AN OPTIMAL ADAPTIVE TENSOR PRODUCT WAVELET SOLVER OF A SPACE-TIME FOSLS FORMULATION OF PARABOLIC EVOLUTION PROBLEMS

NIKOLAOS REKATSINAS AND ROB STEVENSON

ABSTRACT. In this work we construct a well-posed first order system least squares (FOSLS) simultaneously space-time formulation of parabolic PDEs. Using an adaptive wavelet solver, this problem is solved with the best possible rate in linear complexity. Thanks to the use of a basis that consists of tensor products of wavelets in space and time, this rate is equal to that when solving the corresponding stationary problem. Our findings are illustrated by numerical results.

1. INTRODUCTION

After pioneering earlier work in [BJ89, BJ90], in recent years one witnesses a renewed and growing interest in simultaneously space-time solvers for evolutionary PDEs (e.g. [And14, MST14, Ste15, LMN16, ECD16, DFW16]). Instead of reducing the problem to a coupled system of ODEs, as with the method of lines, or to a sequence of stationary PDEs, as with time marching schemes like Rothe's method ([CDD+14]), one aims to solve the problem as a whole. The motivations to do so are to obtain a reduction in the computational complexity or storage requirements by exploiting the product structure of the space-time cylinder for tensor product approximation (see [GO07]); to construct meshes that are optimally adapted to singularities that are local in space and time (improving upon the rather limited possibilities of local time stepping); and, finally, to use the additional time dimension to enhance the possibilities for a massive parallelization (e.g. [GN16]).

When aiming at a convergent or rather optimally convergent adaptive solution method, a major obstruction is that a space-time variational formulation does not lead to a bilinear form that is symmetric or coercive. In [SS09, GK11, KSU16], considering parabolic PDEs, this problem was tackled as follows: By equipping the (Bochner) spaces w.r.t. which the bilinear form is well-posed with (wavelet) Riesz bases, the variational problem has an equivalent formulation as a well-posed bi-infinite matrix-vector formulation. By now forming normal equations one arrives at a well-posed symmetric positive bi-infinite matrix-vector problem whose solution can be approximated at the best possible nonlinear approximation rate using the adaptive wavelet scheme developed in [CDD01, GHS07].

For both the evaluation of the a posteriori error estimator, and for solving the arising Galerkin systems, a key ingredient of these schemes is, for the current finite approximate solution vector, the approximate evaluation of the residual of the infinite system of equations within a tolerance that is equal to a fixed multiple of the current error, and at a cost such that the overall scheme has optimal computational complexity. In the original scheme this was performed by approximating separately both the right-hand side vector and the matrix-vector multiplication each within half this tolerance. For the second task an ingenious scheme was constructed, the so-called **apply**-routine, that approximates each column of the matrix with an accuracy dependent on the modulus of the corresponding entry in the input vector. To have a proper decay of the matrix entries away from the diagonal, it requires that the wavelets have sufficiently many vanishing moments dependent on their order. Even for linear operators the apply routine is non-linear, and unfortunately it turned out to be quantitatively quite demanding.

In [RS18a] we developed an alternative for the **apply** routine. It is based on the observation that if the wavelets are sufficiently smooth, then each entry of the residual vector equals the residual of the equation in mild form integrated against a wavelet. Now it can be shown that if the current approximation is from the span of the wavelets up to some level, then a sufficiently accurate residual vector approximation is obtained by ignoring all entries of this vector that correspond to wavelets whose levels exceed that level by a fixed constant. This observation also applies to non-uniform, adaptive settings as long as the approximate solutions are sought in spans of wavelets whose indices form trees. For this scheme, the number of vanishing moments of the wavelets does not have to grow with their order, and often one vanishing moment suffices. To compute the remaining entries of the residual vector in linear complexity, i.e., those that cannot be ignored, the current approximate solution is first written in a locally finite single-scale basis, then its residual is integrated against single-scale basis functions, after which the transposed multi-to-single scale transformation is applied. Without difficulty the scheme applies to semi-linear PDEs as well.

The key idea behind this alternative residual evaluation scheme is not to split the residual functional into two terms and to approximate them separately, but to represent both terms in one common dictionary before integrating their difference against a wavelet basis. In this way one benefits from the fact that the norm of this functional decreases proportionally to the error in the approximation, meaning that one is allowed to make a fixed relative error in the second step. The approach requires that the application of the PDO to any wavelet lands in L_2, for operators of 2nd order meaning that the (piecewise polynomial) wavelets have to be globally C^1. To avoid this unpleasant condition, in [RS18a] we showed that any well-posed semi-linear 2nd order PDE can be written as a well-posed First Order System Least Squares (FOSLS) problem to which the whole machinery can be applied. In this case, it suffices to have bases of globally continuous (piecewise polynomial) wavelets which are available on arbitrary polytopes.

In the current paper we apply the approach introduced in [RS18a] to parabolic evolutionary PDEs written in a well-posed simultaneously space-time variational formulation. We transform this problem to a well-posed FOSLS formulation, and equip the arising Bochner spaces with bases that are tensor products of temporal and spatial wavelet bases. The advantage of tensor product approximation is that

it allows to solve the full time evolution at an order of computational complexity that is equal as when solving the corresponding stationary problem.

This advantage, however, does not come for free. As a counterpart for the tree approximations with non-tensor product wavelet approximation for stationary problems studied in [RS18a], here we will allow only approximate solutions from spans of tensor-product wavelets whose index pairs form multi-trees (more particular, double trees). With this restriction, we will be able to show that a sufficiently accurate residual vector approximation is obtained as follows: Drop each entry of this residual vector that corresponds to a tensor-wavelet for which one or both its levels exceed by more than a fixed constant the corresponding level of any tensor-wavelet in the multi-tree with which it has an overlapping support.

To compute the remaining entries of the residual vector in linear complexity, we will rely on a generalisation from [KS14] of an algorithm originally developed in [BZ96] in the (non-adaptive) sparse-grid setting. Note that now we cannot rely on a residual evaluation by writing the current approximation in a locally finite single-scale representation, since the complexity of such a representation generally will not be of the order of complexity of the tensor-wavelet representation. The unavailability of such a locally finite single-scale representation has the consequence that we are not able to handle non-linear PDEs.

In a non-FOSLS setting, we studied adaptive tensor product wavelet methods for solving simultaneously space-time variational formulations of parabolic PDEs already in [SS09, CS11]. In those works, we considered unconstrained approximation, i.e. no multi-tree constraints, and relied on the **apply**-routine for the approximate residual evaluation. In [SS09] quantitative aspects were not considered, and in order to get a reasonably fast implementation in [CS11] we considered custom designed wavelets that for a PDO with constant coefficients on rectangular spatial domains yield a stiffness matrix that is truly sparse, and therefore can be applied exactly in optimal complexity.

This paper is organized as follows: In Sect. 2 we give a well-posed FOSLS formulation of a parabolic PDE. One of the residuals in this system will be measured in a dual norm. To circumvent its evaluation, we replace this norm by an equivalent sequence norm of this residual integrated against a wavelet basis.

In Sect. 3 we recall some relevant facts about the adaptive wavelet Galerkin method (**awgm**) for solving general well-posed operator equations.

In Sect. 4 we apply the **awgm** to the FOSLS formulation of the parabolic PDE. We construct tensor product bases for the arising Bochner spaces, specify the multi-tree constraint on the index sets of the bases, and investigate the best possible approximation rate that can be achieved. Most of our efforts will be devoted to the construction of an approximate residual evaluation such that the overall scheme converges with the best possible rate in linear complexity. A number of technical decay estimates will be postponed to the appendix.

In Sect. 5, we test our **awgm** on the heat equation on a two-dimensional L-shaped spatial domain with homogenous Dirichlet boundary conditions. In two examples where the initial condition satisfies the usual condition to vanish at the lateral boundary, we observe the best possible converge rate. In an example where the initial condition violates this lowest order compatibility condition, and thus the exact solution is discontinuous along a two-dimensional manifold, a reduced rate is

observed. We envisage that the 'full' rate will be restored by replacing the isotropic spatial wavelets by (piecewise) tensor product wavelets.

A short conclusion and a brief discussion of issues that are open to further investigations are presented in Sect. 6.

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

For normed linear spaces $\mathscr{A}$ and $\mathscr{B}$, $\mathcal{L}(\mathscr{A}, \mathscr{B})$ will denote the space of bounded linear mappings $\mathscr{A} \to \mathscr{B}$ endowed with the operator norm $\| \cdot \|_{\mathcal{L}(\mathscr{A}, \mathscr{B})}$. The subset of invertible operators in $\mathcal{L}(\mathscr{A}, \mathscr{B})$ with inverses in $\mathcal{L}(\mathscr{B}, \mathscr{A})$ will be denoted as $\mathcal{L}\mathrm{is}(\mathscr{A}, \mathscr{B})$.

For countable sets $\vee_1, \vee_2$, the norms on $\ell_2(\vee_i)$ and on $\mathcal{L}(\ell_2(\vee_1), \ell_2(\vee_2))$ will often simply denoted as $\| \cdot \|$.

2. Well-posed FOSLS formulation of a parabolic PDE

For a bounded domain $\Omega \subset \mathbb{R}^n$, $\mathrm{I} := (0, T)$ for some $T > 0$, and $A = A^\top \in L_\infty(\mathrm{I} \times \Omega)^{n \times n}$ with $\xi^\top A(\cdot)\xi \eqsim \|\xi\|^2$ ($\xi \in \mathbb{R}^n$, a.e. on $\mathrm{I} \times \Omega$), we consider the semi-linear parabolic time evolution problem

$$(2.1) \qquad \begin{cases} \frac{\partial u}{\partial t} - \nabla_x \cdot A\nabla_x u + N(u) = g & \text{on } \mathrm{I} \times \Omega, \\ u = 0 & \text{on } \mathrm{I} \times \partial\Omega, \\ u(0, \cdot) = h & \text{on } \Omega, \end{cases}$$

Multiplying this equation by smooth test functions v of time and space that vanish at $[0, T] \times \partial\Omega$, integrating both sides over time and space, and, applying integration by parts in space, we arrive at the variational problem of finding

$$u \in \boxed{\mathscr{U} := L_2(\mathrm{I}; H_0^1(\Omega)) \cap H^1(\mathrm{I}; H^{-1}(\Omega)),}$$

such that for all

$$v \in \boxed{\mathscr{V} = \mathscr{V}_1 \times \mathscr{V}_2 := L_2(\mathrm{I}; H_0^1(\Omega)) \times L_2(\Omega),}$$

it holds that

$$(Gu)(v) := \int_\mathrm{I} \int_\Omega \big(\frac{\partial u}{\partial t} + N(u) - g\big)v_1 + A\nabla_x u \cdot \nabla_x v_1 \, dx \, dt + \int_\Omega (u(0, \cdot) - h)v_2 \, dx = 0.$$

We assume that

- a solution u exists,
- G, i.e., N, is two times continuously Fréchet differentiable in a neighborhood of u,
- $DG(u) \in \mathcal{L}(\mathscr{U}, \mathscr{V}')$ is a homomorphism with its range,

i.e., that conditions (i)-(iii) from [RS18a, Sect. 2] are satisfied.

Remark 2.1 (linear case). In the case that for some $\vec{b} \in L_\infty(\mathrm{I} \times \Omega)^n$ and $c \in L_\infty(\mathrm{I} \times \Omega)$,

$$(2.2) \qquad\qquad\qquad N(u) = Nu := \vec{b} \cdot \nabla_x u + cu,$$

these conditions are known to be satisfied even with $DG(u) \in \mathcal{L}\mathrm{is}(\mathscr{U}, \mathscr{V}')$, see e.g. [DL92, Ch.XVIII], [SS09]. In this linear case, G is affine and so $DG(u)(v) = G(v) - G(0)$, and the solution u is unique.

For the verification of the conditions in semi-linear cases, we refer to e.g. [Tem97].

Using the general framework outlined in [RS18a, Sect. 2], we write our second order PDE as a *first order system least squares* (FOSLS) problem: With

$$\boxed{\vec{\mathscr{P}} := L_2(\mathrm{I}; L_2(\Omega)^n)},$$

it holds that

$$\vec{p} \mapsto \big((v_1, v_2) \mapsto \int_{\mathrm{I}} \int_{\Omega} \vec{p} \cdot \nabla_x v_1 \, dx\big) \in \mathcal{L}(\vec{\mathscr{P}}, \mathscr{V}'), \qquad u \mapsto A\nabla_x u \in \mathcal{L}(\mathscr{U}, \vec{\mathscr{P}}).$$

Consequently, for any solution u of $G(u) = 0$, it holds that $(u, \vec{p}) := (u, A\nabla_x u) \in \mathscr{U} \times \vec{\mathscr{P}}$ is a zero of the least squares functional

$$(2.3) \qquad Q(u, \vec{p}) := \tfrac{1}{2}\Big(\|H_1(u, \vec{p})\|_{\mathscr{V}_1'}^2 + \|H_2(u, \vec{p})\|_{L_2(\Omega)}^2 + \|H_3(u, \vec{p})\|_{\vec{\mathscr{P}}}^2\Big),$$

where

$$H_1(u, \vec{p}) := v_1 \mapsto \int_{\mathrm{I}} \int_{\Omega} \big(\frac{\partial u}{\partial t} + N(u) - g\big)v_1 + \vec{p} \cdot \nabla_x v_1 \, dx \, dt,$$
$$H_2(u, \vec{p}) := u(0, \cdot) - h, \quad H_3(u, \vec{p}) := \vec{p} - A\nabla_x u,$$

and so a solution of $DQ(u, \vec{p}) = 0$. This latter operator equation is well-posed in the sense that

- $DQ : \mathscr{U} \times \vec{\mathscr{P}} \supset \mathrm{dom}(DQ) \to (\mathscr{U} \times \vec{\mathscr{P}})'$ is continuously Fréchet differentiable in a neighborhood of any solution $(u, \vec{p})$,
- $0 < D^2Q(u, \vec{p}) = D^2Q(u, \vec{p})' \in \mathcal{L}\mathrm{is}(\mathscr{U} \times \vec{\mathscr{P}}, (\mathscr{U} \times \vec{\mathscr{P}})')$, so that any solution is $(u, \vec{p})$ locally unique,
- for $(u, \vec{p})$ being a zero of Q, it holds that

$$(2.4) \qquad\qquad Q(w, \vec{q}) \eqsim \|u - w\|_{\mathscr{U}}^2 + \|\vec{p} - \vec{q}\|_{\vec{\mathscr{P}}}^2$$

  in a neighborhood of $(u, \vec{p})$.

Remark 2.2 (linear case cont'd). If N is linear, and thus G is affine, then DQ is affine, and therefore $D^2Q(u, p)(v, \vec{q}) = DQ(v, \vec{q}) - DQ(0, \vec{0})$; in particular, D^2Q is constant. The equation $DQ(u, \vec{p}) = 0$ is uniquely solvable, and therefore its solution $(u, \vec{p})$ is the zero of Q and u is the solution of $G(u) = 0$. The relation (2.4) holds globally.

Remark 2.3. In [MS01, MS02] a least squares functional similar to (2.3) has been studied with, essentially, $\|\cdot\|_{\mathscr{V}_1'}^2$ being replaced by $\|\cdot\|_{L_2(\mathrm{I}; L_2(\Omega)^n)}^2$. As a consequence, a norm equivalence as in (2.4) could not be demonstrated, but on the other hand the evaluation of the dual norm is avoided.

To deal with the dual norm in the definition of Q, we equip $\mathscr{V}_1$ with a *Riesz basis*

$$\Psi^{\mathscr{V}_1} = \{\psi_\lambda^{\mathscr{V}_1} : \lambda \in \vee_{\mathscr{V}_1}\},$$

meaning that the *analysis operator*

$$\mathcal{F}_{\mathscr{V}_1} : g \mapsto g(\Psi^{\mathscr{V}_1}) := [g(\psi_\lambda^{\mathscr{V}_1})]_{\lambda \in \vee_{\mathscr{V}_1}} \in \mathcal{L}\mathrm{is}(\mathscr{V}_1', \ell_2(\vee_{\mathscr{V}_1})),$$

and so its adjoint, known as the *synthesis operator*,

$$\mathcal{F}_{\mathscr{V}_1}' : \mathbf{v} \mapsto \mathbf{v}^\top \Psi^{\mathscr{V}_1} := \sum_{\lambda \in \vee_{\mathscr{V}_1}} v_\lambda \psi_\lambda^{\mathscr{V}_1} \in \mathcal{L}\mathrm{is}(\ell_2(\vee_{\mathscr{V}_1}), \mathscr{V}_1).$$

In the definition of the least squares functional Q, we now *replace* the standard dual norm on $\mathscr{V}_1'$ by the *equivalent* norm $\|\mathcal{F}_{\mathscr{V}_1} \cdot \|_{\ell_2(\vee_{\mathscr{V}_1})}$, which yields that

$$DQ(u,\vec{p})(w,\vec{q}) = \langle DH_1(u,\vec{p})(w,\vec{q})(\Psi^{\mathscr{V}_1}), H_1(u,\vec{p})(\Psi^{\mathscr{V}_1})\rangle_{\ell_2(\vee_{\mathscr{V}_1})} +$$

$$\langle DH_2(u,\vec{p})(w,\vec{q}), H_2(u,\vec{p})\rangle_{L_2(\Omega)} + \langle DH_3(u,\vec{p})(w,\vec{q}), H_3(u,\vec{p})\rangle_{\vec{\mathscr{P}}}.$$

To solve this operator equation $DQ(u,\vec{p}) = 0$ we are going to apply the adaptive wavelet Galerkin method.

3. The adaptive wavelet Galerkin method (**awgm**)

In this section, we summarize findings about the **awgm** from [Ste14, CS15, RS18a]. Let

(I) $F : \mathscr{H} \supset \mathrm{dom}(F) \to \mathscr{H}'$, with $\mathscr{H}$ being a separable Hilbert space;

(II) $F(z) = 0$;

(III) F be continuously differentiable in a neighborhood of z;

(IV) $0 < DF(z) = DF(z)' \in \mathcal{L}\mathrm{is}(\mathscr{H}, \mathscr{H}')$.

In our applications, the triple $(F, \mathscr{H}, z)$ will read as $(DQ, \mathscr{U} \times \vec{\mathscr{P}}, (u,\vec{p}))$, so that we already know that (I)-(IV) are valid.

Let $\Psi = \{\psi_\lambda : \lambda \in \vee\}$ be a *Riesz basis* for $\mathscr{H}$, with analysis operator $\mathcal{F} : g \mapsto g(\Psi) \in \mathcal{L}\mathrm{is}(\mathscr{H}', \ell_2(\vee))$, and so synthesis operator $\mathcal{F}' : \mathbf{v} \mapsto \mathbf{v}^\top \Psi \in \mathcal{L}\mathrm{is}(\ell_2(\vee), \mathscr{H})$. For any $\Lambda \subset \vee$, we set

$$\ell_2(\Lambda) := \{\mathbf{v} \in \ell_2(\vee) : \mathrm{supp}\, \mathbf{v} \subset \Lambda\}.$$

For satisfying the forthcoming Condition 3.4 that concerns the computational cost, it will be relevant that Ψ is a basis of *wavelet* type.

Writing $z = \mathcal{F}'\mathbf{z}$, and with

$$\mathbf{F} := \mathcal{F}F\mathcal{F}' : \ell_2(\vee) \to \ell_2(\vee),$$

an *equivalent* formulation of $F(z) = 0$ is given by

$$\mathbf{F}(\mathbf{z}) = 0.$$

We are going to approximate $\mathbf{z}$, and so z, by a sequence of Galerkin approximations from the spans of increasingly larger sets of wavelets, which sets are created by an adaptive process. Given $\Lambda \subset \vee$, the Galerkin approximation $\mathbf{z}_\Lambda$, or equivalently, $z_\Lambda := \mathbf{z}_\Lambda^\top \Psi$, are the solutions of $\langle \mathbf{F}(\mathbf{z}_\Lambda), \mathbf{v}_\Lambda\rangle_{\ell_2(\vee)} = 0$ ($\mathbf{v}_\Lambda \in \ell_2(\Lambda)$), i.e., $\mathbf{F}(\mathbf{z}_\Lambda)|_\Lambda = 0$, and $F(z_\Lambda)(v_\Lambda) = 0$ ($v_\Lambda \in \mathrm{span}\{\psi_\lambda : \lambda \in \Lambda\}$), respectively.

In order to be able to construct efficient algorithms, it will be needed to consider only sets Λ from a certain subset of all finite subsets of $\vee$. This collection of so-called *admissible* Λ will specified later. For the moment, it suffices to know that the union of any two admissible sets is again admissible.

To provide a benchmark to evaluate our adaptive algorithm, for $s > 0$, we define the nonlinear *approximation class*

(3.1)

$$\mathcal{A}^s := \Big\{\mathbf{z} \in \ell_2(\vee) : \|\mathbf{z}\|_{\mathcal{A}^s} :=$$

$$\sup_{\varepsilon > 0} \varepsilon \times \min\big\{(\#\Lambda)^s : \Lambda \text{ is admissible}, \inf_{\tilde{\mathbf{z}} \in \ell_2(\Lambda)} \|\mathbf{z} - \tilde{\mathbf{z}}\| \leq \varepsilon\big\} < \infty\Big\}.$$

A vector $\mathbf{z}$ is in $\mathcal{A}^s$ if and only if there exists a sequence of admissible $(\Lambda_i)_i$, with $\lim_{i\to\infty} \#\Lambda_i = \infty$, such that $\sup_i \inf_{\mathbf{z}_i \in \ell_2(\Lambda_i)} (\#\Lambda_i)^s \|\mathbf{z} - \mathbf{z}_i\| < \infty$.

The adaptive wavelet Galerkin method (**awgm**) defined below produces a sequence of increasingly more accurate Galerkin approximations $\mathbf{z}_\Lambda$ to $\mathbf{z}$. The, generally, infinite residual $\mathbf{F}(\mathbf{z}_\Lambda)$ is used as an a posteriori error estimator.

This a posteriori error estimator guides an appropriate enlargement of the current set Λ using a bulk chasing strategy, so that the sequence of approximations converge with the best possible rate to $\mathbf{z}$. To arrive at an implementable method, that is even of optimal computational complexity, both the Galerkin solution and its residual are allowed to be computed inexactly within sufficiently small relative tolerances.

Algorithm 3.1 (awgm).

% Let $0 < \mu_0 \leq \mu_1 < 1$, $\delta, \gamma > 0$ be constants, $\Lambda_0 \subset \vee$ be admissible,
% and $\mathbf{z}_{\Lambda_0} \in \ell_2(\Lambda_0)$. Let $\mathbf{Z}$ be a neighborhood of $\mathbf{z} \in \ell_2(\vee)$.

`for` $i = 0, 1, \ldots$ `do`

(R) $\zeta := \frac{2\delta}{1+\delta}\|\tilde{\mathbf{r}}_{i-1}\|$.   *% (Read $\|\tilde{\mathbf{r}}_{-1}\|$ as some scalar $\approx \|\mathbf{z}\|$.)*
`do` $\zeta := \zeta/2$; *Compute $\tilde{\mathbf{r}}_i \in \ell_2(\vee)$ such that $\|\tilde{\mathbf{r}}_i - \mathbf{F}(\mathbf{z}_{\Lambda_i})\| \leq \zeta$.*
`until` $\zeta \leq \frac{\delta}{1+\delta}\|\tilde{\mathbf{r}}_i\|$.

(B) *Determine an admissible $\Lambda_{i+1} \supset \Lambda_i$ with $\|\tilde{\mathbf{r}}_i|_{\Lambda_{i+1}}\| \geq \mu_0\|\tilde{\mathbf{r}}_i\|$ such that $\#(\Lambda_{i+1} \setminus \Lambda_i) \lesssim \#(\tilde{\Lambda} \setminus \Lambda_i)$ for any admissible $\tilde{\Lambda} \supset \Lambda_i$ with $\|\tilde{\mathbf{r}}_i|_{\tilde{\Lambda}}\| \geq \mu_1\|\tilde{\mathbf{r}}_i\|$.*

(G) *Compute $\mathbf{z}_{\Lambda_{i+1}} \in \ell_2(\Lambda_{i+1}) \cap \mathbf{Z}$ with $\|\mathbf{F}(\mathbf{z}_{\Lambda_{i+1}})|_{\Lambda_{i+1}}\| \leq \gamma\|\tilde{\mathbf{r}}_i\|$.*

`endfor`

In step (R), by means of a loop in which an absolute tolerance is decreased, the true *residual* $\mathbf{F}(\mathbf{z}_{\Lambda_i})$ is approximated within a relative tolerance δ. In step (B), *bulk chasing* is performed on the approximate residual. The idea is to find a smallest admissible $\Lambda_{i+1} \supset \Lambda_i$ with $\|\tilde{\mathbf{r}}_i|_{\Lambda_{i+1}}\| \geq \mu_0\|\tilde{\mathbf{r}}_i\|$. For reasons of computational efficiency, the condition of having a truly smallest Λ_{i+1} has been relaxed. Finally, in step (G), a sufficiently accurate approximation of the *Galerkin* solution w.r.t. the new set Λ_{i+1} is determined.

Convergence of the adaptive wavelet Galerkin method, with the *best possible rate*, is stated in the following theorem.

Theorem 3.2 ([Ste14, Thm. 3.9]). *Let μ_1, γ, δ, $\inf_{\mathbf{v}_{\Lambda_0} \in \ell_2(\Lambda_0)} \|\mathbf{z} - \mathbf{v}_{\Lambda_0}\|$, $\|\mathbf{F}(\mathbf{z}_{\Lambda_0})|_{\Lambda_0}\|$, and the neighborhood $\mathbf{Z}$ of the solution $\mathbf{z}$ all be sufficiently small. Then, for some $\alpha = \alpha[\mu_0] < 1$, the sequence $(\mathbf{z}_{\Lambda_i})_i$ produced by **awgm** satisfies*

$$\|\mathbf{z} - \mathbf{z}_{\Lambda_i}\| \lesssim \alpha^i \|\mathbf{z} - \mathbf{z}_{\Lambda_0}\|.$$

If, for whatever $s > 0$, $\mathbf{z} \in \mathcal{A}^s$, then $\#(\Lambda_{i+1} \setminus \Lambda_0) \lesssim \|\mathbf{z} - \mathbf{z}_{\Lambda_i}\|^{-1/s}$.

The *computation* of the approximate Galerkin solution $\mathbf{z}_{\Lambda_{i+1}}$ can be implemented by performing the simple fixed point iteration

(3.2) $$\mathbf{z}_{\Lambda_{i+1}}^{(j+1)} = \mathbf{z}_{\Lambda_{i+1}}^{(j)} - \omega\mathbf{F}(\mathbf{z}_{\Lambda_{i+1}}^{(j)})|_{\Lambda_{i+1}}.$$

Taking $\omega > 0$ to be a sufficiently small constant and starting with $\mathbf{z}_{\Lambda_{i+1}}^{(0)} = \mathbf{z}_{\Lambda_i}$, a fixed number of iterations suffices to meet the condition $\|\mathbf{F}(\mathbf{z}_{\Lambda_{i+1}}^{(j+1)})|_{\Lambda_{i+1}}\| \leq \gamma\|\tilde{\mathbf{r}}_i\|$. This holds also true when each of the $\mathbf{F}(\cdot)|_{\Lambda_{i+1}}$ evaluations is performed within an absolute tolerance that is a sufficiently small fixed multiple of $\|\tilde{\mathbf{r}}_i\|$.

Optimal *computational* complexity of the **awgm** –meaning that the work to obtain an approximation within a given tolerance $\varepsilon > 0$ can be bounded on some

constant multiple of the bound on its support length from Thm. 3.2,– is guaranteed under the following two conditions concerning the cost of the "bulk chasing" process, and that of the approximate residual evaluation, respectively.

Condition 3.3. The determination of Λ_{i+1} in Algorithm 3.1 is performed in $\mathcal{O}(\# \operatorname{supp} \tilde{\mathbf{r}}_i + \#\Lambda_i)$ operations.

Condition 3.4 (Cost condition)**.** For a sufficiently small, fixed $\varsigma > 0$, there exists a neighborhood $\mathbf{Z}$ of the solution $\mathbf{z}$ of $\mathbf{F}(\mathbf{z}) = 0$, such that for all admissible $\Lambda \subset \vee$, $\tilde{\mathbf{z}} \in \ell_2(\Lambda) \cap \mathbf{Z}$, and $\varepsilon > 0$, there exists an $\tilde{\mathbf{r}} \in \ell_2(\vee)$ with

$$\|\mathbf{F}(\tilde{\mathbf{z}}) - \tilde{\mathbf{r}}\| \leq \varsigma \|\mathbf{z} - \tilde{\mathbf{z}}\| + \varepsilon,$$

that one can compute in $\mathcal{O}(\varepsilon^{-1/s} + \#\Lambda)$ operations. Here $s > 0$ is such that $\mathbf{z} \in \mathcal{A}^s$.

Under both conditions, the **awgm** has optimal computational complexity:

Theorem 3.5. *In the setting of Theorem 3.2, and under Conditions 3.3 and 3.4, not only $\#\mathbf{z}_{\Lambda_i}$, but also the number of arithmetic operations required by* **awgm** *for the computation of $\mathbf{z}_{\Lambda_i}$ is $\mathcal{O}(\|\mathbf{z} - \mathbf{z}_{\Lambda_i}\|^{-1/s})$.*

In the setting of $F = DQ$, and Q being the FOSLS-functional associated to our parabolic time evolution problem, we will be able to verify Condition 3.4 only when DQ is affine, i.e., when our PDO is *linear*, i.e., $N(u) = Nu$ as given in (2.2). In this case where

$$0 < D\mathbf{F}(z)^{\top} = D\mathbf{F}(z) \equiv D\mathbf{F} \colon \tilde{\mathbf{z}} \mapsto \mathbf{F}(\tilde{\mathbf{z}}) - \mathbf{F}(\mathbf{0}) \in \mathcal{L}(\ell_2(\vee), \ell_2(\vee)),$$

first of all the conditions in Theorem 3.2 of $\inf_{\mathbf{v}_{\Lambda_0} \in \ell_2(\Lambda_0)} \|\mathbf{z} - \mathbf{v}_{\Lambda_0}\|$, $\|\mathbf{F}(\mathbf{z}_{\Lambda_0})|_{\Lambda_0}\|$, and the neighborhood $\mathbf{Z}$ being sufficiently small can be dropped.

Secondly, for any $\Lambda \subset \vee$ and $\varepsilon > 0$, our approximate residual $\tilde{\mathbf{r}}$ as meant in Condition 3.4 will be of the form $\mathbf{A}_{\Lambda,\varepsilon} \tilde{\mathbf{z}} + \mathbf{b}_\varepsilon$, where $\mathbf{A}_{\Lambda,\varepsilon} = \mathbf{A}_{\Lambda,\varepsilon}^{\top} \in \mathcal{L}(\ell_2(\Lambda), \ell_2(\Lambda))$. The construction of $\tilde{\mathbf{r}}$ will show that, in addition to $\|\mathbf{F}(\tilde{\mathbf{z}}) - (\mathbf{A}_{\Lambda,\varepsilon} \tilde{\mathbf{z}} + \mathbf{b}_\varepsilon)\| \leq \varsigma \|\mathbf{F}(\tilde{\mathbf{z}})\| + \varepsilon$ for all $\tilde{\mathbf{z}} \in \ell_2(\Lambda)$, and thus $\|\mathbf{F}(\mathbf{0}) - \mathbf{b}_\varepsilon\| \leq \varepsilon$, it holds that $\|D\mathbf{F} - \mathbf{A}_{\Lambda,\varepsilon}\|_{\mathcal{L}(\ell_2(\Lambda),\ell_2(\vee))} \lesssim \zeta \|D\mathbf{F}\|_{\mathcal{L}(\ell_2(\vee),\ell_2(\vee))}$. Consequently, by taking ζ sufficiently small, this $\mathbf{A}_{\Lambda,\varepsilon}$ can be used for solving the arising Galerkin problems by a Krylov iteration assuming that the initial residual is computed sufficiently accurate (see [GHS07, Thm. 2.5] for details). This method is much more efficient than the simple fixed point iteration applicable in the general nonlinear case.

4. Application to the FOSLS formulation

4.1. **Expression for the residual.** As announced before, we apply the **awgm** to solving $DQ(u, \vec{p}) = 0$ where, in order to be able to satisfy the cost condition Condition 3.4, we take $N(u) = Nu$ as given in (2.2). Besides the Riesz basis $\Psi^{\mathscr{V}_1}$ for $\mathscr{V}_1$ introduced earlier, let $\Psi^{\mathscr{U}} := \{\psi_\lambda^{\mathscr{U}} : \lambda \in \vee_{\mathscr{U}}\}$ and $\Psi^{\vec{\mathscr{P}}} := \{\psi_\lambda^{\vec{\mathscr{P}}} : \lambda \in \vee_{\vec{\mathscr{P}}}\}$ be Riesz bases for $\mathscr{U}$ and $\vec{\mathscr{P}}$, respectively. Then

$$(\Psi^{\mathscr{U}}, \vec{0}_{\vec{\mathscr{P}}}) \cup (0_{\mathscr{U}}, \Psi^{\vec{\mathscr{P}}}) := \{(\psi_\lambda^{\mathscr{U}}, \vec{0}_{\vec{\mathscr{P}}}) : \lambda \in \vee_{\mathscr{U}}\} \cup \{(0_{\mathscr{U}}, \psi_\lambda^{\vec{\mathscr{P}}}) : \lambda \in \vee_{\vec{\mathscr{P}}}\}$$

is a Riesz basis for $\mathscr{U} \times \vec{\mathscr{P}}$, with analysis operator $\mathcal{F} \in \mathcal{L}\mathrm{is}((\mathscr{U} \times \vec{\mathscr{P}})', \ell_2(\vee_{\mathscr{U} \times \vec{\mathscr{P}}}))$, with $\vee_{\mathscr{U} \times \vec{\mathscr{P}}} := \vee_{\mathscr{U}} \cup \vee_{\vec{\mathscr{P}}}$. Since with $\mathscr{P} := L_2(\mathrm{I}, L_2(\Omega))$, it holds that $\vec{\mathscr{P}} = \mathscr{P}^n$, we select $\Psi^{\vec{\mathscr{P}}}$ of the form $\{\psi_\lambda^{\mathscr{P}} \vec{e}_i : (\lambda, i) \in \vee_{\vec{\mathscr{P}}} := \vee_{\mathscr{P}} \times \{1, \ldots, n\}\}$ with $\Psi^{\mathscr{P}} := \{\psi_\lambda^{\mathscr{P}} : \lambda \in \vee_{\mathscr{P}}\}$ being a Riesz basis for $\mathscr{P}$. We write $(u, \vec{p}) = \mathcal{F}'[\mathbf{u}^{\top} \ \mathbf{p}^{\top}]^{\top}$.

For the application of the **awgm**, for each $[\mathbf{w}^\top \, \mathbf{q}^\top]^\top$ supported on an admissible (and thus finite) subset of $\vee_{\mathscr{U} \times \vec{\mathscr{P}}}$ we have to construct a computationally efficient approximation for the residual $D\mathbf{Q}([\mathbf{w}^\top \, \mathbf{q}^\top]^\top)$, where $D\mathbf{Q} := \mathcal{F}DQ\mathcal{F}'$. For that goal we impose the condition that $\Psi^{\mathscr{P}} \subset L_2(\mathrm{I}; H^1(\Omega))$, so that

$$(4.1) \qquad\qquad \Psi^{\vec{\mathscr{P}}} \subset L_2(\mathrm{I}; H(\mathrm{div}; \Omega)).$$

Then with $(w, \vec{q}) := (\mathbf{w}^\top \Psi^{\mathscr{U}}, \mathbf{q}^\top \Psi^{\vec{\mathscr{P}}})$, we obtain that

$$(4.2) \quad \boxed{\begin{aligned} &D\mathbf{Q}([\mathbf{w}^\top \, \mathbf{q}^\top]^\top) = \\ &\begin{bmatrix} \langle (\frac{\partial}{\partial t} + N)\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1} \rangle_{L_2(\mathrm{I}\times\Omega)} \\ \langle \Psi^{\vec{\mathscr{P}}}, \nabla_x \Psi^{\mathscr{V}_1} \rangle_{L_2(\mathrm{I}\times\Omega)^n} \end{bmatrix} \langle \Psi^{\mathscr{V}_1}, \frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g \rangle_{L_2(\mathrm{I}\times\Omega)} \\ &+ \begin{bmatrix} \langle \Psi^{\mathscr{U}}(0, \cdot), w(0, \cdot) - h \rangle_{L_2(\Omega)} \\ 0 \end{bmatrix} + \begin{bmatrix} \langle -A\nabla_x \Psi^{\mathscr{U}}, \vec{q} - A\nabla_x w \rangle_{L_2(\mathrm{I}\times\Omega)^n} \\ \langle \Psi^{\vec{\mathscr{P}}}, \vec{q} - A\nabla_x w \rangle_{L_2(\mathrm{I}\times\Omega)^n} \end{bmatrix}, \end{aligned}}$$

where we applied (4.1), and the zero boundary conditions satisfied by $\Sigma^{\mathscr{V}_1}$, to write $\langle \nabla_x \Psi^{\mathscr{V}_1}, \vec{q} \rangle_{L_2(\mathrm{I}\times\Omega)^n}$ as $\langle \Psi^{\mathscr{V}_1}, -\nabla_x \cdot \vec{q} \rangle_{L_2(\mathrm{I}\times\Omega)}$. Note that the residual consists of three terms, each of them being essentially one of the three terms of the least squares functional (2.3) in strong form integrated against a wavelet basis.

In view of (2.4), for the current application of **awgm** the cost condition can be reformulated as follows.

Condition 3.4*. For a sufficiently small, fixed $\varsigma > 0$, for all admissible $\Lambda \subset \vee_{\mathscr{U} \times \vec{\mathscr{P}}}$, $[\mathbf{w}^\top \, \mathbf{q}^\top]^\top \in \ell_2(\Lambda)$, and $\varepsilon > 0$, there exists an $\tilde{\mathbf{r}} \in \ell_2(\vee_{\mathscr{U} \times \vec{\mathscr{P}}})$ with

$$(4.3) \quad \begin{aligned} &\|D\mathbf{Q}([\mathbf{w}^\top \, \mathbf{q}^\top]^\top) - \tilde{\mathbf{r}}\| \leq \\ &\varsigma \Big(\|\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g\|_{\mathscr{V}_1'} + \|w(0, \cdot) - h\|_{L_2(\Omega)} + \|\vec{q} - A\nabla_x w\|_{\vec{\mathscr{P}}} \Big) + \varepsilon, \end{aligned}$$

that one can compute in $\mathcal{O}(\varepsilon^{-1/s} + \#\Lambda)$ operations, where $s > 0$ is such that $[\mathbf{u}^\top \, \mathbf{p}^\top]^\top \in \mathcal{A}^s$.

4.2. **Tensor product bases.** In view of the definitions of $\mathscr{U}$, $\mathscr{V}_1$, and $\mathscr{P}$ as being Bochner spaces, their bases will consist of tensor products of functions of collections of temporal and spatial functions, apart from a normalisation in case of $\mathscr{U}$. For $* \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}\}$, let

$$\Theta^* = \{\theta_\lambda^* \colon \lambda \in \lhd_*\}$$

be collections of *'temporal'* wavelets on I, such that

$$\Theta^{\mathscr{V}_1}, \Theta^{\mathscr{P}} \text{ are Riesz bases for } L_2(\mathrm{I}).$$

We assume that $\Theta^{\mathscr{U}} \subset H^1(\mathrm{I})$, and that

$$\frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}}, \frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{H^1(\mathrm{I})}} \text{ are Riesz bases for } L_2(\mathrm{I}), H^1(\mathrm{I}),$$

respectively. Here with $\Theta^{\mathscr{U}}/\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}$, and similarly for other normalisations or collections, we mean the collection $\{\theta_\lambda^{\mathscr{U}}/\|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})} \colon \lambda \in \lhd_{\mathscr{U}}\}$.

For $* \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}\}$, let

$$\Sigma^* = \{\sigma_\mu^* \colon \mu \in \Diamond_*\}$$

be collections of *'spatial'* wavelets on Ω, such that,

$$\Sigma^{\mathscr{V}_1}, \Sigma^{\mathscr{P}} \text{ are Riesz bases for } H_0^1(\Omega), L_2(\Omega),$$

respectively, $\Sigma^{\mathscr{U}} \subset H_0^1(\Omega)$, and

$$(4.4) \qquad \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^{-1}(\Omega)}}, \; \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}} \text{ are Riesz bases for } H^{-1}(\Omega), H_0^1(\Omega),$$

respectively. An interpolation argument shows that, consequently, $\Sigma^{\mathscr{U}}/\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}$ is a Riesz basis for $L_2(\Omega)$.

Under the above assumptions, we have that

$$(4.5) \qquad \boxed{\Psi^{\mathscr{V}_1} := \Theta^{\mathscr{V}_1} \otimes \Sigma^{\mathscr{V}_1}, \quad \Psi^{\mathscr{P}} := \Theta^{\mathscr{P}} \otimes \Sigma^{\mathscr{P}}, \quad \Psi^{\mathscr{U}} := \frac{\Theta^{\mathscr{U}} \otimes \Sigma^{\mathscr{U}}}{\|\Theta^{\mathscr{U}} \otimes \Sigma^{\mathscr{U}}\|_{\mathscr{U}}}}$$

are Riesz bases for $\mathscr{V}_1$, $\mathscr{P}$, and $\mathscr{U}$ with index sets $\vee_* = \triangleleft_* \times \Diamond_*$ for $*$ being $\mathscr{V}_1$, $\mathscr{P}$, or $\mathscr{U}$, respectively. For the last statement we refer to [GO95].

4.3. Piecewise polynomial spatial and temporal wavelets.
For $* \in \{\mathscr{U}, \mathscr{P}, \mathscr{V}_1\}$, we collect a number of (standard) assumptions on the *spatial* wavelet collections $\Sigma^* = \{\sigma_\lambda^*: \lambda \in \Diamond_*\}$ on Ω. To each $\lambda \in \Diamond_*$, we associate a value $|\lambda| \in \mathbb{N}_0$, which is called the *level* of λ. We will assume that the elements of Σ^* are *locally supported, piecewise polynomial* of some degree m, w.r.t. *dyadically nested partitions* in the following sense:

- (s_1) There exists a collection $\mathcal{O}_\Omega := \{\omega: \omega \in \mathcal{O}_\Omega\}$ of closed polytopes, such that, with $|\omega| \in \mathbb{N}_0$ being the *level* of ω, $\mathrm{meas}(\omega \cap \omega') = 0$ when $|\omega| = |\omega'|$ and $\omega \neq \omega'$; for any $\ell \in \mathbb{N}_0$, $\bar{\Omega} = \cup_{|\omega|=\ell}\omega$; $\mathrm{diam}\,\omega \eqsim 2^{-|\omega|}$; and ω is the union of ω' for some ω' with $|\omega'| = |\omega| + 1$. We call ω the *parent* of its *children* ω'. Moreover, we assume that the $\omega \in \mathcal{O}_\Omega$ are uniformly *shape regular*, in the sense that they satisfy a uniform Lipschitz condition.
- (s_2) $\mathrm{supp}\,\sigma_\lambda^*$ is contained in a connected union of a uniformly bounded number of ω's with $|\omega| = |\lambda|$, and restricted to each of these ω's is σ_λ^* a polynomial of degree m.
- (s_3) Each ω is intersected by the supports of a uniformly bounded number of σ_λ^*'s with $|\lambda| = |\omega|$.
- (s_4) Σ^* has the *cancellation property of order 1* meaning that

$$\Big| \int_\Omega \sigma_\mu^* v \, dx \Big| \lesssim 2^{-|\mu|} \|\sigma_\mu^{\mathscr{U}}\|_{L_1(\Omega)} |v|_{W_\infty^1(\mathrm{supp}\,\sigma_\mu^*)} \quad (\sigma \in \Diamond_{\mathscr{U}}, \, v \in W_\infty^1(\Omega) \cap H_0^1(\Omega)).$$

Generally, the polynomial degree m will be different for the different bases, but otherwise fixed. The collection $\mathcal{O}_\Omega$ is shared among all bases.

In addition to (s_1)–(s_4), we assume that $\Sigma^{\mathscr{U}}$ has the *cancellation properties of order 2*:

($s_4^{\mathscr{U}}$) $\Big| \int_\Omega \sigma_\mu^{\mathscr{U}} v \, dx \Big| \lesssim 4^{-|\mu|} \|\sigma_\mu^{\mathscr{U}}\|_{L_1(\Omega)} |v|_{W_\infty^2(\mathrm{supp}\,\sigma_\mu^{\mathscr{U}})} \; (\sigma \in \Diamond_{\mathscr{U}}, \, v \in W_\infty^2(\Omega) \cap H_0^1(\Omega)).$

Wavelets of in principle arbitrary order that satisfy all these assumptions can be found in e.g. [DS99, NS09].

Remark 4.1. In both (s_4) and ($s_4^{\mathscr{U}}$), $\mathrm{supp}\,\sigma_\mu^*$ could be read as a neighborhood of this support of diameter $2^{-|\mu|}$, which requires an only minor adaptation of some proofs.

Definition 4.2 (tiling). A collection $\mathcal{T} \subset \mathcal{O}_\Omega$ such that $\overline{\Omega} = \cup_{\omega \in \mathcal{T}}\omega$, and for $\omega_1 \neq \omega_2 \in \mathcal{T}$, $\mathrm{meas}(\omega_1 \cap \omega_2) = 0$ will be called a *tiling*. With $\mathcal{P}_m(\mathcal{T})$, we denote the space of piecewise polynomials of degree m w.r.t. $\mathcal{T}$. The smallest common refinement of tilings $\mathcal{T}_1$ and $\mathcal{T}_2$ is denoted as $\mathcal{T}_1 \oplus \mathcal{T}_2$.

To be able to find, in linear complexity, a representation of a function, given as linear combination of wavelets, as a piecewise polynomial w.r.t. a tiling we will impose a tree constraint on the underlying set of wavelet indices:

Definition 4.3 (trees). To each $\lambda \in \Diamond_*$ with $|\lambda| > 0$, we associate one $\mu \in \Diamond_*$ with $|\mu| = |\lambda| - 1$ and $\mathrm{meas}(\mathrm{supp}\,\sigma_\lambda^* \cap \mathrm{supp}\,\sigma_\mu^*) > 0$. We call μ the *parent* of λ, and so λ a *child* of μ.

To each $\lambda \in \Diamond_*$, we associate some neighbourhood $\mathcal{S}(\sigma_\lambda^*)$ of $\mathrm{supp}\,\sigma_\lambda^*$, with diameter $\lesssim 2^{-|\lambda|}$, such that $\mathcal{S}(\sigma_\lambda^*) \subset \mathcal{S}(\sigma_\mu^*)$ when λ is a child of μ.

We call a finite $\Lambda \subset \Diamond_*$ a *tree*, if it contains all $\lambda \in \Diamond_*$ with $|\lambda| = 0$, as well as the parent of any $\lambda \in \Lambda$ with $|\lambda| > 0$.

Remark 4.4. Note that we have parent-child relations on the set $\mathcal{O}_\Omega$ of polytopes as well as on the index sets $\Diamond_*$ (and similarly later on the index sets $\lhd_*$). We trust that no confusion will arise.

For some collections of wavelets, as the Haar or more generally, Alpert wavelets ([Alp93]), it suffices to take $\mathcal{S}(\sigma_\lambda^*) := \mathrm{supp}\,\sigma_\lambda^*$ in order to satisfy the nestedness assumption made in Definition 4.3. The next result shows that, thanks to (s_1)-(s_2), a suitable neighbourhood $\mathcal{S}(\sigma_\lambda^*)$ always exist.

Lemma 4.5. *With $C := \sup_{\lambda \in \Diamond_*} 2^{|\lambda|} \mathrm{diam}\,\mathrm{supp}\,\sigma_\lambda^*$, a valid choice of $\mathcal{S}(\sigma_\lambda^*)$ is given by $\{x \in \Omega\colon \mathrm{dist}(x, \mathrm{supp}\,\sigma_\lambda^*) \leq C2^{-|\lambda|}\}$.*

A proof of the following proposition, as well as an algorithm to apply the multi-to-single-scale transformation that is mentioned, is given in [Ste14, §4.3].

Proposition 4.6 (tree-to-tiling). *Given a tree $\Lambda \subset \Diamond_*$, there exists a tiling $\mathcal{T}(\Lambda) \subset \mathcal{O}_\Omega$ with $\#\mathcal{T}(\Lambda) \lesssim \#\Lambda$ such that $\mathrm{span}\{\sigma_\lambda^*\colon \lambda \in \Lambda\} \subset \mathcal{P}_m(\mathcal{T}(\Lambda))$. Moreover, equipping $\mathcal{P}_m(\mathcal{T}(\Lambda))$ with a basis of functions, each of which supported in ω for one $\omega \in \mathcal{T}(\Lambda)$, the representation of this embedding, known as the multi- to single-scale transform, can be applied in $\mathcal{O}(\#\Lambda)$ operations.*

Conversely, given a tiling, we define an element-tree and, given an integer k, a wavelet-tree:

Definition 4.7 (tiling-to-tree). Given a tiling $\mathcal{T} \subset \mathcal{O}_\Omega$, let $t(\mathcal{T}) \subset \mathcal{O}_\Omega$ be its enlargement by adding all ancestors of all $\omega \in \mathcal{T}$. Given a $k \in \mathbb{N}_0$, we set the *k-neighborhood of $\mathcal{T}$ in $\Diamond_*$* by

$$\Diamond_*(\mathcal{T}, k) := \big\{\lambda \in \Diamond_*\colon \mathrm{meas}\,\big(\mathcal{S}(\sigma_\lambda^*) \cap \bigcup_{\{\omega \in t(\mathcal{T})\colon |\omega| = \max(|\lambda| - k, 0)\}} \omega\big) > 0\big\}.$$

Proposition 4.8. *The set $\Diamond_*(\mathcal{T}, k)$ is a tree, and $\#\Diamond_*(\mathcal{T}, k) \lesssim \#\mathcal{T}$ (dependent on $k \in \mathbb{N}_0$).*

Remark 4.9. The idea behind the definitions related to tilings is the following: With the application of the **awgm** to FOSLS formulations of PDEs where the wavelet bases are of non-tensor product form as studied in [RS18a], the residual consists of terms of the form $\langle \Sigma^*, g\rangle_{L_2(\Omega)}$, where, for some tiling $\mathcal{T}$, $g \in \mathcal{P}_m(\mathcal{T})$ because it is from the span of a set of wavelets with indices from a tree. Now estimates of the form $\lim_{k\to\infty} \sup_{0 \neq g \in \mathcal{P}_m(\mathcal{T})} \frac{\|\langle \Sigma^*, g\rangle_{L_2(\Omega)}|_{\Diamond_* \setminus \Diamond_*(\mathcal{T}, k)}\|}{\|g\|_{*'}} = 0$ were shown, meaning that in order to approximate $\langle \Sigma^*, g\rangle_{L_2(\Omega)}$ within some given relative error it is sufficient to compute this vector on a k-neighborhood of $\mathcal{T}$ in $\Diamond_*$, where k is a suitable

constant. Furthermore, with the aid of multi- to locally single-scale transformations, $\langle \Sigma^*, g \rangle_{L_2(\Omega)}|_{\Diamond_*(\mathcal{T},k)}$ can be exactly evaluated in $\lesssim \#\Diamond_*(\mathcal{T},k) \lesssim \#\mathcal{T}$ operations.

Such results, together with analogous ones for temporal wavelets, will be the basis for the residual approximation in the current setting of the application of tensor product wavelets, where will restrict to approximations from spans of sets of wavelets with indices that from multi-trees.

Moving to the *temporal* wavelet collections, for $* \in \{\mathscr{U}, \mathscr{P}, \mathscr{V}_1\}$ we assume that $\Theta^* = \{\theta^*_\lambda \colon \lambda \in \lhd_*\}$ satisfies conditions (t_1)–(t_3) analogous to (s_1)–(s_3) with $\mathcal{O}_\Omega$ reading as $\mathcal{O}_\mathrm{I} = \{[i2^{-\ell}T, (i+1)2^{-\ell}T] \colon \ell \in \mathbb{N}_0, \, i = 0, \ldots, 2^\ell - 1\}$, and the level $|\omega|$ of $\omega = [i2^{-\ell}T, (i+1)2^{-\ell}T]$ being defined as ℓ. In addition, we assume that Θ^* has the *cancellation property of order 1*:

(t_4) $|\int_\mathrm{I} \theta^*_\mu v \, dx| \lesssim 2^{-|\mu|} \|\theta^*_\mu\|_{L_1(\mathrm{I})} |v|_{W^1_\infty(\mathrm{supp}\,\theta^*_\mu)}$ $(\sigma \in \lhd_*, \, v \in W^1_\infty(\mathrm{I}))$.

Remark 4.10. Compared to (s_4), note that (t_4) is imposed for all $v \in W^1_\infty(\mathrm{I})$ instead of for only $v \in W^1_\infty(\mathrm{I}) \cap H^1_0(\mathrm{I})$.

To each $\lambda \in \lhd_*$ with $|\lambda| > 0$, we associate one $\mu \in \lhd_*$ with $|\mu| = |\lambda| - 1$ and $\mathrm{meas}(\mathrm{supp}\,\theta^*_\lambda \cap \mathrm{supp}\,\theta^*_\mu) > 0$. We call μ the *parent* of λ, and so λ a *child* of μ. To each $\lambda \in \lhd_*$, we associate some neighbourhood $\mathcal{S}(\theta^*_\lambda)$ of $\mathrm{supp}\,\theta^*_\lambda$, with diameter $\lesssim 2^{-|\lambda|}$, such that $\mathcal{S}(\theta^*_\lambda) \subset \mathcal{S}(\theta^*_\mu)$ when λ is a child of μ. We call a finite $\Lambda \subset \lhd_*$ a *tree*, if it contains all $\lambda \in \lhd_*$ with $|\lambda| = 0$, as well as the parent of any $\lambda \in \Lambda$ with $|\lambda| > 0$.

Finally in this subsection, we add one more assumption on our PDE: We assume that its coefficients

(4.6) A, $\vec{b}$, and c (cf. (2.2)) are piecewise polynomial w.r.t. the coarsest possible tiling $\{\omega_\mathrm{I} \times \omega_\Omega \colon (\omega_\mathrm{I}, \omega_\Omega) \in \mathcal{O}_\mathrm{I} \times \mathcal{O}_\Omega, \, |\omega_\mathrm{I}| = |\omega_\Omega| = 0\}$ of $\bar{I} \times \bar{\Omega}$.

4.4. **Alpert wavelets.** Recall the least squares functional Q from (2.3). It consists of three 'residuals' $\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g$, $w(0, \cdot) - h$, and $\vec{q} - A\nabla_x w$ (not to be confused with the residual $D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top)$), whose norms are minimized. A main ingredient of our approximate evaluation of $D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top)$ will consist of representing all terms in each of the three 'residuals' in a common dictionary. If w and $\vec{q}$ were from spans of sets of non-tensor product wavelets whose index sets form trees, then such a dictionary can consist of the piecewise polynomials of some degree w.r.t. a tiling whose cardinality is of the order of the sum of the cardinalities of both trees. This is the setting considered in [RS18a] for the solution of stationary PDEs. In the current setting of tensor product approximation, such a 'single-scale' representation of optimal cardinality does not exist unless we put conditions on the wavelet index sets that are so restrictive that the advantages of tensor product approximation concerning favourable approximation rates are lost.

Instead, focussing to the first and third 'residual', we employ a representation in terms of tensor products of temporal and spatial Alpert wavelets. Unfortunately, this procedure does not apply to nonlinear terms being the reason for our restriction to $N(u) = Nu$ from (2.2).

Definition 4.11 (Alpert wavelets [Alp93]). We let $\Theta^a = \{\theta^a_\lambda \colon \lambda \in \lhd_a\}$ denote an orthonormal basis for $L_2(\mathrm{I})$ such that $\mathrm{supp}\,\theta^a_\lambda = \omega$ for some $\omega \subset \mathcal{O}_\mathrm{I}$ with $|\omega| = \max(|\lambda| - 1, 0)$, and $\mathrm{span}\{\theta^a_\lambda \colon |\lambda| \leq \ell\} = \mathcal{P}_m(\{\omega \in \mathcal{O}_\mathrm{I} \colon |\omega| = \ell\})$.

Similarly, we let $\Sigma^a = \{\sigma_\mu^a \colon \mu \in \Diamond_a\}$ denote an orthonormal basis for $L_2(\Omega)$ such that $\operatorname{supp} \sigma_\mu^a = \omega$ for some $\omega \subset \mathcal{O}_\Omega$ with $|\omega| = \max(|\mu| - 1, 0)$, and $\operatorname{span}\{\sigma_\mu^a \colon |\lambda| \leq \ell\} = \mathcal{P}_m(\{\omega \in \mathcal{O}_\Omega \colon |\omega| = \ell\})$.

We set $\Psi^a := \Theta^a \otimes \Sigma^a$.

4.5. **Multi-tree approximation.** We need a definition of admissible subsets of the index set of our basis for $\mathscr{U} \times \vec{\mathscr{P}}$ that on the one hand is sufficiently restrictive to allow for the evaluation of the approximate residuals in linear complexity, and on the other hand yields the favourable approximation rates known from unconstrained tensor product approximation. For that goal we consider multi-trees as a substitute for the concept of a tree in the non-tensor product case.

Definition 4.12 (multi-trees). For $* \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}, a\}$, $\Lambda \subset \vee_* := \vartriangleleft_* \times \Diamond_*$ is called a *multi-tree* when for any $(\lambda, \mu) \in \Lambda$, $\Lambda_2(\lambda) := \{\gamma \colon (\lambda, \gamma) \in \Lambda\}$ and $\Lambda_1(\mu) := \{\gamma \colon (\gamma, \mu) \in \Lambda\}$ are trees in $\Diamond_*$ and $\vartriangleleft_*$, respectively. We set $\Lambda_2 := \cup_{\lambda \in \vartriangleleft_*} \Lambda_2(\lambda)$, $\Lambda_1 := \cup_{\mu \in \Diamond_*} \Lambda_1(\mu)$.

If $\Lambda \subset \vee_*$ is multi-tree, then Λ_1 and Λ_2, being unions of trees, are trees in $\vartriangleleft_*$ and $\Diamond_*$, respectively.

Simple examples of multi-trees, suited for approximation of functions without local singularities, are sets $\{(\lambda, \mu) \in \vee_* \colon (|\lambda|, |\mu|) \in S\}$ for finite $\emptyset \neq S \subset \mathbb{N}_0^2$ with the property that if $(i, j) \in S$ then $\{(\max(i - 1, 0), j), (i, \max(j - 1, 0))\} \in S$. Examples of such multi-trees are index sets corresponding to 'full' or 'sparse-grids', see [BG04].

Concerning the efficient computation of residuals we recall the following result from [KS14] that builds on earlier work from [BZ96] dealing with sparse-grids: Let a be a bilinear form such that for $u(t, x) = u_1(t)u_2(x)$ and $v(t, x) = v_1(t)v_2(x)$, it holds that $a(u, v) = a_1(u_1, v_1)a_2(u_2, v_2)$, where the a_i are local, i.e., $a_i(u_i, v_i) = 0$ when $|\operatorname{supp} u_i \cap \operatorname{supp} v_i| = 0$, and such that for $\omega \in \mathcal{O}_\Omega$ or $\omega \in \mathcal{O}_\mathrm{I}$ and $p, q \in \mathcal{P}_m(\omega)$ the evaluation of $a_i(p, q)$ can be performed in $\mathcal{O}(1)$ operations. Then for $*, \circ \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}, a\}$, multi-trees $\Lambda_* \subset \vee_*$, $\Lambda_\circ \subset \vee_\circ$, and $\mathbf{w} \in \ell_2(\Lambda_*)$, the matrix-vector product

$$(4.7) \qquad \left(a(\Psi^\circ, \Psi^*)\mathbf{w}\right)|_{\Lambda_\circ}$$

can be evaluated in $\mathcal{O}(\max(\#\Lambda^*, \#\Lambda^\circ))$ operations.

Although the definition in [KS14] of a tree and therefore that of a multi-tree are slightly different from the current definitions, the results from [KS14] carry over to the current setting without much difficulty. For details we refer to [Rek18].

Recalling that the solution of our operator equation $DQ(u, \vec{p}) = 0$ lives in $\mathscr{U} \times \vec{\mathscr{P}}$, being the Cartesian product of $\mathscr{U}$ and the n-fold Cartesian product of $\mathscr{P}$, and which has been equipped with Riesz basis $(\Psi^{\mathscr{U}}, \vec{0}_{\vec{\mathscr{P}}}) \cup (0_{\mathscr{U}}, \Psi^{\vec{\mathscr{P}}})$ with index set $\vee_{\mathscr{U} \times \vec{\mathscr{P}}} = \vee_{\mathscr{U}} \cup \vee_{\vec{\mathscr{P}}} = \vee_{\mathscr{U}} \cup (\vee_{\mathscr{P}}, 1) \cup \cdots \cup (\vee_{\mathscr{P}}, n)$, the following definition is natural.

Definition 4.13 (admissible index sets). A set $\Lambda \subset \vee_{\mathscr{U} \times \vec{\mathscr{P}}}$ is called *admissible* when $\Lambda_{\mathscr{U}} := \Lambda \cap \vee_{\mathscr{U}}$ and, for all $1 \leq i \leq n$, $\Lambda_{\vec{\mathscr{P}}_i} := \{\lambda \colon (\lambda, i) \in \Lambda \cap (\vee_{\mathscr{P}}, i)\} \subset \vee_{\mathscr{P}}$ are multi-trees. We set $\Lambda_{\vec{\mathscr{P}}} := \Lambda \cap \vee_{\vec{\mathscr{P}}}$.

Definition 4.14. For $*, \circ \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}, a\}$, a multi-tree $\Lambda \subset \vee_*$, and $k \in \mathbb{N}_0$, we define its *k-neighborhood in $\vee_\circ$* by

$$\vee_\circ(\Lambda, k) := \{(\lambda', \mu') \in \vee_\circ \colon \exists (\lambda, \mu) \in \Lambda \text{ with } |\lambda| = \max(|\lambda'| - k, 0),$$

$$|\mu| = \max(|\mu'| - k, 0), \text{and meas} \left(\mathcal{S}(\theta^\circ_{\lambda'}) \times \mathcal{S}(\sigma^\circ_{\mu'}) \cap \mathcal{S}(\theta^*_\lambda) \times \mathcal{S}(\sigma^*_\mu)\right) > 0\}.$$

We set $\vee_{\vec{\mathscr{P}}}(\Lambda, k) := \cup_{i=1}^n (\vee_\mathscr{P}(\Lambda, k), i)$.

Proposition 4.15. *The k-neighborhood $\vee_\circ(\Lambda, k)$, defined in Definition 4.14, is a multi-tree, and $\# \vee_\circ (\Lambda, k) \lesssim \#\Lambda$ (dependent on k).*

Proof. Thanks to Λ being a multi-tree, the definition of $\vee_\circ(\Lambda, k)$ would not change if the conditions $|\lambda| = \max(|\lambda'| - k, 0)$, $|\mu| = \max(|\mu'| - k, 0)$ read as $|\lambda| \geq |\lambda'| - k$, $|\mu| \geq |\mu'| - k$. One infers that $\vee_\circ(\Lambda, k)$ is a multi-tree. The statement $\# \vee_\circ (\Lambda, k) \lesssim \#\Lambda$ (dependent on k) follows from the locality of the wavelets. $\qquad\square$

4.6. **Best possible rate.** Let the bases $\Theta^\mathscr{U}$, $\Sigma^\mathscr{U}$, $\Theta^\mathscr{P}$, and $\Sigma^\mathscr{P}$ be of orders $d_{\mathscr{U}_t} > 1$, $d_{\mathscr{U}_x} > 1$, $d_{\mathscr{P}_t} > 0$, and $d_{\mathscr{P}_x} > 0$, respectively. Recalling the definition of the approximation class $\mathcal{A}^s$, the largest value $s_{\max}$ of s for which $[\mathbf{u}^\top \, \mathbf{p}^\top]^\top \in \mathcal{A}^s$ can be expected, and for sufficiently smooth $(u, \vec{p}) \in \mathscr{U} \times \vec{\mathscr{P}}$ actually $[\mathbf{u}^\top \, \mathbf{p}^\top]^\top \in \mathcal{A}^s$, is given by

$$s_{\max} = \min \left(d_{\mathscr{U}_t} - 1, \frac{d_{\mathscr{U}_x} - 1}{n}, d_{\mathscr{P}_t}, \frac{d_{\mathscr{P}_x}}{n}\right),$$

see [SS09, Sect. 7]. This holds true assuming the minimum is not attained for simultaneously $d_{\mathscr{P}_t}$ and $\frac{d_{\mathscr{P}_x}}{n}$, in which case this maximal rate is attained up to a log-factor.

Note that when limited by the orders of the spatial wavelets, the value $s_{\max}$ for the approximation rate of the time-dependent problem is *equal* to the approximate rate for the corresponding stationary problem, being the major advantage of tensor product approximation.

The value of $s_{\max}$ has been derived using 'sparse-grid type' multi-trees assuming sufficient smoothness of the solution. Using multi-trees that are adapted to local singularities of the solution, it can be expected that this rate can be attained to a much larger class of functions. Characterizations of approximation classes corresponding to unconstrained tensor-product approximation (no admissibility condition) as anisotropic Besov spaces can be found in [Nit06, SU09]. Based on results on non-tensor product tree approximation ([CDDD01]), we anticipate that the multi-tree constraint on the index sets makes the approximation classes only 'slightly' smaller. See also [DS10] for results on multi-tree tensor-product approximation of the solution of elliptic PDEs.

Although Besov regularity theory for the heat equation can be found in [AG12], to the best of our knowledge corresponding anisotropic Besov regularity results suitable for tensor product approximation are yet not available.

4.7. **Constructing the approximate residual.** Given a fixed $\varsigma > 0$, for all admissible $\Lambda \subset \vee_{\mathscr{U} \times \vec{\mathscr{P}}}$, $[\mathbf{w}^\top \, \mathbf{q}^\top]^\top \in \ell_2(\Lambda)$, and $\varepsilon > 0$, we construct an approximation $\tilde{\mathbf{r}} \in \ell_2(\vee_{\mathscr{U} \times \vec{\mathscr{P}}})$ to $D\mathbf{Q}([\mathbf{w}^\top \, \mathbf{q}^\top]^\top)$ that satisfies the error bound (4.3) from the cost condition Condition 3.4*.

The first statement of the following lemma shows that with $(w, \vec{q}) := (\mathbf{w}^\top \Psi^\mathscr{U}, \mathbf{q}^\top \Psi^{\vec{\mathscr{P}}})$ it holds that $\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} \in \text{span } \Psi^a|_{\vee_a(\Lambda, 0)}$. The second and third statements of this lemma will imply a similar statement for the third 'residual', being $\vec{q} - A\nabla_x w$.

Lemma 4.16. *Let $\Lambda \subset \vee_{\mathscr{U} \times \mathscr{P}}$ be admissible. Then*

$$\vee_a(\Lambda, 0) := \vee_a(\Lambda_{\mathscr{U}}, 0) \cup \cup_{i=1}^n \vee_a\left(\Lambda_{\mathscr{P}_i}, 0\right)$$

is a multi-tree, $\# \vee_a(\Lambda, 0) \lesssim \#\Lambda$ and

$$\operatorname{span}(\frac{\partial}{\partial t} + N)\Psi^{\mathscr{U}}\big|_{\Lambda_{\mathscr{U}}} + \operatorname{span} \nabla_x \cdot \Psi^{\mathscr{P}}\big|_{\Lambda_{\mathscr{P}}} \subset \operatorname{span} \Psi^a\big|_{\vee_a(\Lambda, 0)},$$

$$\operatorname{span} A\nabla_x\Psi^{\mathscr{U}}\big|_{\Lambda_{\mathscr{U}}} + \operatorname{span} \Psi^{\mathscr{P}}\big|_{\Lambda_{\mathscr{P}}} \subset \prod_{i=1}^n \operatorname{span} \Psi^a\big|_{\vee_a(\Lambda, 0)}\mathbf{e}_i,$$

$$\operatorname{span} A^{\top}A\nabla_x\Psi^{\mathscr{U}}\big|_{\Lambda_{\mathscr{U}}} + \operatorname{span} A^{\top}\Psi^{\mathscr{P}}\big|_{\Lambda_{\mathscr{P}}} \subset \prod_{i=1}^n \operatorname{span} \Psi^a\big|_{\vee_a(\Lambda, 0)}\mathbf{e}_i.$$

Proof. The first statement follows by the L_2-orthogonality of the Alpert wavelets and the fact that, using assumption (4.6), the element of the collections $(\frac{\partial}{\partial t} + N)\Psi^{\mathscr{U}}\big|_{\Lambda_{\mathscr{U}}}$ and $\operatorname{span} \nabla_x \cdot \Psi^{\mathscr{P}}\big|_{\Lambda_{\mathscr{P}}}$ are piecewise polynomials. The proofs of the second and third statements are similar. □

The term $\langle \Psi^{\mathscr{U}}(0, \cdot), w(0, \cdot) - h \rangle_{L_2(\Omega)}$ in $D\mathbf{Q}([\mathbf{w}^{\top} \mathbf{q}^{\top}]^{\top})$, resulting from the second 'residual' $w(0, \cdot) - h$, reads as $\mathbf{E}\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, (\mathbf{E}^{\top}\mathbf{w})^{\top} \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}} - h \rangle_{L_2(\Omega)}$, with the $\vee_{\mathscr{U}} \times \Diamond_{\mathscr{U}}$-matrix $\mathbf{E}$ defined by

$$\mathbf{E}_{(\lambda,\mu),\mu'} = \begin{cases} \frac{\theta_\lambda^{\mathscr{U}}(0)\|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)}}{\|\theta_\lambda^{\mathscr{U}} \otimes \sigma_\mu^{\mathscr{U}}\|_{\mathscr{U}}} & \text{if } (\lambda, \mu) \in \vee_{\mathscr{U}}, \mu' = \mu, \\ 0 & \text{if } \mu \neq \mu' \in \Diamond_{\mathscr{U}}. \end{cases}$$

Its transpose $\mathbf{E}^{\top}$ represents the trace mapping at $t = 0$ w.r.t. to the bases $\frac{\Theta_{\mathscr{U}} \times \Sigma_{\mathscr{U}}}{\|\Theta_{\mathscr{U}} \times \Sigma_{\mathscr{U}}\|_{\mathscr{U}}}$ and $\frac{\Sigma_{\mathscr{U}}}{\|\Sigma_{\mathscr{U}}\|_{L_2(\mathscr{U})}}$ for $\mathscr{U}$ and $L_2(\Omega)$, respectively.

Since $\mathbf{w}$ is finitely supported, $\mathbf{E}^{\top}\mathbf{w}$ can be computed exactly in optimal complexity. The function $w(0, \cdot) = (\mathbf{E}^{\top}\mathbf{w})^{\top} \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}$ is piecewise polynomial w.r.t. some tiling $\mathcal{T}$. Given an $\varepsilon > 0$, h will be approximated within tolerance ε by a piecewise polynomial h_ε w.r.t. some tiling $\mathcal{T}(\varepsilon)$, so that $w(0, \cdot) - h_\varepsilon$ is piecewise polynomial w.r.t. the tiling $\mathcal{T} \oplus \mathcal{T}(\varepsilon)$. Now $\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, w(0, \cdot) - h \rangle_{L_2(\Omega)} \in \ell_2(\Diamond_{\mathscr{U}})$ will be approximated by restricting it to a k-neighborhood of $\mathcal{T} \oplus \mathcal{T}(\varepsilon)$ (cf. Def. 4.7). The remaining issue how to approximate the application of $\mathbf{E}$ is dealt with in the following lemma.

Lemma 4.17. *For $k \in \mathbb{N}_0$ define $\mathbf{E}_k$ by*

$$(\mathbf{E}_k)_{(\lambda,\mu),\mu'} = \begin{cases} (\mathbf{E}_{(\lambda,\mu),\mu'} & \text{if } \big||\lambda| - 2|\mu|\big| \leq k, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathbf{E}, \mathbf{E}_k \in \mathcal{L}(\ell_2(\Diamond_{\mathscr{U}}), \ell_2(\vee_{\mathscr{U}}))$, $\|\mathbf{E} - \mathbf{E}_k\| \lesssim 2^{-k/2}$, and $\# \operatorname{supp} \mathbf{E}_k\mathbf{v} \lesssim \# \operatorname{supp} \mathbf{v}$ (dependent on k).

Proof. By the inequalities

$$(4.8) \qquad \|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})} \lesssim 2^{|\lambda|}\|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})} \lesssim \|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})},$$

$$(4.9) \qquad \|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)} \lesssim 2^{-|\mu|}\|\sigma_\mu^{\mathscr{U}}\|_{H^1(\Omega)}, \quad \|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)} \lesssim 2^{|\mu|}\|\sigma_\mu^{\mathscr{U}}\|_{H^{-1}(\Omega)},$$

which will be demonstrated below, $|\theta_\lambda^{\mathscr{U}}(0)|^2 \lesssim \|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})}\|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})}$ by the trace inequality, and, finally, the definition of $\mathscr{U}$, we obtain

$$\Big|\frac{\theta_\lambda^{\mathscr{U}}(0)\|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)}}{\|\theta_\lambda^{\mathscr{U}} \otimes \sigma_\mu^{\mathscr{U}}\|_{\mathscr{U}}}\Big|^2 \lesssim \Big(\frac{\|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})}}{\|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})}}\frac{\|\sigma_\mu^{\mathscr{U}}\|^2_{H^1(\Omega)}}{\|\sigma_\mu^{\mathscr{U}}\|^2_{L_2(\Omega)}} + \frac{\|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})}}{\|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})}}\frac{\|\sigma_\mu^{\mathscr{U}}\|^2_{H^{-1}(\Omega)}}{\|\sigma_\mu^{\mathscr{U}}\|^2_{L_2(\Omega)}}\Big)^{-1}$$

$$\lesssim \big(2^{-|\lambda|}4^{|\mu|} + 2^{|\lambda|}4^{-|\mu|}\big)^{-1} \le 2^{-|\ell|},$$

when $\ell := |\lambda| - 2|\mu| \in \mathbb{Z}$. From this result, one easily infers the statements of the lemma.

From (s_4), we have

$$\|\sigma_\mu^{\mathscr{U}}\|^2_{L_2(\Omega)} \lesssim 2^{-|\mu|}\|\sigma_\mu^{\mathscr{U}}\|_{L_1(\Omega)}|\sigma_\mu^{\mathscr{U}}|_{W^1_\infty(\Omega)}$$

$$\lesssim 2^{-|\mu|}2^{-|\mu|n/2}\|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)}2^{|\mu|n/2}\|\sigma_\mu^{\mathscr{U}}\|_{H^1(\Omega)},$$

which shows first inequality in (4.9). The second one is a consequence of $\|\sigma_\mu^{\mathscr{U}}\|^2_{L_2(\Omega)} \le \|\sigma_\mu^{\mathscr{U}}\|_{H^1(\Omega)}\|\sigma_\mu^{\mathscr{U}}\|_{H^{-1}(\Omega)} \lesssim 2^{|\mu|}\|\sigma_\mu^{\mathscr{U}}\|_{L_2(\Omega)}\|\sigma_\mu^{\mathscr{U}}\|_{H^{-1}(\Omega)}$.

The first inequality in (4.8) is the inverse inequality for polynomials. Thanks to (t_4), the second one in a consequence of $\|\theta_\lambda^{\mathscr{U}}\|^2_{L_2(\mathrm{I})} \lesssim 2^{-|\lambda|}\|\theta_\lambda^{\mathscr{U}}\|_{L_1(\mathrm{I})}|\theta_\lambda^{\mathscr{U}}|_{W^1_\infty(\mathrm{I})} \eqsim 2^{-|\lambda|}\|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})}|\theta_\lambda^{\mathscr{U}}|_{H^1(\mathrm{I})}$. Note that here ($t_4$) has been used also for $v \notin H^1_0(\mathrm{I})$, cf. Remark 4.10. □

We are ready to specify our approximate evaluation of $D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top)$:

Algorithm 4.18 (approximate residual evaluation).

Input: data $g \in L_2(\mathrm{I}; H^{-1}(\Omega))$, $h \in L_2(\Omega)$, *an admissible* $\Lambda \subset \vee_{\mathscr{U}\times\vec{\mathscr{P}}}$, $[\mathbf{w}^\top \mathbf{q}^\top]^\top \in \ell_2(\Lambda)$, $\varepsilon > 0$, *and* $k_2, k_3, k_4, k_5, k_6 \in \mathbb{N}_0$.

Output: $\tilde{\mathbf{r}} := \tilde{\mathbf{r}}_1 + \tilde{\mathbf{r}}_2 + \tilde{\mathbf{r}}_3 \approx D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top)$ *defined as follows:*

(S1) *Find a multi-tree* $\vee_a(\varepsilon) \subset \vee_a$ *such that*

$$\inf_{g_\varepsilon \in \mathrm{span}\, \Psi^a\big|_{\vee_a(\varepsilon)}} \|g - g_\varepsilon\|_{L_2(\mathrm{I};H^{-1}(\Omega))} \le \varepsilon.$$

Find a tiling $\mathcal{T}(\varepsilon) \in \mathcal{O}_\Omega$ *such that*

$$\inf_{h_\varepsilon \in \mathcal{P}_m(\mathcal{T}(\varepsilon))} \|h - h_\varepsilon\|_{L_2(\Omega)} \le \varepsilon.$$

(S2) *With* $\vee_a(\Lambda, \varepsilon) := \vee_a(\Lambda, 0) \cup \vee_a(\varepsilon)$ *and* $(w, \vec{q}) := [\mathbf{w}^\top \mathbf{q}^\top]\Psi$, *approximate*

$$\mathbf{r}_{\frac{1}{2}} := \langle \Psi^{\mathscr{V}_1}, \frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g\rangle_{L_2(\mathrm{I}\times\Omega)}$$

by $\tilde{\mathbf{r}}_{\frac{1}{2}} := \langle \Psi^{\mathscr{V}_1}, \frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g_\varepsilon\rangle_{L_2(\mathrm{I}\times\Omega)}\big|_{\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k_2)}$.

(S3) *With* $\tilde{r}_{\frac{1}{2}} := \tilde{\mathbf{r}}_{\frac{1}{2}}^\top \Psi^{\mathscr{V}_1}$, *approximate*

$$\mathbf{r}_1 := \begin{bmatrix} \langle (\frac{\partial}{\partial t} + N)\Psi^{\mathscr{U}}, \tilde{r}_{\frac{1}{2}}\rangle_{L_2(\mathrm{I}\times\Omega)} \\ \langle \Psi^{\vec{\mathscr{P}}}, \nabla_x \tilde{r}_{\frac{1}{2}}\rangle_{L_2(\mathrm{I}\times\Omega)^n} \end{bmatrix}$$

by $\tilde{\mathbf{r}}_1 := \mathbf{r}_1|_{\vee_{\mathscr{U}}(\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k_2),k_3)\cup \vee_{\vec{\mathscr{P}}}(\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k_2),k_3)}$.

(S4) *With* $\Lambda_\downarrow := \cup_{\{\lambda \in (\Lambda\cap\vee_{\mathscr{U}})_1 :\, \theta_\lambda^{\mathscr{U}}(0)\neq 0\}}(\Lambda \cap \vee_{\mathscr{U}})_2(\lambda) \subset \Diamond_{\mathscr{U}}$, *set* $\mathcal{T}(\Lambda_\downarrow, \varepsilon) := \mathcal{T}(\varepsilon) \oplus \mathcal{T}(\Lambda_\downarrow)$, *and approximate*

$$\mathbf{r}_{\frac{3}{2}} := \Big\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, w(0, \cdot) - h\Big\rangle_{L_2(\Omega)}$$

$$by \ \tilde{\mathbf{r}}_{\frac{3}{2}} := \left\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, w(0,\cdot) - h_\varepsilon \right\rangle_{L_2(\Omega)} |_{\Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k_4)}.$$

(S5) *Approximate* $\mathbf{r}_2 := \begin{bmatrix} \mathbf{E}\tilde{\mathbf{r}}_{\frac{3}{2}} \\ 0 \end{bmatrix}$ *by* $\tilde{\mathbf{r}}_2 := \begin{bmatrix} \mathbf{E}_{k_5}\tilde{\mathbf{r}}_{\frac{3}{2}} \\ 0 \end{bmatrix}.$

(S6) *Approximate*

$$\mathbf{r}_3 := \begin{bmatrix} \langle \nabla_x \Psi^{\mathscr{U}}, A^\top (A\nabla_x w - \vec{q}) \rangle_{L_2(\mathrm{I}\times\Omega)^n} \\ \langle \Psi^{\vec{\mathscr{P}}}, \vec{q} - A\nabla_x w \rangle_{L_2(\mathrm{I}\times\Omega)^n} \end{bmatrix}$$

$$by \ \tilde{\mathbf{r}}_3 = \tilde{\mathbf{r}}_3(k_6) := \mathbf{r}_3|_{\vee_{\mathscr{U}}(\vee_a(\Lambda,0),k_6)\cup\vee_{\vec{\mathscr{P}}}(\vee_a(\Lambda,0),k_6)}.$$

Theorem 4.19. *For* $g \in L_2(\mathrm{I}; H^{-1}(\Omega))$, $h \in L_2(\Omega)$, *let* $s > 0$ *be such that the solution of* $D\mathbf{Q}([\mathbf{u}^\top \mathbf{p}^\top]^\top) = 0$ *satisfies* $[\mathbf{u}^\top \mathbf{p}^\top]^\top \in \mathcal{A}^s$. *Let* $\vee_a(\varepsilon)$ *and* $\mathcal{T}(\varepsilon)$ *from* (S1) *satisfy* $\max(\#\vee_a(\varepsilon), \#\mathcal{T}(\varepsilon)) \lesssim \varepsilon^{-1/s}$. *Then given an admissible* $\Lambda \subset \vee_{\mathscr{U}\times\vec{\mathscr{P}}}$, $[\mathbf{w}^\top \mathbf{q}^\top]^\top \in \ell_2(\Lambda)$, *and* $\varepsilon > 0$, $\tilde{\mathbf{r}}$ *produced by Algorithm 4.18 satisfies*

(4.10)
$$\|D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top) - \tilde{\mathbf{r}}\| \lesssim 2^{-k/2} \big[\|\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g\|_{L_2(\mathrm{I};H^{-1}(\Omega))} +$$
$$\|w(0,\cdot) - h\|_{L_2(\Omega)} + \|\vec{q} - A\nabla_x w\|_{L_2(\mathrm{I};L_2(\Omega)^n)} \big] + \varepsilon,$$

where $k := \min(k_2, k_3, k_4, k_5, k_6)$, *and its computation requires* $\mathcal{O}(\#\Lambda + \varepsilon^{-1/s})$ *operations. So by taking* k *a sufficiently large constant, Assumption 3.4* *is satisfied.*

Remark 4.20. Collections $\vee_a(\varepsilon)$ and $\mathcal{T}(\varepsilon)$ as in the statement of the theorem exist, and so the condition about them concerns their actual construction.

Indeed, when $[\mathbf{u}^\top \mathbf{p}^\top]^\top \in \mathcal{A}^s$, then given an $\varepsilon > 0$ there exists an admissible $\Lambda_\varepsilon \subset \vee_{\mathscr{U}\times\vec{\mathscr{P}}}$ with $\#\Lambda_\varepsilon \lesssim \varepsilon^{-1/s}$ and a $[\mathbf{u}_\varepsilon^\top \mathbf{p}_\varepsilon^\top]^\top \in \ell_2(\Lambda_\varepsilon)$ with, for a constant $C > 0$ determined below, $\|[\mathbf{u}^\top \mathbf{p}^\top]^\top - [\mathbf{u}_\varepsilon^\top \mathbf{p}_\varepsilon^\top]^\top\| \le \frac{\varepsilon}{C}$, and thus with $(u,\vec{p}) := (\mathbf{u}^\top \Psi^{\mathscr{U}}, \mathbf{p}^\top \Psi^{\vec{\mathscr{P}}})$ and $(u_\varepsilon, \vec{p}_\varepsilon) := (\mathbf{u}_\varepsilon^\top \Psi^{\mathscr{U}}, \mathbf{p}_\varepsilon^\top \Psi^{\vec{\mathscr{P}}})$, $\|u - u_\varepsilon\|_{\mathscr{U}} + \|\vec{p} - \vec{p}_\varepsilon\|_{\vec{\mathscr{P}}} \lesssim \varepsilon$. Using that $g = \frac{\partial}{\partial t}u + Nu - \nabla_x \cdot \vec{p}$, we infer that, by selecting a proper C, with $g_\varepsilon := \frac{\partial}{\partial t}u_\varepsilon + Nu_\varepsilon - \nabla_x \cdot \vec{p}_\varepsilon$, it holds that $\|g - g_\varepsilon\|_{L_2(\mathrm{I};H^{-1}(\Omega))} \le \varepsilon$. Since as shown in Lemma 4.16, $\mathrm{span}(\frac{\partial}{\partial t} + N)\Psi^{\mathscr{U}}|_{(\Lambda_\varepsilon)_{\mathscr{U}}} + \mathrm{span}\,\nabla_x \cdot \Psi^{\vec{\mathscr{P}}}|_{(\Lambda_\varepsilon)_{\vec{\mathscr{P}}}} \subset \mathrm{span}\,\Psi^a|_{\vee_a(\Lambda_\varepsilon,0)}$ and $\#\vee_a(\Lambda_\varepsilon,0) \lesssim \#\Lambda_\varepsilon \lesssim \varepsilon^{-1/s}$, we conclude that a suitable $\vee_a(\varepsilon)$ exists.

Similarly, since $h = u(0,\cdot)$, with $(\Lambda_\varepsilon)_\downarrow$ defined similarly as in (S4), by taking $h_\varepsilon := (\mathbf{u}_\varepsilon|_{(\Lambda_\varepsilon)_\downarrow})^\top \Sigma^{\mathscr{U}}$ and possibly adjusting C, it holds that $\|h - h_\varepsilon\|_{L_2(\Omega)} \le \varepsilon$. The collection $(\Lambda_\varepsilon)_\downarrow$ is a tree in $\Diamond_{\mathscr{U}}$, and so $h_\varepsilon \in \mathcal{P}_m(\mathcal{T}(\varepsilon))$ with, thanks to Proposition 4.6, $\#\mathcal{T}(\varepsilon) \lesssim \#(\Lambda_\varepsilon)_\downarrow \lesssim \#\Lambda_\varepsilon \lesssim \varepsilon^{-1/s}$.

Proof of Theorem 4.19. The expression for $D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top)$ given in (4.2), and the definitions of $\mathbf{r}_i$ and $\tilde{\mathbf{r}}_i$, for $i \in \{\frac{1}{2}, 1, \frac{3}{2}, 2, 3\}$, show that

$$D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top) - \tilde{\mathbf{r}} =$$

$$\mathbf{r}_1 - \tilde{\mathbf{r}}_1 + \begin{bmatrix} \langle \frac{\partial}{\partial t}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1} \rangle_{L_2(\Omega)} \\ \langle \Psi^{\vec{\mathscr{P}}}, \nabla_x \Psi^{\mathscr{V}_1} \rangle_{L_2(\Omega)^n} \end{bmatrix} (\mathbf{r}_{\frac{1}{2}} - \tilde{\mathbf{r}}_{\frac{1}{2}}) + \mathbf{r}_2 - \tilde{\mathbf{r}}_2 + \mathbf{E}(\mathbf{r}_{\frac{3}{2}} - \tilde{\mathbf{r}}_{\frac{3}{2}}) + \mathbf{r}_3 - \tilde{\mathbf{r}}_3.$$

The boundedness of $\mathbf{E}$ (cf. Lemma 4.17) and, by the Riesz basis properties of $\Psi^{\mathscr{U}}$, $\Psi^{\mathscr{V}_1}$ and $\Psi^{\vec{\mathscr{P}}}$, that of $\langle \Psi^{\vec{\mathscr{P}}}, \nabla_x \Psi^{\mathscr{V}_1} \rangle_{L_2(\mathrm{I}\times\Omega)^n}$ and $\langle \frac{\partial}{\partial t}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1} \rangle_{L_2(\mathrm{I}\times\Omega)}$, show that

$$\|D\mathbf{Q}([\mathbf{w}^\top \mathbf{q}^\top]^\top) - (\tilde{\mathbf{r}}_1 + \tilde{\mathbf{r}}_2 + \tilde{\mathbf{r}}_3)\| \lesssim \sum_{i\in\{\frac{1}{2},1,\frac{3}{2},2,3\}} \|\mathbf{r}_i - \tilde{\mathbf{r}}_i\|.$$

Below, we will show that all five terms at the right-hand side are bounded by a multiple of the right-hand side of (4.10).

For $g_\varepsilon \in \operatorname{span} \Psi^a|_{\vee_a(\varepsilon)}$, we write

$$\mathbf{r}_{\frac{1}{2}} - \tilde{\mathbf{r}}_{\frac{1}{2}} = \langle \Psi^{\mathscr{V}_1}, \frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g_\varepsilon \rangle_{L_2(\mathrm{I}\times\Omega)}\big|_{\vee_{\mathscr{V}_1}\setminus\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k)} + \langle \Psi^{\mathscr{V}_1}, g_\varepsilon - g \rangle_{L_2(\mathrm{I}\times\Omega)}.$$

From the first statement of Lemma 4.16, we know that $\frac{\partial}{\partial t}w + Nw - \nabla_x \cdot \vec{q} - g_\varepsilon \in \operatorname{span} \Psi^a|_{\vee_a(\Lambda,\varepsilon)}$. An application of the forthcoming Theorem A.3 shows that consequently the norm of the first term is $\lesssim 2^{-k/2}\|\frac{\partial}{\partial t}w + Nw - \nabla_x \cdot \vec{q} - g_\varepsilon\|_{L_2(\mathrm{I};H^{-1}(\Omega))}$. From $\Psi^{\mathscr{V}_1}$ being a Riesz basis for $L_2(\mathrm{I};H_0^1(\Omega))$, it follows that the norm of the second term is $\lesssim \|g - g_\varepsilon\|_{L_2(\mathrm{I};H^{-1}(\Omega))}$. From (S1), we infer that

$$\|\mathbf{r}_{\frac{1}{2}} - \tilde{\mathbf{r}}_{\frac{1}{2}}\| \lesssim 2^{-k/2}\|\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g\|_{L_2(\mathrm{I};H^{-1}(\Omega))} + \varepsilon.$$

Applications of the forthcoming Corollaries A.7 and A.9 show that $\|\mathbf{r}_1 - \tilde{\mathbf{r}}_1\| \lesssim 2^{-k/2}\|\mathbf{r}_{\frac{1}{2}}\|$, whereas $\|\mathbf{r}_{\frac{1}{2}}\| \le \|\mathbf{r}_{\frac{1}{2}} - \tilde{\mathbf{r}}_{\frac{1}{2}}\| + \|\tilde{\mathbf{r}}_{\frac{1}{2}}\|$ and $\|\tilde{\mathbf{r}}_{\frac{1}{2}}\| \lesssim \|\frac{\partial w}{\partial t} + Nw - \nabla_x \cdot \vec{q} - g\|_{L_2(\mathrm{I};H^{-1}(\Omega))}$.

For $h_\varepsilon \in \mathcal{P}_m(\mathcal{T}(\varepsilon))$, we write

$$\mathbf{r}_{\frac{3}{2}} - \tilde{\mathbf{r}}_{\frac{3}{2}} = \Big\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, w(0,\cdot) - h_\varepsilon \Big\rangle_{L_2(\Omega)}\Big|_{\Diamond_{\mathscr{U}}\setminus\Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k)} + \Big\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, h_\varepsilon - h \Big\rangle_{L_2(\Omega)}.$$

Note that $w(0,\cdot) \in \operatorname{span} \Sigma^{\mathscr{U}}|_{\Lambda_\downarrow}$, and that $\Lambda_\downarrow$ is a tree in $\Diamond_{\mathscr{U}}$. Using that $\Sigma^{\mathscr{U}}$ satisfies (s_1)-(s_4), analogously to [RS18a, Prop. A4 (first statement)] one shows that that the norm of the first term is $\lesssim 2^{-k/2}\|w(0,\cdot) - h_\varepsilon\|_{L_2(\Omega)}$. From $\Sigma^{\mathscr{U}}/\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}$ being a Riesz basis it follows that the norm of the second term is $\lesssim \|h - h_\varepsilon\|_{L_2(\Omega)}$. From (S1), we infer that $\|\mathbf{r}_{\frac{3}{2}} - \tilde{\mathbf{r}}_{\frac{3}{2}}\| \lesssim 2^{-k/2}\|w(0,\cdot) - h\|_{L_2(\Omega)} + \varepsilon$.

An application of Lemma 4.17 shows that $\|\mathbf{r}_2 - \tilde{\mathbf{r}}_2\| \lesssim 2^{-k/2}\|\tilde{\mathbf{r}}_{\frac{3}{2}}\|$, whereas $\|\tilde{\mathbf{r}}_{\frac{3}{2}}\| \lesssim \|w(0,\cdot) - h\|_{L_2(\Omega)}$.

From the second and third statements of Lemma 4.16, we know that $\vec{q} - A\nabla_x w$, $A^\top(A\nabla_x w - \vec{q}) \in \prod_{i=1}^n \operatorname{span} \Psi^a|_{\vee_a(\Lambda,0)}\mathbf{e}_i$. Now an application of the forthcoming Corollary A.11 shows that $\|\mathbf{r}_3 - \tilde{\mathbf{r}}_3\| \lesssim 2^{-k/2}\|\vec{q} - A\nabla_x w\|_{L_2(\mathrm{I}\times\Omega)^n}$, which completes the proof of (4.10).

The computation of $\tilde{\mathbf{r}}_{\frac{1}{2}}$, $\tilde{\mathbf{r}}_1$, $\tilde{\mathbf{r}}_{\frac{3}{2}}$, $\tilde{\mathbf{r}}_2$ or $\tilde{\mathbf{r}}_3$ requires a number of operations that is of the order

$$\# \vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k),$$
$$\# \vee_{\mathscr{U}}(\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k),k) + \# \vee_{\mathscr{P}}(\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k),k),$$
$$\# \Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k) + \#\Lambda,$$
$$k\# \Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k),$$
$$\# \vee_{\mathscr{U}}(\vee_a(\Lambda,0),k) + \# \vee_{\mathscr{P}}(\vee_a(\Lambda,0),k),$$

respectively. Each of these numbers can be bounded by a multiple (dependent on k) of $\#\Lambda + \varepsilon^{-1/s}$ which proves the statement about the total complexity of computing $\tilde{\mathbf{r}}$.

The statement about the cost of computing $\tilde{\mathbf{r}}_{\frac{1}{2}}$ follows by expressing $\frac{\partial}{\partial t}w + Nw - \nabla_x \cdot \vec{q} - g_\varepsilon$ in terms of $\Psi^a|_{\vee_a(\Lambda,\varepsilon)}$, and then by applying the statement

about the cost of the evaluation of (4.7). A similar argument applies for the computation of $\tilde{\mathbf{r}}_1$ when $\langle N\Psi^{\mathscr{U}}, \tilde{r}_{\frac{1}{2}}\rangle_{L_2(I\times\Omega)}$ is written as $\langle\Psi^{\mathscr{U}}, \Psi^a\rangle_{L_2(I\times\Omega)}\mathbf{d}$ for some $\mathbf{d} \in \vee_a(\vee_{\mathscr{V}_1}(\vee_a(\Lambda,\varepsilon),k),0)$, as well as for the computation of $\tilde{\mathbf{r}}_3$. The evaluation of $\tilde{\mathbf{r}}_{\frac{3}{2}}$ requires first the computation of $\mathbf{E}^\top\mathbf{w}$, which takes $\mathcal{O}(\#\Lambda)$ operations, then h_ε needs to be subtracted taking $\mathcal{O}(\#\Lambda_\downarrow + \#\mathcal{T}(\varepsilon))$ operations, and finally $\left\langle\frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{L_2(\Omega)}}, w(0,\cdot) - h_\varepsilon\right\rangle_{L_2(\Omega)}|_{\Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k)}$ needs to be evaluated which takes $\mathcal{O}(\#\Diamond_{\mathscr{U}}(\mathcal{T}(\Lambda_\downarrow,\varepsilon),k))$ operations (cf. Remark 4.9). The statement about the cost of the evaluation of $\tilde{\mathbf{r}}_2$ follows directly from the definition of $\mathbf{E}_k$. $\qquad\square$

Finally, recall that in addition to the cost condition Condition 3.4* that has been verified in Theorem 4.19, another condition, Condition 3.3, is required to conclude by Theorem 3.5 that the **awgm** is optimal. This condition requires the determination of an admissible $\tilde{\Lambda} \supset \Lambda$ with essentially quasi-minimal $\#(\tilde{\Lambda}\setminus\Lambda)$ such that $\|\tilde{\mathbf{r}}|_{\tilde{\Lambda}}\| \geq \mu_0\|\tilde{\mathbf{r}}\|$. Unfortunately in our setting of multi-tree approximation, we are not aware of an algorithm that is guaranteed to yield such a $\tilde{\Lambda}$. In our numerical experiments, as a first step we constructed some set $\hat{\Lambda} \supset \Lambda$ with quasi-minimal $\#(\hat{\Lambda}\setminus\Lambda)$ such that $\|\tilde{\mathbf{r}}|_{\hat{\Lambda}}\| \geq \mu_0\|\tilde{\mathbf{r}}\|$ by applying a bucket sort procedure on the entries of $\tilde{\mathbf{r}}$. Secondly, we enlarged $\hat{\Lambda}$ to an admissible set. In experiments we observed that $\#(\tilde{\Lambda}\setminus\hat{\Lambda})$ is at most a small multiple of $\#(\hat{\Lambda}\setminus\Lambda)$ which means that Condition 3.3 is satisfied, but we do not have a proof of this.

5. Numerical results

We consider the *heat equation*, i.e. $A = \mathrm{Id}$ and $N = 0$, on the L-shaped domain $\Omega = (0,1) \setminus [\frac{1}{2}, 1)^2$ and $I = (0,1)$. For our convenience we take $g = 1$, and consider three different initial conditions $h = 0$, $h = 1$, and $h(x,y) = 50x(x-1)(x-\frac{1}{2})y(y-1)(y-\frac{1}{2})$. The resulting solutions will exhibit a spatial singularity caused by the re-entrant corner, as well as, for the last two initial conditions, temporal-spatial singularities caused by the incompatibility of the initial and boundary conditions (most severely for $h = 1$) at the intersection of the bottom and lateral boundary.

We select the spatial wavelet collections Σ^* for $* \in \{\mathscr{U}, \mathscr{V}_1, \mathscr{P}\}$ s.t. σ_λ^* for $|\lambda| = \ell$ is piecewise polynomial w.r.t. the triangulation of Ω indicated in Figure 1, which specifies the collection $\mathcal{O}_\Omega$. We take $\Sigma^{\mathscr{V}_1}$ to be the continuous piecewise linear
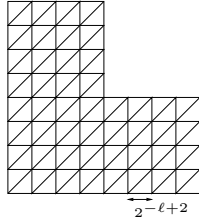

FIGURE 1. Partition of Ω on level $\ell \in \mathbb{N}_0$.

three-point wavelet basis from [Ste98], satisfying homogenous boundary conditions, and normalized such that it is a Riesz basis for $H_0^1(\Omega)$, and for $\Sigma^{\mathscr{P}}$ we select this three-point wavelet basis, now without boundary conditions, and normalized such that it is a Riesz basis for $L_2(\Omega)$. For $\Sigma^{\mathscr{U}}$ we take a continuous piecewise quadratic

wavelet basis that will be described in [RS18b], and that when normalized in $H^1(\Omega)$ or in $H^{-1}(\Omega)$ is a Riesz basis for $H_0^1(\Omega)$ or $H^{-1}(\Omega)$, respectively.

For $\Theta^{\mathscr{V}_1}$ and $\Theta^{\mathscr{P}}$ we take a discontinuous $L_2(I)$-orthonormal piecewise linear wavelet basis, and for $\Theta^{\mathscr{U}}$ the continuous piecewise linear three-point wavelet basis (without boundary conditions).

These collections satisfy all conditions (s_1)-(s_4), $(s_4^{\mathscr{U}})$, and (t_1)-(t_4), and using them we build the tensor product wavelet bases $\Psi^{\mathscr{V}_1}$, $\Psi^{\mathscr{P}}$ and $\Psi^{\mathscr{U}}$ as in (4.5).

It holds that $d_{\mathscr{U}_t} = d_{\mathscr{P}_t} = d_{\mathscr{P}_x} = 2$ and $d_{\mathscr{U}_x} = 3$, meaning that the best possible rate $s_{\max} = 1$.

The auxiliary temporal and spatial Alpert wavelet bases Θ^a and Σ^a were taken of degrees $m = 1$ and $m = 2$, respectively, which are the smallest values such that the inclusions from Lemma 4.16 hold.

Recall that $D\mathbf{Q}$ is affine, so that its (symmetric positive definite) linear part is given by $D^2\mathbf{Q} \colon [\mathbf{w}^\top \, \mathbf{q}^\top]^\top \mapsto D\mathbf{Q}[\mathbf{w}^\top \, \mathbf{q}^\top]^\top - D\mathbf{Q}[0\,0]^\top$. We have approximated the condition number of this system matrix by restricting it to the square block corresponding to all wavelets with indices $\lambda \in \vee_{\mathscr{U}}$, $(\mu, i) \in \vee_{\mathscr{P}} \times \{1, 2\}$ with $|\lambda|, |\mu|$ less than some integer. Even these Galerkin matrices cannot be evaluated exactly because they are of the form $\mathbf{A}_1^\top \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3$ where the rows of $\mathbf{A}_1$ run over the infinite index set $\vee_{\mathscr{V}_1}$. Similarly to Step (S2) in Algorithm 4.18 we made an approximation by omitting all rows corresponding to indices $\gamma \in \vee_{\mathscr{V}_1}$ for which $|\gamma|$ exceeds any value of $|\lambda|$ or $|\mu|$ by more than a constant k that was taken sufficiently large so that it hardly affected the computed condition numbers. These numbers, illustrated in Figure 2 indicate that the condition number of the infinite system matrix can be expected to be of the order of 700.
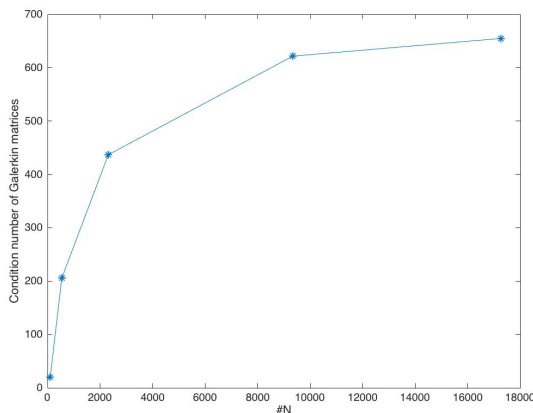

FIGURE 2. Approximate condition numbers of the Galerkin matrices vs. their dimension.

In [RS18a] for an analogous FOSLS formulation of the corresponding stationary operator, i.e. the Laplace operator, and with the same spatial wavelets (and thus without temporal wavelets), we found a condition number of the order of 550.

We applied the **awgm** given in Algorithm 3.1. In step (R) of this algorithm, instead of performing a loop we simply computed the approximated residual by one application of Algorithm 4.18. In step (B) we applied bulk chasing as explained in the last paragraph of the previous Section 4 with parameter $\mu = 0.5$. The Galerkin

matrices in step (G) were approximated using Algorithm 4.18, and approximately solved with parameter $\gamma = 0.2$ using CG iteration as explained in the last paragraph of Section 3.

The parameters k_2, k_3 and k_6 in Algorithm 4.18 were chosen to be equal to 1. Furthermore, (k_4, k_5) was taken to be equal to $(1, 3)$, $(4, 7)$ or $(1, 5)$ for $h = 0$, $h = 1$, and $h(x, y) = 50x(x - 1)(x - \frac{1}{2})y(y - 1)(y - \frac{1}{2})$, respectively. Since g, and in all three cases, h are global polynomials, ε in (S1) equals 0, where $\vee_a(0)$ corresponds to the Cartesian product of the indices corresponding to the temporal and spatial 'scaling functions', and $\mathcal{T}(0)$ is the initial triangulation of Ω. For details about the implementation we refer to [Rek18].

In the left pictures in Figures 3-5, for right-hand side $g = 1$, and initial conditions $h = 0$, $h = 1$, and $h(x, y) = 50x(x-1)(x-\frac{1}{2})y(y-1)(y-\frac{1}{2})$, respectively, the ℓ_2-norm of the (approximate) residual is given vs. the number of wavelets from the basis for $\mathscr{U} \times \vec{\mathscr{P}} = L_2(\mathrm{I}; H_0^1(\Omega)) \cap H^1(\mathrm{I}; H^{-1}(\Omega)) \times L_2(\mathrm{I}; L_2(\Omega)^2)$. This norm is equivalent
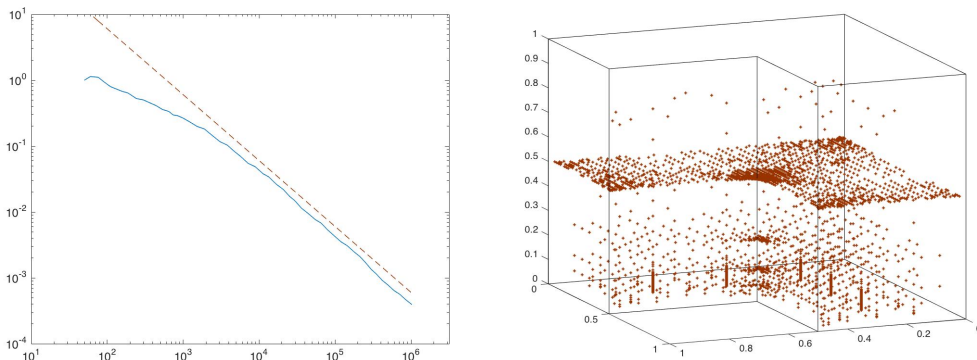

FIGURE 3. Norm residual vs. number of wavelets (dashed line has slope -1), and centers supports of the selected 4500 wavelets for $h = 0$ and $g = 1$.

to the $\mathscr{U} \times \vec{\mathscr{P}}$-norm in the error of the approximation to $(u, \vec{p}) = (u, \nabla_x u)$. In the right pictures one finds the centers of the supports of the tensor product wavelets that were selected by the adaptive method.

For $h = 0$ and $h(x, y) = 50x(x - 1)(x - \frac{1}{2})y(y - 1)(y - \frac{1}{2})$, one observes that the **awgm** converges with the best possible rate $s = 1$. Moreover, thanks to the tensor product approximation not only the rate but also in an absolute sense the results are rather close to the results we found in [Rek18, Figure 4] for the corresponding stationary problem with errors in $(u, \nabla_x u)$ measured in $H_0^1(\Omega) \times L_2(\Omega)^2$. This means that one obtains the additional time dimension nearly for free.

For $h = 1$, the observed rate indicates that the exact solution is only in $\mathcal{A}^{\frac{1}{2}}$. This reduction in the best approximation rate can be understood as follows: With this initial condition, the solution u is discontinuous at the full intersection of the lateral boundary and the bottom of the space-time space cylinder, inducing strong refinements near this intersection, as illustrated in the right picture of Figure 4. Although the solution is smooth in the direction tangential to this intersection, since
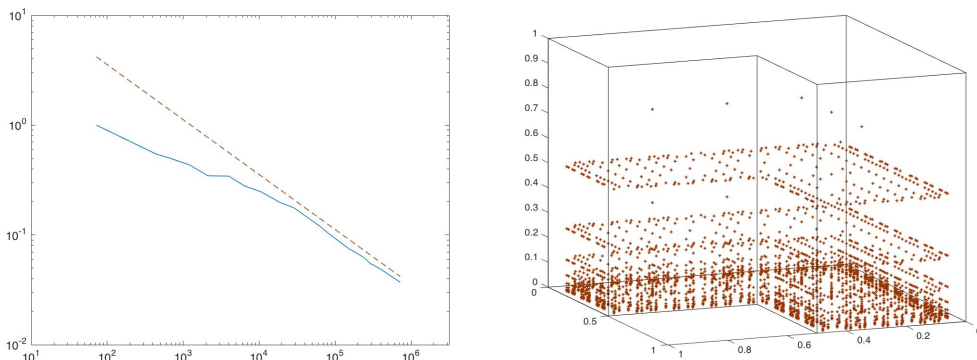
FIGURE 4. Norm residual vs. number of wavelets (dashed line has slope $-\frac{1}{2}$), and centers supports of the selected 4131 wavelets for $h = 1$ and $g = 1$.
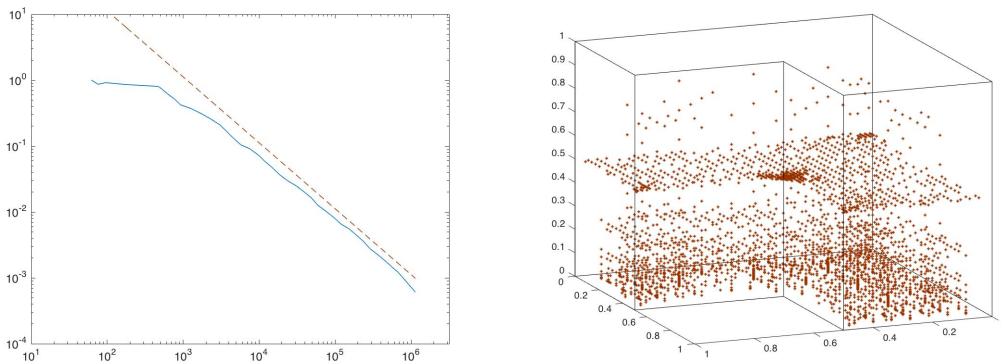

FIGURE 5. Norm residual vs. number of wavelets (dashed line has slope -1), and centers supports of the selected 4345 wavelets for $h(x, y) = 50x(x - 1)(x - \frac{1}{2})y(y - 1)(y - \frac{1}{2})$ and $g = 1$.

the spatial wavelets are isotropic the method cannot benefit from this smoothness causing the reduced best approximation rate.

We expect that if we had applied a spatial piecewise tensor product wavelet basis as constructed in [CDFS13], this reduction would not have occurred, and also for this problem we would have obtained rates as if we would solve a one-dimensional problem. We have however chosen for the current isotropic spatial wavelets because their (relatively) easy construction, and because they apply to any polygon. Moreover, parabolic problems are usually studied assuming that the data satisfies the lowest order compatibility condition of a vanishing initial condition at the homogenous Dirichlet boundary.

Note that the initial condition $h(x, y) = 50x(x - 1)(x - \frac{1}{2})y(y - 1)(y - \frac{1}{2})$ does vanish at $\partial\Omega$, but does not satisfy the next compatibility condition of $-\Delta h = g(= 1)$

at $\partial\Omega$. Nevertheless, it seems to give rise to the best possible rate allowed by the polynomial orders that were applied.

6. CONCLUSION

In this paper an optimal adaptive wavelet solver has been developed for solving a simultaneously space-time FOSLS formulation of parabolic PDEs. Thanks to the use of tensor products of wavelets in space and time the whole time evolution can be solved at a complexity of solving the corresponding stationary problem, which has been illustrated by numerical results.

A theoretical issue that has not yet been satisfactorily solved is that of bulk chasing under a multi-tree constraint (cf. last paragraph of Sect. 4). It may require a generalisation to multi-trees of the tree approximation routines given in [BD04].

Other than in our preceding paper [RS18a] dealing with stationary PDEs and non-tensor product approximation, in order to construct an approximate residual evaluation of linear complexity we had to restrict ourselves to linear PDOs. It would be interesting to circumvent this restriction.

In [SS17], we constructed well-posed simultaneously space-time saddle-point formulations of instationary Stokes and Navier-Stokes equations. Using the approach from [RS18a], these formulations can be recast as well-posed FOSLS so that, modulo the treatment of the nonlinear term in NSE, the adaptive scheme from the current work applies.

Adaptive finite element (afem) schemes usually have better quantitative properties than adaptive wavelet schemes. To the best of our knowledge, for simultaneously space-time variational formulations, currently there are no afem schemes available that are proven to converge, let alone to be optimal or to give rates as for stationary problems. Our FOSLS formulation might give an opening towards such results because it gives a well-posed symmetric positive definite problem.

APPENDIX A. DECAY ESTIMATES

In this appendix we prove the technical results Theorem A.3, Corollaries A.7, A.9, and A.11 that were used in the proof of Theorem 4.19.

The following lemma is an application of Schur's lemma that is often used to bound the spectral norm of a matrix whose the row and column indices run over index sets of multi-level bases.

Lemma A.1. *For index sets J, J', let $|\cdot| \colon J \cup J' \to \mathbb{N}_0$, and let $M := [m_{\lambda',\lambda}]_{(\lambda',\lambda)\in J'\times J}$ be such that for some $\xi \geq 0$, $\varrho > 0$,*

$$\#\{\lambda' \colon m_{\lambda',\lambda} \neq 0, |\lambda'| = |\lambda| + k\} \lesssim 2^{\xi k} \quad (\lambda \in J, k \in \mathbb{N}_0),$$
$$\#\{\lambda \colon m_{\lambda',\lambda} \neq 0, |\lambda'| = |\lambda| + k\} \lesssim 1 \quad (\lambda' \in J', k \in \mathbb{N}_0),$$

and

$$|m_{\lambda',\lambda}| \lesssim 2^{(|\lambda|-|\lambda'|)(\varrho+\frac{\xi}{2})} \quad (|\lambda'| \geq |\lambda|).$$

Then

$$\||M\big|_{\{(\lambda',\lambda)\colon |\lambda'|>|\lambda|+k\}}\|| \lesssim 2^{-\varrho k},$$

where $(M|_{\{(\lambda',\lambda)\colon |\lambda'|>|\lambda|+k\}})_{\lambda',\lambda} := \begin{cases} m_{\lambda',\lambda} & \text{when } |\lambda'| > |\lambda| + k \\ 0 & \text{otherwise} \end{cases}$, and $\|\cdot\|$ denotes the matrix spectral norm, i.e., here the norm on $\mathcal{L}(\ell_2(J), \ell_2(J'))$. The absolute value refers to taking entry-wise absolute value. (Similar notations will be used at other occasions.)

Proof. With $I_{\ell',\ell} := [|m_{\lambda',\lambda}|]_{\{(\lambda',\lambda)\colon |\lambda'|=\ell',|\lambda|=\ell\}}$, we have

$$\||M|_{\{(\lambda',\lambda)\colon |\lambda'|>|\lambda|+k\}}\|\|^2 \lesssim \max_{\ell'} \sum_{\ell<\ell'-k} \|I_{\ell',\ell}\| \times \max_{\ell} \sum_{\ell'>\ell+k} \|I_{\ell',\ell}\|,$$

where ℓ, ℓ' run over $\mathbb{N}_0$.

The number of non-zero entries in each column or row of $I_{\ell',\ell}$ is $\lesssim 2^{\xi(\ell'-\ell)}$ or $\lesssim 1$, respectively. Using $\|\cdot\|^2 \leq \|\cdot\|_1 \|\cdot\|_\infty$, we infer that $\|I_{\ell',\ell}\|^2 \lesssim 2^{\xi(\ell'-\ell)} \cdot 2^{(\frac{\xi}{2}+\varrho)(\ell-\ell')} \cdot 1 \cdot 2^{(\frac{\xi}{2}+\varrho)(\ell-\ell')} = 4^{\varrho(\ell-\ell')}$. $\qquad\square$

The next lemma concerns near-sparsity of a generalized mass matrix corresponding to two temporal wavelet bases.

Lemma A.2. *For $k \in \mathbb{N}_0$, $\Theta^*, \Theta^\circ \in \{\Theta^{\mathscr{V}_1}, \Theta^{\mathscr{P}}, \Theta^a, \Theta^{\mathscr{U}}/\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}\}$ we have*

$$\||\langle\Theta^*,\Theta^\circ\rangle_{L_2(\mathrm{I})}|_{\{(\lambda',\lambda)\colon |\lambda'|>|\lambda|+k\}}\|\| \lesssim 2^{-k/2}.$$

Proof. Using that Θ^* satisfies (t_1)–(t_4), being the counterparts of (s_1)-(s_4) for the spatial wavelets, we split the matrix into $B^r + B^s$, where B^r contains all its entries $\langle\theta_{\lambda'}^*,\theta_\lambda^\circ\rangle_{L_2(\mathrm{I})}$ for which $\operatorname{supp}\theta_{\lambda'}^*$ is contained in ω for some $\omega \in \mathcal{O}_{\mathrm{I}}$ with $|\omega| = |\lambda|$ (the 'regular' entries), and where B^s contains the remaining ('singular') entries.

The number of non-zero entries with $|\lambda'| = \ell'$ and $|\lambda| = \ell$ in each column or row of B^r is $\lesssim 2^{\ell'-\ell}$ or $\lesssim 1$, respectively. Thanks to (t_4), for each of these entries we have $|\langle\theta_{\lambda'}^*,\theta_\lambda^\circ\rangle_{L_2(\mathrm{I})}| \leq \|\theta_{\lambda'}^*\|_{L_1(\mathrm{I})} 2^{-\ell'}|\theta_\lambda^\circ|_{W_\infty^1(\operatorname{supp}\theta_{\lambda'}^*)} \lesssim 2^{3(\ell-\ell')/2}$. An application of Lemma A.1 with $\xi = \varrho = 1$ shows that $\||B^r\|\| \lesssim 2^{-k}$.

The number of non-zero entries with $|\lambda'| = \ell'$ and $|\lambda| = \ell$ in each column or row of B^s is $\lesssim 1$. For each of these entries, we have $|\langle\theta_{\lambda'}^*,\theta_\lambda^\circ\rangle_{L_2(\mathrm{I})}| \leq \|\theta_{\lambda'}^*\|_{L_1(\mathrm{I})}\|\theta_\lambda^\circ\|_{L_\infty(\mathrm{I})} \lesssim 2^{(\ell-\ell')/2}$. An application of Lemma A.1 with $\xi = 0$, $\varrho = \frac{1}{2}$ shows that $\||B^s\|\| \lesssim 2^{-k/2}$. $\qquad\square$

The following theorem provides the main ingredient for bounding $\|\mathbf{r}_{\frac{1}{2}} - \tilde{\mathbf{r}}_{\frac{1}{2}}\|$.

Theorem A.3. *Let $\Lambda^a \subset \vee_a$ be a multi-tree, and $r \in \operatorname{span}\Psi^a|_{\Lambda^a}$ For $k \in \mathbb{N}_0$, it holds that*

$$(A.1)\qquad \|\langle\Psi^{\mathscr{V}_1}, r\rangle_{L_2(\mathrm{I}\times\Omega)}|_{\vee_{\mathscr{V}_1}\setminus\vee_{\mathscr{V}_1}(\Lambda^a,k)}\| \lesssim 2^{-k/2}\|r\|_{L_2(\mathrm{I};H^{-1}(\Omega))}.$$

Proof. We write $r = \sum_{(\lambda,\mu)\in\Lambda^a} r_{\lambda\mu}\theta_\lambda^a \otimes \sigma_\mu^a$, and let

$$\delta_\lambda(\mu') := \begin{cases} 0 & |\mu'| \leq \max\{|\mu|\colon \mu \in \Lambda_2^a(\lambda), \operatorname{meas}(\mathcal{S}(\sigma_{\mu'}^{\mathscr{V}_1}) \cap \operatorname{supp}\sigma_\mu^a) > 0\} + k \\ 1 & \text{elsewhere} \end{cases},$$

Writing $\Lambda^{\mathcal{V}_1} := \vee_{\mathcal{V}_1}(\Lambda^a, k)$, from $\frac{1}{2}(a+b)^2 \leq (a^2+b^2)$, we have that

$$
\begin{aligned}
&\frac{1}{2}\|\langle \Psi^{\mathcal{V}_1}, r\rangle_{L_2(\mathrm{I}\times\Omega)}|_{\vee_{\mathcal{V}_1}\setminus\Lambda^{\mathcal{V}_1}}\|^2 \leq \\
\text{(A.2)}\quad & \sum_{\mu'\in\Diamond_{\mathcal{V}_1}} \sum_{\lambda'\in\lhd_{\mathcal{V}_1}\setminus\Lambda_1^{\mathcal{V}_1}(\mu')} \big| \sum_{|\lambda'|>|\lambda|+k} \langle\theta_{\lambda'}^{\mathcal{V}_1},\theta_\lambda^a\rangle_{L_2(\mathrm{I})}\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2 + \\
& \sum_{\mu'\in\Diamond_{\mathcal{V}_1}} \sum_{\lambda'\in\lhd_{\mathcal{V}_1}\setminus\Lambda_1^{\mathcal{V}_1}(\mu')} \big| \sum_{|\lambda|\geq|\lambda'|-k} \langle\theta_{\lambda'}^{\mathcal{V}_1},\theta_\lambda^a\rangle_{L_2(\mathrm{I})}\delta_\lambda(\mu')\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2
\end{aligned}
$$

Here we could insert the factor $\delta_\lambda(\mu')$ in the second sum because of the following reason: Let $(\lambda',\mu')\in\vee_{\mathcal{V}_1}\setminus\Lambda^{\mathcal{V}_1}$ and $\lambda\in\Lambda_1^a$ with $|\lambda|\geq|\lambda'|-k$. If $\mathrm{meas}(\mathcal{S}(\theta_{\lambda'}^{\mathcal{V}_1})\cap \mathrm{supp}\,\theta_\lambda^a)=0$, then the value of $\delta_\lambda(\mu')$ is irrelevant. If $\mathrm{meas}(\mathcal{S}(\theta_{\lambda'}^{\mathcal{V}_1})\cap\mathrm{supp}\,\theta_\lambda^a)>0$, then the definition of $\Lambda^{\mathcal{V}_1}=\vee_{\mathcal{V}_1}(\Lambda^a,k)$ shows that $|\mu'|>|\mu|+k$ for all $\mu\in\Lambda_2^a(\lambda)$ with $\mathrm{meas}(\mathcal{S}(\sigma_{\mu'}^{\mathcal{V}_1})\cap\mathrm{supp}\,\sigma_\mu^a)>0$, meaning that $\delta_\lambda(\mu')=1$.

Using Lemma A.2 for $(\Theta^*,\Theta^\circ)=(\Theta^{\mathcal{V}_1},\Theta^a)$, the first sum can be bounded on a multiple of

$$
\begin{aligned}
&\sum_{\mu'\in\Diamond_{\mathcal{V}_1}} 2^{-k} \sum_{\lambda\in\lhd_a} \big|\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2 \\
&= 2^{-k}\sum_{\lambda\in\lhd_a}\sum_{\mu'\in\Diamond_{\mathcal{V}_1}} \big|\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2 \eqsim 2^{-k}\sum_{\lambda\in\lhd_a}\|\sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\|_{H^{-1}(\Omega)}^2 \\
&\eqsim 2^{-k}\|\sum_{\lambda\in\lhd_a}\theta_\lambda^a\otimes\sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\|_{L_2(\mathrm{I};H^{-1}(\Omega))}^2 = 2^{-k}\|r\|_{L_2(\mathrm{I};H^{-1}(\Omega))}^2,
\end{aligned}
$$

where we used that $\Sigma^{\mathcal{V}_1}$ is a Riesz basis for $H_0^1(\Omega)$, and that Θ^a is a Riesz basis for $L_2(\mathrm{I})$.

To bound the second sum, recall that for $\mu\in\Diamond_a$, it holds that $\mathrm{supp}\,\sigma_\mu^a=\omega_\mu$ for some $\omega_\mu\in\mathcal{O}_\Omega$ with $|\omega_\mu|=\max(|\mu|-1,0)$. Define the tiling $\mathcal{T}(\lambda)\in\mathcal{O}_\Omega$ as the union, over the leaves μ of the tree $\Lambda_2^a(\lambda)$, of the children of ω_μ when $|\mu|>0$, or of ω_μ itself when $|\mu|=0$. Then $\mathrm{span}\{\sigma_\mu^a\colon \mu\in\Lambda_2^a(\lambda)\}=\mathcal{P}_m(\mathcal{T}(\lambda))$, and $\{\mu'\in\Diamond_{\mathcal{V}_1}\colon \delta_\lambda(\mu')=1\}=\Diamond_{\mathcal{V}_1}\setminus\Diamond_{\mathcal{V}_1}(\mathcal{T}(\lambda),k)$, cf. Definition 4.7.

Since Θ^a and $\Theta^{\mathcal{V}_1}$ are Riesz bases for $L_2(\mathrm{I})$, and so $\langle\Theta^{\mathcal{V}_1},\Theta^a\rangle_{L_2(\mathrm{I})}\in\mathcal{L}(\ell_2(\lhd_a),\ell_2(\lhd_{\mathcal{V}_1}))$, invoking [RS18a, Prop. A.1] using that $\Sigma^{\mathcal{V}_1}$ satisfies (s_1)–(s_4), the second sum can be bounded on a multiple of

$$
\begin{aligned}
&\sum_{\mu'\in\Diamond_{\mathcal{V}_1}}\sum_{\lambda\in\lhd_a} \big|\delta_\lambda(\mu')\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2 \\
&= \sum_{\lambda\in\lhd_a}\sum_{\mu'\in\Diamond_{\mathcal{V}_1}} \big|\delta_\lambda(\mu')\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}\big|^2 \lesssim \sum_{\lambda\in\lhd_a} 4^{-k}\|\sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\|_{H^{-1}(\Omega)}^2 \\
&\eqsim 4^{-k}\|\sum_{\lambda\in\lhd_a}\theta_\lambda^a\otimes\sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\|_{L_2(\mathrm{I};H^{-1}(\Omega))}^2 = 4^{-k}\|r\|_{L_2(\mathrm{I};H^{-1}(\Omega))}^2,
\end{aligned}
$$

where we used that Θ^a is a Riesz basis for $L_2(\mathrm{I})$. $\square$

If Θ^a was a Riesz basis for $H^{-1}(\Omega)$, then in the proof of Theorem A.3 it would have been natural to write $\langle\sigma_{\mu'}^{\mathcal{V}_1}, \sum_{\mu\in\Lambda_2^a(\lambda)} r_{\lambda\mu}\sigma_\mu^a\rangle_{L_2(\Omega)}$ as $\langle\Sigma^{\mathcal{V}_1},\Sigma^a\rangle_{L_2(\Omega)}[r_{\lambda\mu}]_{\mu\in\Lambda_2^a(\lambda)}$. In this case the approach of the insertion of the factor $\delta_\lambda(\mu')$ would have given the

bound

$$\tfrac{1}{2}\sqrt{2}\,\big\|\langle\Psi^{\mathscr{V}_1},\Psi^a\rangle_{L_2(\mathrm{I}\times\Omega)}\big|_{\vee_{\mathscr{V}_1}\setminus\Lambda^{\mathscr{V}_1}(\Lambda^a,k)\times\Lambda^a}\big\|\leq$$

$$\big\|\langle\Theta^{\mathscr{V}_1},\Theta^a\rangle_{L_2(\mathrm{I})}\big|_{\{(\lambda',\lambda):\,|\lambda'|>|\lambda|+k\}}\big\|\big\|\langle\Sigma^{\mathscr{V}_1},\Sigma^a\rangle_{L_2(\Omega)}\big\|+$$

$$\big\|\langle\Theta^{\mathscr{V}_1},\Theta^a\rangle_{L_2(\mathrm{I})}\big\|\big\|\langle\Sigma^{\mathscr{V}_1},\Sigma^a\rangle_{L_2(\Omega)}\big|_{\{(\mu',\mu):\,|\mu'|>|\mu|+k\}}\big\|.$$

Although in the current setting where $\langle\Sigma^{\mathscr{V}_1},\Sigma^a\rangle_{L_2(\Omega)}\notin\mathcal{L}(\ell_2(\Diamond_a),\ell_2(\Diamond_{\mathscr{V}_1}))$, this estimate makes not much sense, for other collections this result, formulated in the next proposition, is going to be useful.

Proposition A.4. *For* $*,\circ\in\{\mathscr{U},\mathscr{V}_1,\mathscr{P},a\}$*, let* $M_{\lhd}:=[m_{\lambda',\lambda}^{\lhd}]_{(\lambda',\lambda)}\in\mathcal{L}(\ell_2(\lhd_*),\ell_2(\lhd_\circ))$*,* $M_{\Diamond}:=[m_{\mu',\mu}^{\Diamond}]_{(\mu',\mu)}\in\mathcal{L}(\ell_2(\Diamond_*),\ell_2(\Diamond_\circ))$*, where* $m_{\lambda',\lambda}^{\lhd}=0$ *when* $\mathrm{meas}(\mathcal{S}(\theta_{\lambda'}^\circ)\cap\mathrm{supp}\,\theta_\lambda^*)=0$*, and* $m_{\mu',\mu}^{\Diamond}=0$ *when* $\mathrm{meas}(\mathcal{S}(\sigma_{\mu'}^\circ)\cap\mathrm{supp}\,\sigma_\mu^*)=0$*. Then for a multi-tree* $\Lambda^*\subset\vee_*$*, and* $k\in\mathbb{N}_0$*, it holds that*

$$\tfrac{1}{2}\sqrt{2}\,\|M_{\lhd}\otimes M_{\Diamond}\big|_{(\vee_\circ\setminus\Lambda^\circ(\Lambda^*,k))\times\Lambda^*}\|\leq$$

$$\|M_{\lhd}\big|_{\{(\lambda',\lambda):\,|\lambda'|>|\lambda|+k\}}\|\|M_{\Diamond}\|+\|M_{\lhd}\|\|M_{\Diamond}\big|_{\{(\mu',\mu):\,|\mu'|>|\mu|+k\}}\|.$$

The remaining of this appendix will consist of various applications of Proposition A.4 for which in several lemmas we estimate norms of type $\|M_{\lhd}\big|_{\{(\lambda',\lambda):\,|\lambda'|>|\lambda|+k\}}\|$ or $\|M_{\lhd}\big|_{\{(\lambda',\lambda):\,|\lambda'|>|\lambda|+k\}}\|$. The next lemma deals with the first task.

Lemma A.5. *For* $k\in\mathbb{N}_0$*, it holds that*

$$\left\|\left|\left\langle\frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^{-1}(\Omega)}},\Sigma^{\mathscr{V}_1}\right\rangle_{L_2(\Omega)}\right|_{\{(\mu',\mu):\,|\mu'|>|\mu|+k\}}\right\|\lesssim 2^{-k/2},$$

$$\left\|\left|\left\langle\frac{\frac{\partial}{\partial x_i}\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}},\Sigma^a\right\rangle_{L_2(\Omega)}\right|_{\{(\mu',\mu):\,|\mu'|>|\mu|+k\}}\right\|\lesssim 2^{-k/2}.$$

Proof. For proving the first inequality, we split the matrix into B^r+B^s, where B^r contains all its entries $\left\langle\frac{\sigma_{\mu'}^{\mathscr{U}}}{\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}},\sigma_\mu^{\mathscr{V}_1}\right\rangle_{L_2(\Omega)}$ for which $\mathrm{supp}\,\sigma_{\mu'}^{\mathscr{U}}$ is contained in ω for some $\omega\in\mathcal{O}_\Omega$ with $|\omega|=|\mu|$ (the 'regular' entries), and where B^s contains the remaining ('singular') entries.

Thanks to $(s_4^{\mathscr{U}})$, for the regular entries we can estimate

$$|\langle\sigma_{\mu'}^{\mathscr{U}},\sigma_\mu^{\mathscr{V}_1}\rangle_{L_2(\Omega)}|\lesssim\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_1(\Omega)}4^{-|\mu'|}|\sigma_\mu^{\mathscr{V}_1}|_{W_\infty^2(\mathrm{supp}\,\sigma_{\mu'}^{\mathscr{U}})}$$

$$\lesssim 4^{-|\mu'|}2^{-|\mu'|\frac{n}{2}}\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_2(\Omega)}2^{|\mu|}2^{|\mu|\frac{n}{2}}\|\sigma_\mu^{\mathscr{V}_1}\|_{H^1(\Omega)}$$

$$\eqsim 2^{(|\mu|-|\mu'|)(1+\frac{n}{2})}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}$$

where we used $\|\sigma_\mu^{\mathscr{V}_1}\|_{H^1(\Omega)}\eqsim 1$, and $\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_2(\Omega)}^2\leq\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}\lesssim 2^{|\mu'|}\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_2(\Omega)}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}$. An application of Lemma A.1 with $\xi=n$ and $\varrho=1$ shows that $\||B^r\||\lesssim 2^{-k}$.

Since the wavelets $\sigma_\mu^{\mathscr{V}_1}$ are piecewise polynomial functions in $H^1(\Omega)$, they are contained in $W_\infty^1(\Omega)$. Using (s_4), for the remaining singular entries we estimate

$$|\langle\sigma_{\mu'}^{\mathscr{U}},\sigma_\mu^{\mathscr{V}_1}\rangle_{L_2(\Omega)}|\lesssim\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_1(\Omega)}2^{-|\mu'|}|\sigma_\mu^{\mathscr{V}_1}|_{W_\infty^1(\mathrm{supp}\,\sigma_{\mu'}^{\mathscr{U}})}$$

$$\eqsim 2^{(|\mu|-|\mu'|)\frac{n}{2}}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}$$

again by $\|\sigma_\mu^{\mathscr{V}_1}\|_{H^1(\Omega)} \eqsim 1$, and $\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_2(\Omega)} \lesssim 2^{|\mu'|}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^{-1}(\Omega)}$ (cf. (4.9)). An application of Lemma A.1 with $\xi = n-1$ and $\varrho = 1/2$ shows that $\||B^s|\| \lesssim 2^{-k/2}$.

Moving to the second inequality, we split the matrix into $B^r + B^s$, where B^r contains all its entries $\left\langle \frac{\frac{\partial}{\partial x_i}\sigma_{\mu'}^{\mathscr{U}}}{\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}}, \sigma_\mu^a \right\rangle_{L_2(\Omega)}$ for which $\operatorname{supp}\sigma_{\mu'}^{\mathscr{U}}$ is contained in $\omega\cap\Omega$ for some $\omega \in \mathcal{O}_\Omega$ with $|\omega| = |\mu|$ (the 'regular' entries), and where B^s contains the remaining ('singular') entries.

Thanks to (s_4), for the regular entries we can estimate

$$|\langle\frac{\partial}{\partial x_i}\sigma_{\mu'}^{\mathscr{U}}, \sigma_\mu^a\rangle_{L_2(\Omega)}| = |\langle\sigma_{\mu'}^{\mathscr{U}}, \frac{\partial}{\partial x_i}\sigma_\mu^a\rangle_{L_2(\Omega)}| \lesssim \|\sigma_{\mu'}^{\mathscr{U}}\|_{L_1(\Omega)}2^{-|\mu'|}|\sigma_\mu^a|_{W_\infty^2(\operatorname{supp}\sigma_{\mu'}^{\mathscr{U}})}$$

$$\lesssim 4^{-|\mu'|}2^{-|\mu'|\frac{n}{2}}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}4^{|\mu|}2^{|\mu|\frac{n}{2}}\|\sigma_\mu^a\|_{L_2(\Omega)} \eqsim 2^{(|\mu|-|\mu'|)(2+\frac{n}{2})}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)},$$

where we used that $\|\sigma_{\mu'}^{\mathscr{U}}\|_{L_2(\Omega)} \lesssim 2^{-|\mu'|}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}$ ((4.9)). An application of Lemma A.1 with $\xi = n$ and $\varrho = 2$ shows that $\||B^r|\| \lesssim 4^{-k}$.

For the remaining singular entries we estimate

$$|\langle\frac{\partial}{\partial x_i}\sigma_{\mu'}^{\mathscr{U}}, \sigma_\mu^a\rangle_{L_2(\Omega)}| \lesssim \|\sigma_{\mu'}^{\mathscr{U}}\|_{W_1^1(\Omega)}\|\sigma_\mu^a\|_{L_\infty(\operatorname{supp}\sigma_{\mu'}^{\mathscr{U}})}$$

$$\lesssim 2^{(|\mu|-|\mu'|)\frac{n}{2}}\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}\|\sigma_\mu^a\|_{L_2(\operatorname{supp}\sigma_{\mu'}^{\mathscr{U}})}.$$

An application of Lemma A.1 with $\xi = n-1$ and $\varrho = 1/2$ shows that $\||B^s|\| \lesssim 2^{-k/2}$. $\square$

Lemma A.6. *For $k \in \mathbb{N}_0$, it holds that*

$$\left\|\left|\left\langle\frac{(\Theta^{\mathscr{U}})'}{\|(\Theta^{\mathscr{U}})'\|_{L_2(\mathrm{I})}}, \Theta^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I})}\right|_{\{(\lambda',\lambda)\colon |\lambda'|>|\lambda|+k\}}\right|\right\| \lesssim 2^{-k/2}.$$

Proof. We split the matrix into $B^r + B^s$, where B^r contains all ('regular') entries $\left\langle\frac{(\theta_{\lambda'}^{\mathscr{U}})'}{\|(\theta_{\lambda'}^{\mathscr{U}})'\|_{L_2(\mathrm{I})}}, \theta_\lambda^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I})}$ for which $\operatorname{supp}\theta_{\lambda'}^{\mathscr{U}}$ is contained in $\omega \cap I$ for some $\omega \in \mathcal{O}_\mathrm{I}$ with $|\omega| = |\lambda|$ (so that in particular $\theta_{\lambda'}^{\mathscr{U}}$ vanishes on ∂I), and where B^s contains the remaining ('singular') entries.

For the regular entries, we can estimate

$$|\langle(\theta_{\lambda'}^{\mathscr{U}})', \theta_\lambda^{\mathscr{V}_1}\rangle_{L_2(\mathrm{I})}| = |\langle\theta_{\lambda'}^{\mathscr{U}}, (\theta_\lambda^{\mathscr{V}_1})'\rangle_{L_2(\mathrm{I})}| \lesssim \|\theta_{\lambda'}^{\mathscr{U}}\|_{L_1(\mathrm{I})}2^{-|\lambda'|}\|\theta_\lambda^{\mathscr{V}_1}\|_{W_\infty^2(\operatorname{supp}\theta_{\lambda'}^{\mathscr{U}})}$$

$$\lesssim 2^{-|\lambda'|/2}2^{-|\lambda'|}\|(\theta_{\lambda'}^{\mathscr{U}})'\|_{L_2(\mathrm{I})}2^{-|\lambda'|}4^{|\lambda|}2^{|\lambda|/2}\|\theta_\lambda^{\mathscr{V}_1}\|_{L_2(\mathrm{I})} \eqsim 2^{\frac{5}{2}(|\lambda|-|\lambda'|)}\|(\theta_{\lambda'}^{\mathscr{U}})'\|_{L_2(\mathrm{I})},$$

where we used (t_4), Poincaré's inequality, an inverse inequality, and $\|\theta_\lambda^{\mathscr{V}_1}\|_{L_2(\mathrm{I})} \eqsim 1$. An application of Lemma A.1 with $\xi = 1$ and $\varrho = 2$ shows that $\||B^r|\| \lesssim 4^{-k}$.

For the remaining singular entries, we estimate

$$|\langle(\theta_{\lambda'}^{\mathscr{U}})', \theta_\lambda^{\mathscr{V}_1}\rangle_{L_2(\mathrm{I})}| \leq \|(\theta_{\lambda'}^{\mathscr{U}})'\|_{L_1(\mathrm{I})}\|\theta_\lambda^{\mathscr{V}_1}\|_{L_\infty(\mathrm{I})} \lesssim 2^{-\frac{1}{2}(|\lambda|-|\lambda'|)}\|(\theta_{\lambda'}^{\mathscr{U}})'\|_{L_2(\mathrm{I})}.$$

An application of Lemma A.1 with $\xi = 0$ and $\varrho = \frac{1}{2}$ shows that $\||B^s|\| \lesssim 2^{-k/2}$. $\square$

The following Corollary will be used to bound $\|(\mathbf{r}_1 - \tilde{\mathbf{r}}_1)|_{\vee_\mathscr{U}}\|$.

Corollary A.7. *Let $\Lambda^{\mathscr{V}_1} \subset \vee_{\mathscr{V}_1}$ be a multi-tree. Then for $k \in \mathbb{N}_0$,*

$$\left.\begin{array}{l}\|\langle\frac{\partial}{\partial t}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1}\rangle_{L_2(\mathrm{I}\times\Omega)}|_{(\vee_\mathscr{U}\setminus\vee_\mathscr{U}(\Lambda^{\mathscr{V}_1},k))\times\Lambda^{\mathscr{V}_1}}\| \\ \|\langle b_i\frac{\partial}{\partial x_i}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1}\rangle_{L_2(\mathrm{I}\times\Omega)}|_{(\vee_\mathscr{U}\setminus\vee_\mathscr{U}(\Lambda^{\mathscr{V}_1},k))\times\Lambda^{\mathscr{V}_1}}\| \\ \|\langle c\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1}\rangle_{L_2(\mathrm{I}\times\Omega)}|_{(\vee_\mathscr{U}\setminus\vee_\mathscr{U}(\Lambda^{\mathscr{V}_1},k))\times\Lambda^{\mathscr{V}_1}}\|\end{array}\right\} \lesssim 2^{-k/2}.$$

Proof. (a). From $\|\theta_\lambda^{\mathscr{U}} \otimes \sigma_\mu^{\mathscr{U}}\|_{\mathscr{U}} \geq \|\theta_\lambda^{\mathscr{U}}\|_{H^1(\mathrm{I})}\|\sigma_\mu^{\mathscr{U}}\|_{H^{-1}(\Omega)}$, for the first inequality it is sufficient to prove that

$$\left\|\left\langle \frac{(\Theta^{\mathscr{U}})'}{\|\Theta^{\mathscr{U}}\|_{H^1(\mathrm{I})}} \otimes \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^{-1}(\Omega)}}, \Theta^{\mathscr{V}_1} \otimes \Sigma^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I}\times\Omega)}\Big|_{(\vee_{\mathscr{U}}\setminus\vee_{\mathscr{U}}(\Lambda^{\mathscr{V}_1},k))\times\Lambda^{\mathscr{V}_1}}\right\| \lesssim 2^{-k/2}.$$

From $\Theta^{\mathscr{U}}/\|\Theta^{\mathscr{U}}\|_{H^1(\mathrm{I})}$, $\Theta^{\mathscr{V}_1}$, $\Sigma^{\mathscr{U}}/\|\Sigma^{\mathscr{U}}\|_{H^{-1}(\Omega)}$, and $\Sigma^{\mathscr{V}_1}$ being Riesz bases for $H^1(\mathrm{I})$, $L_2(\mathrm{I})$, $H^{-1}(\Omega)$, and $H_0^1(\Omega)$, we have

$$\left\langle \frac{(\Theta^{\mathscr{U}})'}{\|\Theta^{\mathscr{U}}\|_{H^1(\mathrm{I})}}, \Theta^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I})} \in \mathcal{L}(\ell_2(\triangleleft_{\mathscr{V}_1}), \ell_2(\triangleleft_{\mathscr{U}})),$$

$$\left\langle \frac{\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^{-1}(\Omega)}}, \Sigma^{\mathscr{V}_1}\right\rangle_{L_2(\Omega)} \in \mathcal{L}(\ell_2(\lozenge_{\mathscr{V}_1}), \ell_2(\lozenge_{\mathscr{U}})).$$

The proof of the first inequality is completed by applications of Proposition A.4 and Lemmata A.5(first statement)–A.6.

(b). From span $\frac{\partial}{\partial x_i}b_i\Psi^{\mathscr{V}_1}|_{\Lambda^{\mathscr{V}_1}} \subset$ span $\Psi^a|_{\vee_a(\Lambda^{\mathscr{V}_1},0)}$ (similar to Lemma 4.16), for $\mathbf{c} \in \ell_2(\Lambda^{\mathscr{V}_1})$ there exists a $\mathbf{d} \in \ell_2(\vee_a(\Lambda^{\mathscr{V}_1},0))$ such that

$$\left\langle b_i\frac{\partial}{\partial x_i}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I}\times\Omega)}\mathbf{c} = \left\langle \Psi^{\mathscr{U}}, -\mathbf{c}^\top\frac{\partial}{\partial x_i}(b_i\Psi^{\mathscr{V}_1})\right\rangle_{L_2(\mathrm{I}\times\Omega)}$$

$$= \left\langle \Psi^{\mathscr{U}}, \mathbf{d}^\top\Psi^a\right\rangle_{L_2(\mathrm{I}\times\Omega)} = \left\langle \Psi^{\mathscr{U}}, \Psi^a\right\rangle_{L_2(\mathrm{I}\times\Omega)}\mathbf{d},$$

where

$$\|\mathbf{d}\| \eqsim \|\mathbf{d}^\top\Psi^a\|_{L_2(\mathrm{I}\times\Omega)} = \|\mathbf{c}^\top\frac{\partial}{\partial x_i}(b_i\Psi^{\mathscr{V}_1})\|_{L_2(\mathrm{I}\times\Omega)} \lesssim \|\mathbf{c}\|.$$

From $\|\theta_\lambda^{\mathscr{U}}\otimes\sigma_\mu^{\mathscr{U}}\|_{\mathscr{U}} \geq \|\theta_\lambda^{\mathscr{U}}\|_{L_2(\mathrm{I})}\|\sigma_\mu^{\mathscr{U}}\|_{H^1(\Omega)}$, and $\vee_{\mathscr{U}}(\Lambda^{\mathscr{V}_1},k) = \vee_{\mathscr{U}}(\vee_a(\Lambda^{\mathscr{V}_1},0),k)$ it remains to be proven that

$$\left\|\left\langle \frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}}\otimes\frac{\frac{\partial}{\partial x_i}\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}}, \Theta^a\otimes\Sigma^a\right\rangle_{L_2(\mathrm{I}\times\Omega)}\Big|_{(\vee_{\mathscr{U}}\setminus\vee_{\mathscr{U}}(\vee_a(\Lambda^{\mathscr{V}_1},0),k))\times\vee_a(\Lambda^{\mathscr{V}_1},0)}\right\| \lesssim 2^{-k/2}.$$

Indeed, this gives

$$\left\|\left(\left\langle b_i\frac{\partial}{\partial x_i}\Psi^{\mathscr{U}}, \Psi^{\mathscr{V}_1}\right\rangle_{L_2(\mathrm{I}\times\Omega)}\mathbf{c}\right)\Big|_{\vee_{\mathscr{U}}\setminus\vee_{\mathscr{U}}(\Lambda^{\mathscr{V}_1},k)}\right\|$$

$$= \left\|\left(\left\langle \Psi^{\mathscr{U}}, \Psi^a\right\rangle_{L_2(\mathrm{I}\times\Omega)}\mathbf{d}\right)\Big|_{\vee_{\mathscr{U}}\setminus\vee_{\mathscr{U}}(\Lambda^{\mathscr{V}_1},k)}\right\| \lesssim 2^{-k/2}\|\mathbf{d}\| \lesssim 2^{-k/2}\|\mathbf{c}\|,$$

showing the second inequality.

From $\Theta^{\mathscr{U}}/\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}$, Θ^a, $\Sigma^{\mathscr{U}}/\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}$, and Σ^a being Riesz bases for $L_2(\mathrm{I})$, $L_2(\mathrm{I})$, $H_0^1(\Omega)$, and $L_2(\Omega)$, we have

$$\left\langle \frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}}, \Theta^a\right\rangle_{L_2(\mathrm{I})} \in \mathcal{L}(\ell_2(\triangleleft_a), \ell_2(\triangleleft_{\mathscr{U}})),$$

$$\left\langle \frac{\frac{\partial}{\partial x_i}\Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}}, \Sigma^a\right\rangle_{L_2(\Omega)} \in \mathcal{L}(\ell_2(\lozenge_{\mathscr{V}_a}), \ell_2(\lozenge_{\mathscr{U}})).$$

The proof of the remaining inequality is completed by applications of Proposition A.4, Lemma A.2 for $(\Theta^*, \Theta^\circ) = (\frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{L_2(\mathrm{I})}}, \Theta^a)$, and the second statement from Lemma A.5.

(c). A subset of the arguments that showed the second inequality gives the third one. $\qquad\square$

Lemma A.8. *For $k \in \mathbb{N}_0$ and $1 \le i \le n$, it holds that*

$$\left\| \left| \left\langle \Sigma^{\mathscr{P}}, \partial_i \Sigma^{\mathscr{V}_1} \right\rangle_{L_2(\Omega)^n} \big|_{\{(\mu', \mu) \,:\, |\mu'| > |\mu| + k\}} \right| \right\| \lesssim 2^{-k/2},$$

Proof. We split the matrix into $B^r + B^s$, where B^r contains all its entries $\left\langle \sigma_{\mu'}^{\mathscr{P}}, \partial_i \sigma_\mu^{\mathscr{V}_1} \right\rangle_{L_2(\Omega)^n}$ for which supp $\sigma_{\mu'}^{\mathscr{P}}$ is contained in $\omega \cap \Omega$ for some $\omega \in \mathcal{O}_\Omega$ with $|\omega| = |\mu|$ (the 'regular' entries), and where B^s contains the remaining ('singular') entries.

For the regular entries using (s_4) we can estimate

$$|\langle \sigma_{\mu'}^{\mathscr{P}}, \partial_i \sigma_\mu^{\mathscr{V}_1} \rangle_{L_2(\Omega)^n}| \lesssim \| \sigma_{\mu'}^{\mathscr{P}} \|_{L_1(\Omega)^n} 2^{-|\mu'|} |\sigma_\mu^{\mathscr{V}_1}|_{W_\infty^2(\mathrm{supp}\, \sigma_{\mu'}^{\mathscr{P}})}$$

$$\eqsim 2^{(|\mu| - |\mu'|)(1 + \frac{n}{2})}$$

An application of Lemma A.1 with $\xi = n$ and $\varrho = 1$ shows that $\| |B^r| \| \lesssim 2^{-k}$.

Since the wavelets $\sigma_\mu^{\mathscr{V}_1}$ are piecewise polynomial, and functions in $H^1(\Omega)$, they are contained in $W_\infty^1(\Omega)$. For the remaining singular entries we estimate

$$|\langle \sigma_{\mu'}^{\mathscr{P}}, \partial_i \sigma_\mu^{\mathscr{V}_1} \rangle_{L_2(\Omega)^n}| \lesssim \| \sigma_{\mu'}^{\mathscr{P}} \|_{L_1(\Omega)^n} |\sigma_\mu^{\mathscr{V}_1}|_{W_\infty^1(\mathrm{supp}\, \sigma_{\mu'}^{\mathscr{P}})} \lesssim 2^{(|\mu| - |\mu'|)\frac{n}{2}}.$$

An application of Lemma A.1 with $\xi = n - 1$ and $\varrho = 1/2$ shows that $\| |B^s| \| \lesssim 2^{-k/2}$. $\qquad\square$

The following Corollary will be used to bound $\| (\mathbf{r}_1 - \tilde{\mathbf{r}}_1) |_{\vee_{\mathscr{B}}} \|$.

Corollary A.9. *Let $\Lambda^{\mathscr{V}_1} \subset \vee_{\mathscr{V}_1}$ be a multi-tree. Then for $k \in \mathbb{N}_0$,*

$$\left\| \left\langle \Psi^{\vec{\mathscr{P}}}, \nabla_x \Psi^{\mathscr{V}_1} \right\rangle_{L_2(\mathrm{I} \times \Omega)^n} \big|_{(\vee_{\mathscr{B}} \setminus \vee_{\mathscr{B}}(\Lambda^{\mathscr{V}_1}, k)) \times \Lambda^{\mathscr{V}_1}} \right\| \lesssim 2^{-k/2}.$$

Proof. Using Lemma A.2 for $(\Theta^*, \Theta^\circ) = (\Theta^{\mathscr{P}}, \Theta^{\mathscr{V}_1})$, Lemma A.8, $\left\langle \Theta^{\mathscr{P}}, \Theta^{\mathscr{V}_1} \right\rangle_{L_2(\mathrm{I})} \in \mathcal{L}(\lhd_{\mathscr{V}_1}, \lhd_{\mathscr{P}})$, and $\left\langle \Sigma^{\mathscr{P}}, \partial_i \Sigma^{\mathscr{V}_1} \right\rangle_{L_2(\Omega)^n} \in \mathcal{L}(\Diamond_{\mathscr{V}_1}, \Diamond_{\mathscr{P}})$, the proof follows from Proposition A.4. $\qquad\square$

Lemma A.10. *For $k \in \mathbb{N}_0$, $1 \le i \le n$, it holds that*

$$\left\| \left| \left\langle \frac{\nabla \Sigma^{\mathscr{U}}}{\| \Sigma^{\mathscr{U}} \|_{H^1(\Omega)}}, \Sigma^a \mathbf{e}_i \right\rangle_{L_2(\Omega)^n} \big|_{\{(\mu', \mu) \,:\, |\mu'| > |\mu| + k\}} \right| \right\| \lesssim 2^{-k/2},$$

$$\left\| \langle \Sigma^{\mathscr{P}}, \Sigma^a \rangle_{L_2(\Omega)^n} \big|_{\{(\mu', \mu) \,:\, |\mu'| > |\mu| + k\}} \right\| \lesssim 2^{-k/2}.$$

Proof. For proving the first inequality, we split the matrix into $B^r + B^s$, where B^r contains all its entries $\left\langle \frac{\nabla \sigma_{\mu'}^{\mathscr{U}}}{\| \sigma_{\mu'}^{\mathscr{U}} \|_{H^1(\Omega)}}, \sigma_\mu^a \mathbf{e}_i \right\rangle_{L_2(\Omega)^n}$ for which supp $\sigma_{\mu'}^{\mathscr{U}}$ is contained in $\omega \cap \Omega$ for some $\omega \in \mathcal{O}_\Omega$ with $|\omega| = |\mu|$ (the 'regular' entries), and where B^s contains the remaining ('singular') entries.

For the regular entries, using (s_4) and the first inequality in (4.9), we estimate

$$\left| \left\langle \frac{\partial_i \sigma_{\mu'}^{\mathscr{U}}}{\| \sigma_{\mu'}^{\mathscr{U}} \|_{H^1(\Omega)}}, \sigma_\mu^a \right\rangle_{L_2(\Omega)^n} \right| = \left| \left\langle \frac{\sigma_{\mu'}^{\mathscr{U}}}{\| \sigma_{\mu'}^{\mathscr{U}} \|_{H^1(\Omega)}}, \partial_i \sigma_\mu^a \right\rangle_{L_2(\Omega)^n} \right|$$

$$\lesssim 2^{-|\mu'|} \frac{\| \sigma_{\mu'}^{\mathscr{U}} \|_{L_1(\Omega)}}{\| \sigma_{\mu'}^{\mathscr{U}} \|_{H^1(\Omega)}} |\sigma_\mu^a|_{W^2(\mathrm{supp}\, \sigma_{\mu'}^{\mathscr{U}})} \lesssim 2^{-|\mu'|} 2^{-|\mu'| n/2} 2^{-|\mu'|} 2^{|\mu| n/2} 2^{2|\mu|}$$

$$= 2^{(2 + n/2)(|\mu'| - |\mu|)}.$$

An application of Lemma A.1 with $\xi = n$ and $\varrho = 2$ shows that $\| |B^r| \| \lesssim 4^{-k}$.

For the singular entries, we estimate

$$\left| \left\langle \frac{\partial_i \sigma_{\mu'}^{\mathscr{U}}}{\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}}, \sigma_{\mu}^a \right\rangle_{L_2(\Omega)^n} \right| \lesssim \frac{\|\partial_i \sigma_{\mu'}^{\mathscr{U}}\|_{L_1(\Omega)}}{\|\sigma_{\mu'}^{\mathscr{U}}\|_{H^1(\Omega)}} \|\sigma_{\mu}^a\|_{L_\infty(\Omega)} \lesssim 2^{(|\mu'|-|\mu|)n/2}.$$

An application of Lemma A.1 with $\xi = n-1$ and $\varrho = 1/2$ shows that $\||B^s|\| \lesssim 2^{-k/2}$.

The proof of the second inequality proceeds along the by now well-known steps. Using assumption (s_4) on $\Sigma^{\mathscr{P}}$ one shows that $\||B^r|\| \lesssim 2^{-k}$, whereas $\||B^s|\| \lesssim 2^{-k/2}$. $\qquad\square$

The following Corollary will be used to bound $\|\mathbf{r}_3 - \tilde{\mathbf{r}}_3\|$.

Corollary A.11. *Let $\Lambda^a \subset \vee_a$ be a multi-tree. Then for $k \in \mathbb{N}_0$,*

$$\left\| \left\langle \nabla_x \Psi^{\mathscr{U}}, \Psi^a \mathbf{e}_i \right\rangle_{L_2(I \times \Omega)^n} \big|_{(\vee_{\mathscr{U}} \setminus \vee_{\mathscr{U}}(\Lambda^a, k)) \times \Lambda^a} \right\| \lesssim 2^{-k/2}.$$

$$\left\| \left\langle \Psi^{\vec{\mathscr{P}}}, \Psi^a \mathbf{e}_i \right\rangle_{L_2(I \times \Omega)^n} \big|_{(\vee_{\mathscr{P}} \setminus \vee_{\mathscr{P}}(\Lambda^a, k)) \times \Lambda^a} \right\| \lesssim 2^{-k/2}.$$

Proof. From $\|\theta_\lambda^{\mathscr{U}} \otimes \sigma_\mu^{\mathscr{U}}\|_{\mathscr{U}} \geq \|\theta_\lambda^{\mathscr{U}}\|_{L_2(I)} \|\sigma_\mu^{\mathscr{U}}\|_{H^1(\Omega)}$, in order to prove the first result it suffices to show that

$$\left\| \left\langle \frac{\Theta^{\mathscr{U}}}{\|\Theta^{\mathscr{U}}\|_{L_2(I)}} \otimes \frac{\nabla \Sigma^{\mathscr{U}}}{\|\Sigma^{\mathscr{U}}\|_{H^1(\Omega)}}, \Theta^a \otimes \Sigma^a \mathbf{e}_i \right\rangle_{L_2(I \times \Omega)^n} \Big|_{\vee_{\mathscr{U}} \setminus \vee_{\mathscr{U}}(\Lambda^a, k) \times \Lambda^a} \right\| \lesssim 2^{-k/2}.$$

This and the second result follow from applications of Proposition A.4, Lemma A.2, and Lemma A.10. $\qquad\square$

References

[AG12]   H. Aimar and I. Gómez, *Parabolic Besov regularity for the heat equation*, Constr. Approx. **36**(1), 145–159 (2012).

[Alp93]   B. Alpert, *A class of bases in L^2 for the sparse representation of integral operators*, SIAM J. Math. Anal. **24**, 246–262 (1993).

[And14]   R. Andreev, *Space-time discretization of the heat equation*, Numer. Algorithms **67**(4), 713–731 (2014).

[BD04]   P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math. **97**(2), 193 – 217 (2004).

[BG04]   H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer. **13**, 147–269 (2004).

[BJ89]   I. Babuška and T. Janik, *The h-p version of the finite element method for parabolic equations. I. The p-version in time*, Numer. Methods Partial Differential Equations **5**(4), 363–399 (1989).

[BJ90]   I. Babuška and T. Janik, *The h-p version of the finite element method for parabolic equations. II. The h-p version in time*, Numer. Methods Partial Differential Equations **6**(4), 343–369 (1990).

[BZ96]   R. Balder and C. Zenger, *The solution of multidimensional real Helmholtz equations on sparse grids*, SIAM J. Sci. Comput. **17**(3), 631–646 (1996).

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

[CDD+14]   P. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter and R. Schilling, *Convergence analysis of spatially adaptive Rothe methods*, Found. Comput. Math. **14**(5), 863–912 (2014).

[CDDD01]   A. Cohen, W. Dahmen, I. Daubechies and R. DeVore, *Tree approximation and optimal encoding*, Appl. Comput. Harmon. Anal. **11**(2), 192–226 (2001).

[CDFS13]   N. Chegini, S. Dahlke, U. Friedrich and R. Stevenson, *Piecewise tensor product wavelet bases by extensions and approximation rates*, Math. Comp. **82**, 2157–2190 (2013).

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

[CS15]   N. Chegini and R. Stevenson, *An adaptive wavelet method for semi-linear first order system least squares*, Comput. Math. Appl. (August 2015), DOI: 10.1515/cmam-2015-0023.

[DFW16]   W. Dörfler, S. Findeisen and C. Wieners, *Space-time discontinuous Galerkin discretizations for linear first-order hyperbolic evolution systems*, Comput. Methods Appl. Math. **16**(3), 409–428 (2016).

[DL92]   R. Dautray and J.-L. Lions, *Mathematical analysis and numerical methods for science and technology. Vol. 5*, Springer-Verlag, Berlin, 1992, Evolution problems I.

[DS99]   W. Dahmen and R. Stevenson, *Element-by-element construction of wavelets satisfying stability and moment conditions*, SIAM J. Numer. Anal. **37**(1), 319–352 (1999).

[DS10]   M. Dauge and R. Stevenson, *Sparse tensor product wavelet approximation of singular functions*, SIAM J. Math. Anal. **42**(5), 2203–2228 (2010).

[ECD16]   T. Ellis, J. Chan and L. Demkowicz, Robust DPG methods for transient convection-diffusion, in *Building bridges: connections and challenges in modern approaches to numerical partial differential equations*, volume 114 of *Lect. Notes Comput. Sci. Eng.*, pages 179–203, Springer, [Cham], 2016.

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

[GK11]   M. Gunzburger and A. Kunoth, *Space-Time Adaptive Wavelet Methods for Control Problems Constrained by Parabolic Evolution Equations*, SIAM J. Contr. Optim. **49**(3), 1150–1170 (2011).

[GN16]   M. Gander and M. Neumüller, *Analysis of a new space-time parallel multigrid algorithm for parabolic problems*, SIAM J. Sci. Comput. **38**(4), A2173–A2208 (2016).

[GO95]   M. Griebel and P. Oswald, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Adv. Comput. Math. **4**(1–2), 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).

[KS14]   S. Kestler and R. Stevenson, *Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions*, J. Comput. Appl. Math. **260**, 103–116 (2014).

[KSU16]   S. Kestler, K. Steih and K. Urban,   *An efficient space-time adaptive wavelet Galerkin method for time-periodic parabolic partial differential equations*, Math. Comp. **85**(299), 1309–1333 (2016).

[LMN16]   U. Langer, S. Moore and M. Neumüller, *Space-time isogeometric analysis of parabolic evolution problems*, Comput. Methods Appl. Mech. Engrg. **306**, 342–363 (2016).

[MS02]   M. Majidi and G. Starke, *Least-squares Galerkin methods for parabolic problems. II. The fully discrete case and adaptive algorithms*, SIAM J. Numer. Anal. **39**(5), 1648–1666 (2001/02).

[MS01]   M. Majidi and G. Starke, *Least-squares Galerkin methods for parabolic problems. I. Semidiscretization in time*, SIAM J. Numer. Anal. **39**(4), 1302–1323 (2001).

[MST14]   M. Messner, M. Schanz and J. Tausch, *A fast Galerkin method for parabolic space-time boundary integral equations*, J. Comput. Phys. **258**, 15–30 (2014).

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

[NS09]   H. Nguyen and R. Stevenson, *Finite element wavelets with improved quantitative properties*, J. Comput. Appl. Math. **230**(2), 706–727 (2009).

[Rek18]   N. Rekatsinas, *Optimal Adaptive Wavelet Methods for solving First Order System Least Squares*, PhD thesis, University of Amsterdam, 2018.

[RS18a]   N. Rekatsinas and R. Stevenson, *Optimal adaptive wavelet methods for first order system least squares*, Numer. Math. (2018).

[RS18b]   N. Rekatsinas and R. Stevenson, *A quadratic finite element wavelet Riesz basis*, Int. J. Wavelets Multiresolut. Inf. Process (2018).

[SS09]   C. Schwab and R. Stevenson, *A space-time adaptive wavelet method for parabolic evolution problems*, Math. Comp. **78**, 1293–1318 (2009).

[SS17]   C. Schwab and R. Stevenson,   *Fractional space-time variational formulations of (Navier)-Stokes equations*, SIAM J. Math. Anal. **49**(4), 2442–2467 (2017).

[Ste98]   R. Stevenson, *Stable Three-Point Wavelet Bases on General Meshes*, Numer. Math. **80**, 131–158 (1998).

[Ste14]   R. Stevenson, *Adaptive wavelet methods for linear and nonlinear least-squares problems*, Found. Comput. Math. **14**(2), 237–283 (2014).

[Ste15]   O. Steinbach, *Space-Time Finite Element Methods for Parabolic Problems*, Comput. Methods Appl. Math. **15**(4), 551–566 (2015).

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

[Tem97]   R. Temam, *Infinite-dimensional dynamical systems in mechanics and physics*, volume 68 of *Applied Mathematical Sciences*, Springer-Verlag, New York, second edition, 1997.

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

E-mail address: n.rekatsinas@uva.nl, r.p.stevenson@uva.nl