad \check{\boldsymbol{\Lambda}}_{\mathrm{Red}} := \bigcup_{i=0}^{n} \{\boldsymbol{\lambda} \in \check{\boldsymbol{\mathcal{J}}} : \exists \boldsymbol{\mu} \in \widehat{\boldsymbol{\Lambda}} \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{i,j}\,\ell$$
$$\text{and } \mathrm{dist}\big(\mathrm{supp}\,\check{\psi}_{\lambda_j}^{(j)}, \mathrm{supp}\,\widehat{\psi}_{\mu_j}^{(j)}\big) \leq \mathrm{D}_{\check{\Psi}^{(j)}} 2^{-|\lambda_j|}\}.$$

(iii) $\check{\boldsymbol{\Lambda}}_{\mathrm{Temp}} = \mathbf{TemporalStableExpansion}(\widehat{\boldsymbol{\Lambda}}, \ell)$: consists of only *temporal* higher level extensions, i.e.,

$$(8.15) \qquad \check{\boldsymbol{\Lambda}}_{\mathrm{Temp}} := \big\{\boldsymbol{\lambda} \in \check{\boldsymbol{\mathcal{J}}} : \exists \boldsymbol{\mu} \in \widehat{\boldsymbol{\Lambda}} \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{0,j}\,\ell$$
$$\text{and } \mathrm{dist}\big(\mathrm{supp}\,\check{\psi}_{\lambda_j}^{(j)}, \mathrm{supp}\,\widehat{\psi}_{\mu_j}^{(j)}\big) \leq \mathrm{D}_{\check{\Psi}^{(j)}} 2^{-|\lambda_j|}\big\}.$$

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 $[\check{\boldsymbol{\Lambda}}] = \mathbf{FullStableExpansion}[\widehat{\boldsymbol{\Lambda}}, \ell]$

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

Choice of sets $\widehat{\boldsymbol{\Xi}}_k$, $\check{\boldsymbol{\Xi}}_k$. The proposed index set reads

$$\widehat{\boldsymbol{\Xi}}_k = \mathbf{ReducedMultiTreeCone}(\widehat{\boldsymbol{\Lambda}}, \ell)$$

$$(8.16) \qquad := \bigcup_{i=0}^{n} \{\boldsymbol{\lambda} \in \widehat{\boldsymbol{\mathcal{J}}} : \exists \boldsymbol{\mu} \in \widehat{\boldsymbol{\Lambda}}_k \text{ s.t. for all } j = 0, \ldots, n : |\lambda_j| \leq |\mu_j| + \delta_{i,j}\,\ell$$
$$\text{and } \mathrm{dist}\big(\mathrm{supp}\,\widehat{\psi}_{\lambda_j}^{(j)}, \mathrm{supp}\,\widehat{\psi}_{\mu_j}^{(j)}\big) \leq \mathrm{D}_{\widehat{\Psi}^{(j)}} 2^{-|\lambda_j|}\}$$

It was shown in [KS12, KS13] that this index set for $\ell = 1$ and the analogously defined $\mathbf{FullMultiTreeCone}(\widehat{\boldsymbol{\Lambda}}, 1)$ are adequate choices for an accurate approximation of the primal residual in the Galerkin setting, where $\widehat{\boldsymbol{\Psi}}^{\mathcal{X}} = \check{\boldsymbol{\Psi}}^{\mathcal{Y}}$, $\widehat{\boldsymbol{\Xi}} = \check{\boldsymbol{\Xi}}$.

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 **FullResConstruction** and **OptimResConstruction**, see Figure 1. For the primal residual (i.e., in $\mathcal{Y}$), we expand $\widehat{\boldsymbol{\Lambda}}_k$ to $\widehat{\boldsymbol{\Xi}}_k^{\mathrm{tmp}} = \mathbf{ReducedMultiTreeCone}(\widehat{\boldsymbol{\Lambda}}_k, \ell)$ and obtain the desired $\check{\boldsymbol{\Xi}}_k$ 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 Fig. 1(a), we take the set $\check{\boldsymbol{\Xi}}_k$ as above and set $\widehat{\boldsymbol{\Xi}}_k = \mathbf{FullStableExpansion}(\check{\boldsymbol{\Xi}}_k, \ell)$ (with obvious inverted roles of primal and dual basis). Then, $\widehat{\boldsymbol{\Xi}}_k$ is the smallest multitree containing $\mathrm{supp}\,\mathbf{B}_\eta^\top \check{\mathbf{r}}_k$ for sufficiently small η. The second approach uses the by far smaller set $\widehat{\boldsymbol{\Xi}}_k = \widehat{\boldsymbol{\Xi}}_k^{\mathrm{tmp}}$ as indicated in Fig. 1(b), [KS13].

(a) **FullResConstruction**   (b) **OptimResConstruction**
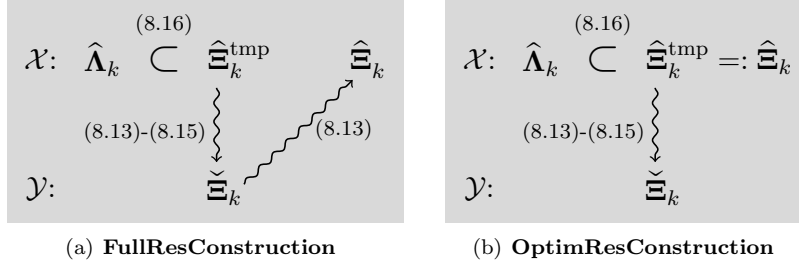
FIGURE 1. Constructions of index sets $\widehat{\breve{\boldsymbol{\Xi}}}_k$, $\breve{\boldsymbol{\Xi}}_k$ for residual approximation.

9. NUMERICAL EXPERIMENTS

We report numerical examples for time-periodic problems of type (2.1). We focus on the *stability of the arising normal equations* (8.10) in view of different choices for $\breve{\boldsymbol{\Lambda}}_k$. 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 $\boldsymbol{\Omega} = (0,1)$), since we employ an $L_2(0,1)$-orthonormal (multi-)wavelet basis $\boldsymbol{\Sigma} = \Sigma$ (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 $\mathbf{B}^\top \mathbf{B}$ and of $_{\widehat{\boldsymbol{\Lambda}}}\mathbf{B}_{\widehat{\boldsymbol{\Lambda}}\breve{\boldsymbol{\Lambda}}}^\top \mathbf{B}_{\widehat{\boldsymbol{\Lambda}}}$ 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 Θ^{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 Θ (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 $\delta = 0.7^3$, $\gamma_{\mathrm{ls}} = 0.01$ and, if not indicated differently, $\ell = 1$ for the stable extensions from Section 8.5. We obtain qualitatively similar results for choosing Σ as in [Dij09] for $d_x = \tilde{d}_x = 2$ even though they do not satisfy our assumptions.[4]

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 $\widehat{\boldsymbol{\Lambda}}_{SG,J}$, $\breve{\boldsymbol{\Lambda}}_{SG,J}$, $J = 0,1,\ldots$, as in (8.11), (8.12), e.g. [Zen91, BG04].

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

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

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 Fig. 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 δ 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 $s_{\max} = d - 1 = 1$, whereas uniform SG suffers from the lack of smoothness of the solution. We observe in Figure 2(b) that the iteration
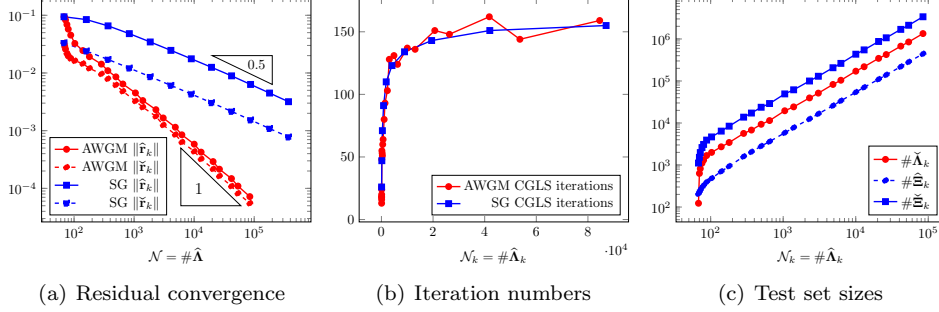

(a) Residual convergence   (b) Iteration numbers   (c) Test set sizes

FIGURE 2. Heat Equation Example: Comparison of **LS–AWGM** (AWGM) and Sparse Grids (SG).

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 $\check{\mathbf{\Lambda}}_k = \check{\mathbf{\Lambda}}_{\mathrm{Full}}$ yields stability. Figure 2(c) shows the cardinalities of the test sets. They grow only linearly with $\#\hat{\mathbf{\Lambda}}_k$, so that both $\mathbf{w}_{\hat{\mathbf{\Lambda}}_k}$ and $\hat{\mathbf{r}}_k$ can be computed within linear complexity in each iteration (cf. (**GAL**), (**RES**)). These results are based

(a) Convergence of dual residual   (b) Index set sizes $\#\hat{\mathbf{\Xi}}_k$
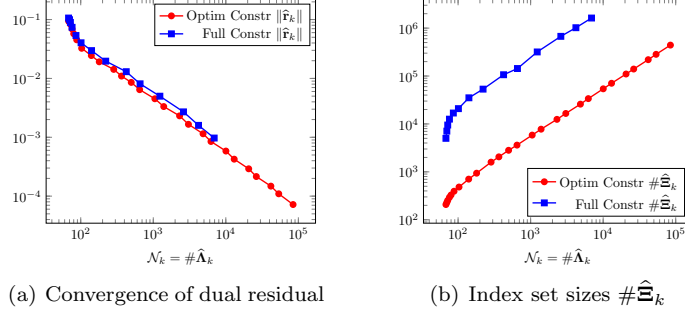
FIGURE 3. Heat Equation Example: Comparison of Residual Constructions

on **OptimResConstruction** for $\hat{\mathbf{\Xi}}_k$. In Figure 3 **FullResConstruction** is used. As $\hat{\mathbf{\Xi}}_k$ hardly impacts $\check{\mathbf{r}}_k$, we monitor only the dual residual. Since using a larger index set, $\|\hat{\mathbf{r}}_k\|_{\ell_2(\hat{\mathcal{J}})}$ is slightly increased (as expected), but it exhibits the same behaviour as **OptimResConstruction** (Fig. 3(a)). This marginal improvement comes at a high cost, $\#\hat{\mathbf{\Xi}}_k$ is 40-50 times larger, see Fig. 3(b).

Finally, in Figure 4, we compare the stable expansion types (Full, Reduced, Temporal). We find no discernible differences in the residual (Fig. 4(a)) and only a very slight increase in the iteration numbers in **GALSOLVE** (Fig. 4(b)). It seems that choosing $\check{\mathbf{\Xi}}_k = \mathbf{TemporalStableExpansion}(\hat{\mathbf{\Xi}}_k^{\mathrm{tmp}}, 1)$ yields results that are comparable to the other extensions, which could not be deduced from [KS12, KS13].

(a) Residual convergence   (b) Iteration numbers   (c) Test set Sszes (normal eq.)
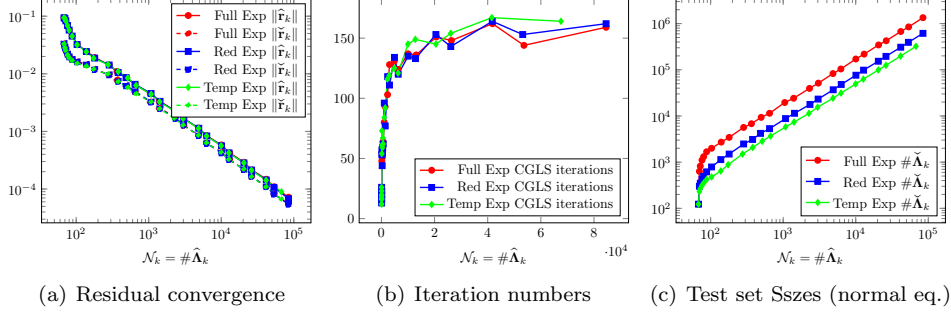
FIGURE 4. Heat Equation Example: Comparison of Stable Expansions

All three methods seem stable, and we can reduce the size of the test sets by a factor of about 3.4 for $\breve{\boldsymbol{\Lambda}}_k$ (and likewise by 2.5 for $\breve{\boldsymbol{\Xi}}_k$).

9.2. **Convection-Diffusion-Reaction Equation.** As a second example, we consider the convection-diffusion-reaction (CDR) equation

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

for a $f(t,x)$ 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.

(a) Solution $u(t,x)$ of the CDR example   (b) Support centres of basis functions in $\widehat{\boldsymbol{\Lambda}}_k$ for $k = 12$ ($\#\widehat{\boldsymbol{\Lambda}}_{12} = 9445$).
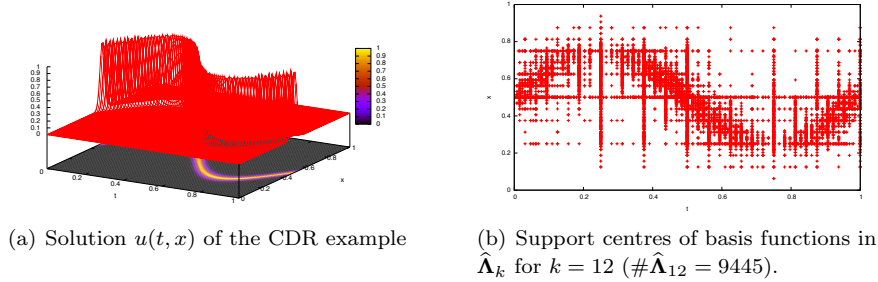
FIGURE 5. CDR Example: Solution and adaptive refinement

The support centers (i.e., the centers of supp $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda}}$, $\boldsymbol{\lambda} \in \widehat{\boldsymbol{\Lambda}}_k$) 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 (Fig. 6(b)), employing the optimized construction of $\widehat{\boldsymbol{\Xi}}_k$ and only temporal stable expansions for $\breve{\boldsymbol{\Lambda}}_k$, $\breve{\boldsymbol{\Xi}}_k$. 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.
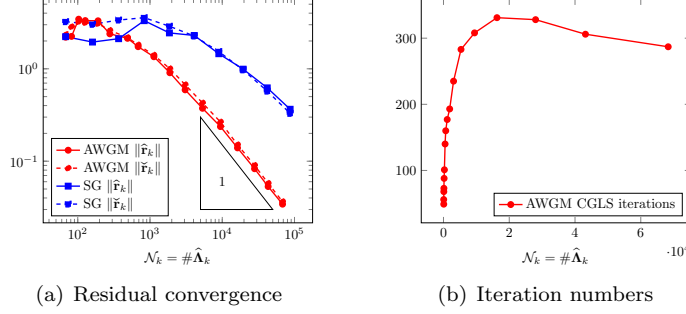
(a) Residual convergence   (b) Iteration numbers

FIGURE 6. CDR Example: Convergence and stability of **LS-AWGM** (AWGM) and Sparse Grids (SG)

(a) Residual convergence $\|\widehat{\mathbf{r}}_k\|$ (b) Reduction Factors $\widecheck{\mathbf{\Lambda}}_k$, $\widebreve{\mathbf{\Xi}}_k$ (c) Index set sizes $\#\widehat{\mathbf{\Xi}}_k$
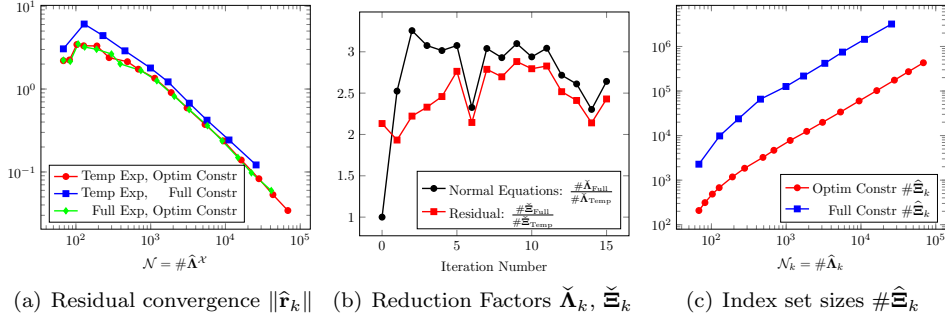
FIGURE 7. CDR Example: Comparison of different index set constructions.

Finally, we compare the above AWGM results with those obtained for larger sets, i.e., using full stable expansions and the **FullResConstruction** for $\widehat{\mathbf{\Xi}}_k$. As before, we see in Figures 7(a), 7(b) that we can reduce the size of the test sets $\widecheck{\mathbf{\Lambda}}_k$, $\widebreve{\mathbf{\Xi}}_k$ by factors 2 to 3 without losing accuracy. Likewise, the full construction of $\widehat{\mathbf{\Xi}}_k$ yields index sets that are approximately 20 times as large as for the optimized version (cf. Fig. 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 $\|\cdot\|_{\mathcal{X}}$, $\|\cdot\|_{\mathcal{Y}}$ as well as Cauchy-Schwarz's, Hölder's and Young's inequalities.
(2) Inf-sup condition. We consider an arbitrary $0 \neq w \in \mathcal{X}$ and define $z_w(t) := (\mathcal{A}(t)^*)^{-1} \dot{w}(t)$ for the adjoint $\mathcal{A}(t)^*$ of $\mathcal{A}(t)$. The bound $\|(\mathcal{A}(t)^*)^{-1}\| \leq \alpha^{-1}$ then yields for $v_w(t) := z_w(t) + w(t)$ that $\|v_w\|_{\mathcal{Y}} \leq \sqrt{2}\max\{1, \alpha^{-1}\}\|w\|_{\mathcal{X}} < \infty$. By definition of z_w and (2.2), $\langle z_w(t), \dot{w}(t)\rangle_{V \times V'} = \langle z_w(t), \mathcal{A}(t)[z_w(t)]\rangle_{V \times V'} \geq \alpha\|z_w(t)\|_V^2 \geq \frac{\alpha}{\gamma^2}\|\dot{w}(t)\|_{V'}^2$. Since $w \in \mathcal{X}$ is periodic, we have $\int_0^T \langle w, \dot{w}\rangle_{V \times V'} + \langle z_w, \mathcal{A}(t)[w]\rangle_{V \times V'} dt = \int_0^T \langle w, \dot{w}\rangle_{V \times V'} + \int_0^T \langle \dot{w}, w\rangle_{V \times V'} = \int_0^T \frac{d}{dt}\|w(t)\|_H^2 dt = \|w(T)\|_H^2 - \|w(0)\|_H^2 = 0$, so that we finally get $b(w, v_w) \geq \alpha \min\{1, \gamma^{-2}\}\|w\|_{\mathcal{X}}^2 \geq \frac{\alpha \min\{1, \gamma^{-2}\}}{\sqrt{2}\max\{1, \alpha^{-1}\}}\|w\|_{\mathcal{X}}\|v_w\|_{\mathcal{Y}} > 0$.

(3) Surjectivity. Let $0 \neq v \in \mathcal{Y}$. We aim to construct $z \in \mathcal{X}$ with $\langle w(t), \dot{z}(t) \rangle_{V \times V'} + \langle w(t), \mathcal{A}(t)[z(t)] \rangle_{V \times V'} = \langle w(t), \mathcal{A}(t)[v(t)] \rangle_{V \times V'}$ for all $w \in \mathcal{Y}$, and t a.e. on $(0, T)$, as then $b(z, v) = \int_0^T \langle v(t), \mathcal{A}(t)[v(t)] \rangle_{V \times V'} \geq \alpha \|v\|_{\mathcal{Y}}^2 > 0$, 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, $V_n := \text{span}\{\phi_i, i = 1, \ldots, n\}$, $z_n(t) := \sum_{i=1}^n z_i^{(n)}(t) \phi_i$. Then the linear system of ODEs $\langle w_n, \dot{z}_n(t) \rangle_{V \times V'} + \langle w_n, \mathcal{A}(t)[z_n(t)] \rangle_{V \times V'} = \langle w_n, \mathcal{A}(t)[v(t)] \rangle_{V \times V'}$, $z_n(0) = z_{n0}$, 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) A-priori estimates. (i), (2.2) and Young's inequality with some $\varepsilon < \frac{\alpha}{\gamma}$ yield

$$(A.1) \qquad \frac{1}{2} \frac{d}{dt} \|z_n(t)\|_H^2 + \alpha \|z_n(t)\|_V^2 \leq \gamma \varepsilon \|z_n(t)\|_V^2 + \frac{\gamma}{4\varepsilon} \|v(t)\|_V^2$$

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 $\sup_{s \in [0,T]} \|z_n(s)\|_H^2 < \infty$ and $\{z_n\}_{n \in \mathbb{N}}$ is uniformly bounded in $L_\infty(0, T; H)$. Similarly, we can conclude that $2(\alpha - \gamma \varepsilon) \|z_n\|_{\mathcal{Y}} \leq \|z_n(0)\|_H^2 - \|z_n(T)\|_H^2 + \frac{\gamma}{2\varepsilon} \|v\|_{\mathcal{Y}}^2 < \infty$, so that $\{z_n\}_{n \in \mathbb{N}}$ is also uniformly bounded in $\mathcal{Y}$.

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

$$(A.2) \qquad \|z_n(T)\|_H^2 \leq e^{-\bar{\alpha} T} \|z_n(0)\|_H^2 + \bar{c} e^{-\bar{\alpha} T} \int_0^T e^{\bar{\alpha} t} \|v(t)\|_V^2 \, dt.$$

Set $M := \{z \in V_n : \|z\|_H \leq R := K^{\frac{1}{2}} (1 - e^{-\bar{\alpha} T})^{-\frac{1}{2}}\}$, $K := \bar{c} e^{-\bar{\alpha} T} \int_0^T e^{\bar{\alpha} t} \|v(t)\|_V^2 dt$. 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 $S : M \to M$, $z_n(0) \mapsto z_n(T)$, is continuous, the existence of a fixed-point $S(\bar{z}_n) = \bar{z}_n \in M$ follows from Brouwer's fixed-point theorem. By the a-priori estimates, the sequence $\{\bar{z}_n\}_{n \in \mathbb{N}}$ 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 $z_{n0} = \bar{z}_n$. From the a-priori estimates, we have that $\{z_n\}$ is uniformly bounded in the separable space $\mathcal{Y}$, so that there exists a subsequence (also denoted $\{z_n\}$) converging weakly to some z in $\mathcal{Y}$. 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 $j = 1, \ldots, n$ $-\langle \dot{\theta} \phi_j, z_n \rangle = \langle \theta(0) \phi_j - \theta(T) \phi_j, \bar{z}_n \rangle_H + \langle \theta \phi_j, \mathcal{A}(t)[v - z_n] \rangle$. As $z_n \rightharpoonup z$ in $\mathcal{Y}$ and $\bar{z}_n \rightharpoonup \bar{z}$ in H, we can pass to the limit $n \to \infty$ and obtain

$$(A.3) \qquad -\langle \dot{\theta} \phi_j, z \rangle = \langle \theta(0) \phi_j - \theta(T) \phi_j, \bar{z} \rangle_H + \langle \theta \phi_j, \mathcal{A}(t)[v - z] \rangle.$$

This particularly holds true for all $\theta \in \mathcal{D}(I)$, so that $\dot{z} = \mathcal{A}(\cdot)(v - z)$ in the distributional sense and hence $\dot{z} \in L_2(0, T; V')$. Moreover, (A.3) implies that for $w \in C^1(0, T; V)$, we have $-\langle \dot{w}, z \rangle - \langle w(0) - w(T), \bar{z} \rangle = \langle w, \mathcal{A}(t)[v - z] \rangle = \langle \dot{z}, w \rangle = -\langle \dot{w}, z \rangle + \langle w(T), z(T) \rangle_H - \langle w(0), z(0) \rangle_H$, so that indeed $\bar{z} = z(0) = z(T)$ in H and hence $z \in \mathcal{X}$. With this z, the surjectivity condition is fulfilled.

Appendix B. Auxiliary wavelet compression results

Here, we report two facts for $\mathbf{B}$ 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 $\eta < 1$, $\mathbf{B}_\eta \in \mathcal{L}(\ell_2(\widehat{\mathcal{J}}), \ell_2(\check{\mathcal{J}}))$ and $\mathbf{B}_\eta^\top \mathbf{B}_\eta \in \mathcal{L}(\ell_2(\widehat{\mathcal{J}}), \ell_2(\widehat{\mathcal{J}}))$ are boundedly invertible with bounds depending on η.*

Proposition B.2 ([KS13])**.** *Let the assumptions of Theorem 8.6 hold. Then, there exists a constant ω_{ls} such that $\|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{w}_{\widehat{\boldsymbol{\Lambda}}}) - \widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})} \leq \omega_{\mathrm{ls}}\|\widehat{\mathbf{r}}\|_{\ell_2(\widehat{\mathcal{J}})}$, $\widehat{\mathbf{r}} := \mathbf{B}_\eta^\top \check{\mathbf{r}}$.*

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(2):829–857, May 2006.

[And13] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.*, 33(1):242–260, 2013.

[BG04] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comput.*, 70(233):27–75, 2001.

[CDD02] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - beyond the elliptic case. *Found. Comput. Math.*, 2:203–245, 2002.

[CDD03] A. Cohen, W. Dahmen, and R. DeVore. Sparse evaluation of compositions of functions using multiscale expansions. *SIAM J. Math. Anal.*, 35(2):279–303, 2003.

[CS11] N. G. Chegini and R. P. Stevenson. Adaptive wavelet schemes for parabolic problems: sparse matrices and numerical results. *SIAM J. Numer. Anal.*, 49(1):182–212, 2011.

[CS12] N. G. Chegini and R. P. Stevenson. The adaptive tensor product wavelet scheme: Sparse matrices and the application to singularly perturbed problems. *IMA J. Numer. Anal.*, 32(1):75–104, 2012.

[Dah97] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica*, 6:55–228, 1997.

[DeV98] R. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[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. Evolution Problems I*, volume 5. Springer-Verlag, Berlin, 1992.

[DSS09] T. J. Dijkema, C. Schwab, and R. P. Stevenson. An adaptive wavelet method for solving high-dimensional elliptic PDEs. *Constr. Approx.*, 30(3):423–455, 2009.

[GHS07] T. Gantumur, H. Harbrecht, and R. P. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comput.*, 76(258):615–629, 2007.

[GO95] M. Griebel and P. Oswald. Tensor product type subspace splitting and multilevel iterative methods for anisotropic problems. *Adv. Comp. Math.*, 4:171–206, 1995.

[GO07] M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing*, 81(1):1–34, 2007.

[HV95] G. Horton and S. Vandewalle. A space-time multigrid method for parabolic partial differential equations. *SIAM J. Sci. Comput.*, 16(4):848–864, July 1995.

[JPSU07] D. Jürgens, M. Palm, S. Singer, and K. Urban. Numerical optimization of the Voith-Schneider® Propeller. *ZAMM*, 87(10):698–710, 2007.

[KB06] Y. Kawajiri and L. T. Biegler. Optimization strategies for simulated moving bed and PowerFeed processes. *AIChE Journal*, 52(4):1343–1350, 2006.

[Kes13] S. Kestler. *On the adaptive tensor product wavelet Galerkin method with applications in finance*. PhD thesis, University of Ulm, 2013.

[KS12] S. Kestler and R. P. Stevenson. Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions. *Preprint (submitted)*, 2012.

[KS13] S. Kestler and R. P. Stevenson. An efficient approximate residual evaluation in the adaptive tensor product wavelet method. *to appear in J. Sci. Comp.*, 2013.

[MV07]   D. Meidner and B. Vexler. Adaptive space-time finite element methods for parabolic optimization problems. *SIAM J. Contr. Optim.*, 46(1):116–142, 2007.

[Nit06]   P.-A. Nitsche. Best N-term approximation spaces for tensor product wavelet bases. *Const. Approx.*, 24(1):49–70, 2006.

[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. P. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comput.*, 78(267):1293–1318, 2009.

[SS11]   C. Schwab and R. P. Stevenson. Fast evaluation of nonlinear functionals of tensor product wavelet expansions. *Numer. Math.*, 119(4):765–786, 2011.

[Ste09]   R. P. Stevenson. Adaptive wavelet methods for solving operator equations: An overview. In R. DeVore and A. Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation*, pages 543–598. Springer (Berlin), 2009.

[SU09]   W. Sickel and T. Ullrich. Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Th.*, 161:748–786, 2009.

[SZ99]   H.L. Smith and X.-Q. Zhao. Dynamics of a Periodically Pulsed Bio-reactor Model. *J. Diff. Eq.*, 155:368–404, 1999.

[UP12]   K. Urban and A. T. Patera. A new error bound for reduced basis approximation of parabolic partial differential equations. *C.R. Math.*, 350(3–4):203 – 207, 2012.

[UP13]   K. Urban and A.T. Patera. An improved error bound for reduced basis approximation of linear parabolic problems. *Math. Comput.*, June 2013. to appear.

[Urb09]   K. Urban. *Wavelet methods for elliptic partial differential equations.* Oxford University Press, 2009.

[Zen91]   C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel algorithms for partial differential equations*, Notes on numer. fluid mech., pages 241–251. Vieweg, 1991.

Sebastian Kestler, Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany

E-mail address: sebastian.kestler@uni-ulm.de

Kristina Steih, Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany

E-mail address: kristina.steih@uni-ulm.de

Karsten Urban, Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse 20, D-89069 Ulm, Germany

E-mail address: karsten.urban@uni-ulm.de