ADAPTIVE WAVELET SCHEMES FOR PARABOLIC PROBLEMS: SPARSE MATRICES AND NUMERICAL RESULTS*

NABI CHEGINI† AND ROB STEVENSON‡

Abstract. A simultaneous space-time variational formulation of a parabolic evolution problem is solved with an adaptive wavelet method. This method is shown to converge with the best possible rate in linear complexity. Thanks to the use of tensor product bases, there is no penalty in complexity due to the additional time dimension. Special wavelets are designed such that the bi-infinite system matrix is sparse. This sparsity largely simplifies the implementation and improves the quantitative properties of the adaptive wavelet method. Numerical results for an ODE and the heat equation are presented.

AMS subject classifications. 35K15, 41A25, 42C40, 65F50, 65T60.

Key words. Adaptive wavelet scheme, parabolic equations, simultaneous space-time variational formulation, tensor product approximation, optimal computational complexity.

1. Introduction. This paper is concerned with the numerical solution of linear parabolic evolution equations. In general, even for a smooth source term and initial condition, the solution of such a problem has a reduced smoothness near the bottom and walls of the space-time cylinder. As a consequence, standard discretization methods based on uniform meshes will converge at reduced rates.

A general approach to retrieve the best possible rate allowed by the order of the discretization is to apply *adaptive* methods. Standard methods for solving time evolution problems first discretize in space and then in time (method of lines), or first in time and then in space (Rothe's method). As a consequence, with these time marching methods it seems hard to be able to arrive at an optimal distribution of the "mesh-points" or degrees of freedom simultaneously over space *and* time.

In this paper, we consider a simultaneous space-time variational formulation of the parabolic problem. This formulation is well-posed in the sense that it defines a boundedly invertible operator between a Hilbert space and the dual of another Hilbert space. We equip both Hilbert spaces – being Bochner spaces or intersections of those – by Riesz bases that are tensor products of temporal and spatial wavelet bases. In this way, we arrive at an equivalent, well-posed bi-infinite matrix vector problem. We solve this problem with an *adaptive wavelet method* applied to the *normal equations*.

The advantages of our approach are two-fold: Firstly, thanks to the tensor-product construction of the basis, there is a nearly neglectable penalty in asymptotic computational complexity due to the additional time dimension, an effect that is well-known for so-called *sparse-grid* or *hyperbolic cross* approximation methods. Secondly, the adaptive wavelet method is proven to converge at the best possible rate, in linear complexity.

Our approach was investigated earlier in [SS09], and on a more heuristical level, in [GO07]. Compared to the first work, here we use an alternative variational formulation in which the initial condition is incorporated more naturally. Furthermore, we design

*The first author was supported by the Netherlands Organisation for Scientific Research (NWO), contract no. 613.000.902.

†Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands (n.godarzvandchegini@uva.nl), and University of Tafresh, Tafresh 39518-79611, Iran.

‡Korteweg - de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands (R.P.Stevenson@uva.nl)

a new wavelet basis with respect to which any parabolic differential operator of second order with constant coefficient gives rise to a bi-infinite system matrix that is *truly sparse*. Having such a matrix largely simplifies the implementation and improves the quantitative properties of the adaptive wavelet method. We present numerical results for an ordinary differential equation and for the heat equation.

This paper is organized as follows: In Sect. 2, a space-time variational formulation of the parabolic problem is derived, and mapping properties of the resulting operator are given. By equipping the occurring Hilbert spaces with tensor products of temporal and spatial wavelet bases, in Sect. 3 an equivalent bi-infinite matrix-vector problem is derived. In Sect. 4, it is investigated what the best possible approximation rates are from the tensor product bases in various situations. The adaptive wavelet method for solving the bi-infinite matrix-vector problem is described in Sect. 5. Spatial wavelets that give rise to a sparse representation of the elliptic part of the operator were designed in [DS10b, CS10]. In Sect. 6, temporal test and trial wavelets are constructed such that the parabolic problem with respect to the tensor product wavelets is sparse. Numerical results for an ODE obtained with these temporal wavelets are presented in Sect. 7. Finally, in Sect. 8 numerical results are given that are obtained by the overall scheme applied to the heat equation.

In this paper, 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 $C \gtrsim D$.

2. Parabolic problem in variational form. Let V, H be separable Hilbert spaces, for convenience over $\mathbb{R}$, such that $V \hookrightarrow H$ with dense embedding. Identifying H with its dual, we obtain the Gelfand triple $V \hookrightarrow H \hookrightarrow V'$. We use the notation $\langle \cdot, \cdot \rangle_H$ both to denote the scalar product on $H \times H$ and its unique extension by continuity to the duality pairing on $V' \times V$.

Let $0 < T < \infty$ and, for a.e. $t \in \mathrm{I} := (0, T)$, let $a(t; \cdot, \cdot)$ denote a bilinear form on $V \times V$ such that for any $\eta, \zeta \in V$, $t \mapsto a(t; \eta, \zeta)$ is measurable on I, and such that for some constant $\lambda_0 \in \mathbb{R}$ and for a.e. $t \in \mathrm{I}$,

$$|a(t; \eta, \zeta)| \lesssim \|\eta\|_V \|\zeta\|_V \quad (\eta, \zeta \in V) \quad (\textit{boundedness}), \tag{2.1}$$

$$a(t; \eta, \eta) + \lambda_0 \|\eta\|_H^2 \gtrsim \|\eta\|_V^2 \quad (\eta \in V) \quad (\textit{Gårding inequality}). \tag{2.2}$$

For a.e. $t \in \mathrm{I}$, let $A(t) \in \mathcal{L}(V, V')$ be defined by

$$\langle A(t)\eta, \zeta \rangle_H = a(t; \eta, \zeta).$$

Given $g \in L_2(\mathrm{I}; V')$ and $u_0 \in H$, we are interested in solving the parabolic problem of finding, for a.e. $t \in \mathrm{I}$, $u(t) \in V$ such that

$$\dot{u}(t) + A(t)u(t) = g(t) \quad \text{in } V', \quad u(0) = u_0 \text{ in } H. \tag{2.3}$$

Remark 2.1. In particular, we have in mind $A(t)$ being a linear, scalar differential or integrodifferential operator of order $2m \geq 0$ on a bounded domain $\Omega \subset \mathbb{R}^n$ in variational form (systems of equations will not impose any difficulties apart from a more involved notation). Then, $H = L_2(\Omega)$ and $V = H^m(\Omega)$ or a closed subspace incorporating homogeneous Dirichlet boundary conditions.

By multiplying (2.3) by smooth test functions v of time and space that vanish at $t = T$, integrating both sides over time and space, and, other than in [SS09], by

applying integration by parts to the first term, we end up with the variational problem of finding u such that for all such v

$$b(u,v) = f(v), \tag{2.4}$$

where

$$b(w,v) := \int_{\mathrm{I}} -\langle w(t), \dot{v}(t)\rangle_H + a(t; w(t), v(t))dt$$
$$f(v) := \int_{\mathrm{I}} \langle g(t), v(t)\rangle_H dt + \langle u_0, v(0)\rangle_H.$$

For symmetric a, specifically for $a(t;\eta,\zeta) = \int_\Omega \nabla\eta\cdot\nabla\zeta$, this variational formulation was also studied in [BJ89].

THEOREM 2.2. *With $\mathcal{X} := L_2(\mathrm{I};V)$ and $\mathcal{Y} := L_2(\mathrm{I};V)\cap H^1_{0,\{T\}}(\mathrm{I};V')$, the operator $B\in\mathcal{L}(\mathcal{X},\mathcal{Y}')$ defined by $(Bw)(v) = b(w,v)$ is boundedly invertible.*

Here $H^1_{0,\{T\}}(\mathrm{I})$ denotes the closure in $H^1(\mathrm{I})$ of the space of smooth functions on I that vanish at $t=T$, and $\mathcal{Y}$, being the intersection of the Hilbert spaces $L_2(\mathrm{I};V)$ and $H^1_{0,\{T\}}(\mathrm{I};V')$, is a Hilbert space with squared norm $\|\cdot\|^2_{L_2(\mathrm{I};V)} + \|\cdot\|^2_{H^1_{0,\{T\}}(\mathrm{I};V')}$.

Remark 2.3. Since for $v\in\mathcal{Y}$, one has $v(0)\in H$ with $\|v(0)\|_H \lesssim \|v\|_{\mathcal{Y}}$ (see [SS09, §5] and the references cited there), for $u_0\in H$ and, say, $g\in L_2(\mathrm{I};V')$, it holds that $f\in\mathcal{Y}'$ with $\|f\|_{\mathcal{Y}'}\lesssim \|g\|_{L_2(\mathrm{I};V')} + \|u_0\|_H$.

Theorem 2.2 is proved by checking the following three conditions:

$$\sup_{0\neq w\in\mathcal{X},\, 0\neq v\in\mathcal{Y}} \frac{|b(w,v)|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} < \infty \quad (\textit{continuity}), \tag{2.5}$$

$$\inf_{0\neq v\in\mathcal{Y}} \sup_{0\neq w\in\mathcal{X}} \frac{|b(w,v)|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} > 0 \quad (\textit{inf sup-condition}), \tag{2.6}$$

$$\forall 0\neq w\in\mathcal{X},\ \sup_{0\neq v\in\mathcal{Y}} |b(w,v)| > 0 \quad (\textit{non-degeneracy}), \tag{2.7}$$

This can be done similarly to [SS09, Appendix A]. In [SS09], a different bilinear form b and spaces $\mathcal{X}$ and $\mathcal{Y}$ were applied, because there a variational formulation was derived without performing integration by parts. With the current approach, the condition $u(0) = u_0$ is incorporated in the variational formulation as a natural boundary condition instead of an essential one.

Under some additional conditions, a result similar to Theorem 2.2 is also valid after making a shift of the smoothness indices in space:

THEOREM 2.4. *Let $W\hookrightarrow V$ with dense embedding, for every $t\in\mathrm{I}$, $A(\cdot)'\in C([0,T],\mathcal{L}(W,H))$, and for λ_0 as in (2.2), $A(t)'+\lambda_0 I : W\to H$ be boundedly invertible. Then with $\mathcal{X} := L_2(\mathrm{I};H)$ and $\mathcal{Y} := L_2(\mathrm{I};W)\cap H^1_{0,\{T\}}(\mathrm{I};H)$, $B\in\mathcal{L}(\mathcal{X},\mathcal{Y}')$ is boundedly invertible.*

Indeed, by making the standard transformation $u(t) = \hat{u}(t)e^{\lambda_0 t}$, w.l.o.g. we may assume that $\lambda_0 = 0$. Membership of $B\in\mathcal{L}(\mathcal{X},\mathcal{Y}')$ is equivalent to $B'\in\mathcal{L}(\mathcal{Y},\mathcal{X}')$. In view of $(B'v)(w) = b(w,v)$, the latter is easily verified.

To infer that $(B')^{-1}\in\mathcal{L}(\mathcal{X}',\mathcal{Y})$, that is equivalent to $B^{-1}\in\mathcal{L}(\mathcal{Y}',\mathcal{X})$, note that, for $h\in L_2(\mathrm{I};V')$, the adjoint variational problem of finding $z\in L_2(\mathrm{I};V)\cap H^1_{0,\{T\}}(\mathrm{I};V')$ with

$$(B'z)(w) = b(w,z) = \int_{\mathrm{I}} \langle w(t), -\dot{z}(t) + A'(t)z(t)\rangle_H dt = h(w) \quad (w\in L_2(\mathrm{I};V))$$

is the weak formulation of the problem of finding, for a.e. $t \in \mathrm{I}$, $z(t) \in V$ such that

$$-\dot{z}(t) + A(t)'z(t) = h(t) \quad \text{in } V', \quad z(T) = 0,$$

or, with $\underline{z}(\cdot) := z(T - \cdot)$, $\underline{A}(\cdot) := A(T - \cdot)'$, and $\underline{h}(\cdot) := h(T - \cdot)$,

$$\dot{\underline{z}}(t) + \underline{A}(t)\underline{z}(t) = \underline{h}(t) \quad \text{in } V', \quad \underline{z}(0) = 0. \tag{2.8}$$

For any fixed $s \in [0, T]$, the bounded invertibility of $\underline{A}(s) : W \to H$ and its coercivity show that $-\underline{A}(s)$ generates an analytic semigroup in H, and that for any $\tilde{h} \in \mathcal{X}$, the solution $w(t)$ of, for a.e. $t \in \mathrm{I}$,

$$\dot{w}(t) + \underline{A}(s)w(t) = \tilde{h}(t) \quad \text{in } V', \quad w(0) = 0$$

is in $\mathcal{Y}$ ([dS64], cf. also [Woo07]). As is shown in [PS01, Theorem 2.5], the latter results together with the assumption that $\underline{A}(\cdot) \in C([0, T], \mathcal{L}(W, H))$ show that the solution of (2.8) is in $\mathcal{Y}$ with $\|\underline{z}\|_{\mathcal{Y}} \lesssim \|\underline{h}\|_{\mathcal{X}}$, which was left to show.

Remark 2.5. In the situation of Remark 2.1, and with $A(t)$ being a linear, scalar differential operator of order $2m \geq 0$, with $W := H^{2m}(\Omega) \cap V$ the additional conditions of Theorem 2.4 are fulfilled when both the coefficients of the differential operator and the boundary of Ω are sufficiently smooth. For $m = 1$ and Ω being convex, sufficient smoothness of $\partial\Omega$ reduces to Lipschitz continuity.

3. Equivalent bi-infinite matrix vector equations. Let $\Psi^{\mathcal{X}} = \{\psi^{\mathcal{X}}\}$ and $\Psi^{\mathcal{Y}} = \{\psi^{\mathcal{Y}}\}$ be *Riesz bases* for $\mathcal{X}$ and $\mathcal{Y}$, respectively, where we have in mind *wavelet* bases, and where the pair $(\mathcal{X}, \mathcal{Y})$ is as in either Theorem 2.2 or Theorem 2.4. Then (2.4) can be equivalently formulated as

$$\mathbf{B}\mathbf{u} = \mathbf{f}, \tag{3.1}$$

where $\mathbf{u}$ is the vector of coefficients of u with respect to $\Psi^{\mathcal{X}}$, $\mathbf{f} := [f(\psi^{\mathcal{Y}})]_{\psi^{\mathcal{Y}} \in \Psi^{\mathcal{Y}}}$ and $\mathbf{B} := b(\Psi^{\mathcal{X}}, \Psi^{\mathcal{Y}}) := [b(\psi^{\mathcal{X}}, \psi^{\mathcal{Y}})]_{\psi^{\mathcal{Y}} \in \Psi^{\mathcal{Y}}, \psi^{\mathcal{X}} \in \Psi^{\mathcal{X}}}$. Thanks to Theorems 2.2 or 2.4, the vector $\mathbf{f}$ is in ℓ_2 and both $\mathbf{B}$ and its inverse are bounded mappings on ℓ_2 (cf. [Ste09, §2.1]). In particular, it holds that

$$\|\mathbf{B}\| \leq \|B\|_{\mathcal{X} \to \mathcal{Y}'} \Lambda_{\mathcal{X}}(\Psi^{\mathcal{X}}) \Lambda_{\mathcal{Y}}(\Psi^{\mathcal{Y}}), \ \|\mathbf{B}^{-1}\| \leq \|B^{-1}\|_{\mathcal{Y}' \to \mathcal{X}} \lambda_{\mathcal{X}}(\Psi^{\mathcal{X}})^{-1} \lambda_{\mathcal{Y}}(\Psi^{\mathcal{Y}})^{-1}. \tag{3.2}$$

Here, for a Riesz basis Π for a Hilbert space U, we define the *Riesz constants* $\Lambda_U(\Pi) := \sqrt{\|\langle \Pi, \Pi \rangle_U\|}$ and $\lambda_U(\Pi) := \sqrt{\|\langle \Pi, \Pi \rangle_U^{-1}\|^{-1}}$. With the expressions $\Lambda_U(|\Pi|)$ and $\lambda_U(|\Pi|)$, we will mean the Riesz constants of Π normalized in U.

Another equivalent formulation is given by the normal equations

$$\mathbf{B}^\top \mathbf{B}\mathbf{u} = \mathbf{B}^\top \mathbf{f}. \tag{3.3}$$

The bi-infinite matrix $\mathbf{B}^\top \mathbf{B}$ is boundedly invertible, symmetric and positive definite, and so the *adaptive wavelet schemes* proposed in [CDD01, CDD02] can be applied to (3.3).

The idea of [CDD02] is to apply some convergent iterative scheme to (3.3), or, when such a scheme is available, directly to (3.1). The best possible rate in linear complexity, i.e., *quasi-optimality*, is realized by a recurrent application of *coarsening*.

The approach from [CDD01] is to solve a sequence of Galerkin approximations $(\mathbf{B}^\top \mathbf{B})|_{\Lambda_i \times \Lambda_i} \mathbf{u}^{(i)} = (\mathbf{B}^\top \mathbf{f})|_{\Lambda_i}$, where the expansion of Λ_i to Λ_{i+1} is guided by the

a posteriori error estimator $\mathbf{B}^\top\mathbf{B}\mathbf{u}^{(i)} - \mathbf{B}^\top\mathbf{f}$. As shown in [GHS07], with this approach quasi-optimality is realized without coarsening. We focus on the method from [CDD01] since it turns out to be quantitatively better. We refer to this method as the *adaptive wavelet-Galerkin method* (**AWGM**). For a recent overview of adaptive wavelet methods, we refer to [Ste09]. More details about the **AWGM** will be given in Sect. 5.

Since the spaces $\mathcal{X} = L_2(\mathrm{I}; H)$ and $\mathcal{Y} = L_2(\mathrm{I}; W) \cap H^1_{0,\{T\}}(\mathrm{I}; H)$ from Theorem 2.4 are (intersections of) tensor products of spaces in time and space, a natural construction of Riesz bases for these spaces is as follows: Let $\Theta^{\mathcal{X}}$, $\Theta^{\mathcal{Y}}$, and $\Sigma^{\mathcal{X}}$, $\Sigma^{\mathcal{Y}}$ be collections of temporal or spatial functions such that, normalized in the corresponding norms,

$$\begin{array}{llll} \Theta^{\mathcal{X}} & \text{is a Riesz basis for} & L_2(\mathrm{I}), & \\ \Theta^{\mathcal{Y}} & \quad\text{"} & L_2(\mathrm{I}) & \text{and for } H^1_{0,\{T\}}(\mathrm{I}), \\ \Sigma^{\mathcal{X}} & \quad\text{"} & H, & \\ \Sigma^{\mathcal{Y}} & \quad\text{"} & W & \quad\text{"} \quad H. \end{array}$$

Then, normalized in the corresponding norms,

$$\begin{array}{lll} \Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}} & \text{is a Riesz basis for} & \mathcal{X}, \\ \Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}} & \quad\text{"} & L_2(\mathrm{I}; W), H^1_{0,\{T\}}(\mathrm{I}; H), \text{ and so for } \mathcal{Y}, \end{array}$$

cf. [GO95] for the last statement. In particular, as shown in [GO95],

$$\begin{aligned} \Lambda_{\mathcal{Y}}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|) \leq & \max\Big(\Lambda_{L_2(\mathrm{I};W)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|), \Lambda_{H^1_{0,\{T\}}(\mathrm{I};H)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|)\Big) \\ \lambda_{\mathcal{Y}}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|) \geq & \min\Big(\lambda_{L_2(\mathrm{I};W)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|), \lambda_{H^1_{0,\{T\}}(\mathrm{I};H)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|)\Big), \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \Lambda_{L_2(\mathrm{I};W)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|) = & \Lambda_{L_2(\mathrm{I})}(|\Theta^{\mathcal{Y}}|)\Lambda_W(|\Sigma^{\mathcal{Y}}|), \\ \lambda_{L_2(\mathrm{I};W)}(|\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}|) = & \lambda_{L_2(\mathrm{I})}(|\Theta^{\mathcal{Y}}|)\lambda_W(|\Sigma^{\mathcal{Y}}|), \end{aligned} \tag{3.5}$$

and similar for the Riesz constants of the other normalized tensor product bases for tensor product spaces.

Similarly, for $\mathcal{X} = L_2(\mathrm{I}; V)$ and $\mathcal{Y} = L_2(\mathrm{I}; V) \cap H^1_{0,\{T\}}(\mathrm{I}; V')$ as in Theorem 2.2, if $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ are as above, and if, when normalized in the corresponding norms, $\Sigma^{\mathcal{X}}$ is a Riesz basis for V, and $\Sigma^{\mathcal{Y}}$ is a Riesz basis for V' and V, then, when normalized in the corresponding norms, $\Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}$ is a Riesz basis for $\mathcal{X}$, and $\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}$ is a Riesz basis for $L_2(\mathrm{I}; V)$ and $H^1_{0,\{T\}}(\mathrm{I}; V')$, and so for $\mathcal{Y}$.

With, for $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$, $\mathbf{D}_{\mathcal{Z}} := \mathrm{diag}\{\|\theta^{\mathcal{Z}} \otimes \sigma^{\mathcal{Z}}\|_{\mathcal{Z}} : \theta^{\mathcal{Z}} \in \Theta^{\mathcal{Z}}, \sigma^{\mathcal{Z}} \in \Sigma^{\mathcal{Z}}\}$, for either choice of the pair $(\mathcal{X}, \mathcal{Y})$ it holds that $\mathbf{f} = \mathbf{D}_{\mathcal{Y}}^{-1}[f(\theta^{\mathcal{Y}} \otimes \sigma^{\mathcal{Y}})]_{\theta^{\mathcal{Y}} \in \Theta^{\mathcal{Y}}, \sigma^{\mathcal{Y}} \in \Sigma^{\mathcal{Y}}}$ and

$$\mathbf{B} = \mathbf{D}_{\mathcal{Y}}^{-1}\big[-\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})} \otimes \langle \Sigma^{\mathcal{X}}, \Sigma^{\mathcal{Y}}\rangle_H + \int_{\mathrm{I}} a(t; \Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}, \Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}})dt\big]\mathbf{D}_{\mathcal{X}}^{-1}. \tag{3.6}$$

Furthermore, if the bilinear form a is of the form $a(t; \eta, \zeta) = a_1(t)a_2(\eta, \zeta)$, then

$$\int_{\mathrm{I}} a(t; \Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}, \Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}})dt = \langle a_1(\cdot)\Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})} \otimes a_2(\Sigma^{\mathcal{X}}, \Sigma^{\mathcal{Y}}). \tag{3.7}$$

The diagonal matrix $\mathbf{D}_{\mathcal{Y}}^{-1}$, however, is not of tensor product type.

4. Best possible rates. Suppose that for some $s > 0$, for any N, the solution u of (2.4) *can* be approximated in $\mathcal{X}$ from the span of N (adaptively chosen) elements from $\Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}$ within tolerance $\mathcal{O}(N^{-s})$. Recall that $\mathcal{X} = L_2(\mathrm{I}; Z)$ with either $Z = H$ (Thm. 2.4) or $Z = V$ (Thm. 2.2). Then the aim of the adaptive wavelet schemes is to produce an approximation to u from $\operatorname{span} \Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}$ within a given tolerance ε in $L_2(\mathrm{I}; Z)$ taking not more than $\mathcal{O}(\varepsilon^{-1/s})$ operations.

In this section, for $Z \in \{H, V\}$ we investigate what the best possible rate

$$s_{\max} = s_{\max}(\Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}, L_2(\mathrm{I}; Z))$$

is that can be expected, where we consider u that are *sufficiently smooth.* Then, in order to do so, it is sufficient to consider linear approximation, i.e., the choice of the aforementioned (sequence of) N elements will not depend on u. We will write Θ and Σ for $\Theta^{\mathcal{X}}$ and $\Sigma^{\mathcal{X}}$.

Remark 4.1. For any $s \in (0, s_{\max}]$, the class of functions that can be approximated at rate s with nonlinear approximation is much larger than the corresponding class of functions that can be approximated with that rate with linear approximation, being the reason to consider adaptive algorithms in the first place. A characterization of the nonlinear approximation classes for tensor product approximation in Sobolev spaces in terms of certain tensor products of Besov spaces can be found in [Nit06, SU09]. For elliptic problems, in particular for the Poisson problem, corresponding regularity theory has been developed in [DS10a]. For parabolic problems, it seems that such a regularity theory still has to be investigated.

Let the best possible approximation rate from Σ in Z be $s_{\max} = s_{\max}(\Sigma, Z)$. This means that there exists a sequence $\Sigma_1 \subset \Sigma_2 \subset \cdots$ with $\cup_i \Sigma_i = \Sigma$ and $\#\Sigma_i \eqsim \rho(\Sigma)^i$ for some constant $\rho(\Sigma) > 1$, such that for some densely embedded subspace $\hat{Z}$ of Z, and with the (biorthogonal) projector $Q^{\Sigma_i} : Z \to \operatorname{span} \Sigma_i : u = \sum_{\sigma \in \Sigma} \mathbf{u}_\sigma \sigma \mapsto \sum_{\sigma \in \Sigma_i} \mathbf{u}_\sigma \sigma$, it holds that $\|I - Q^{\Sigma_i}\|_{\hat{Z} \to Z} \lesssim (\#\Sigma_i)^{-s_{\max}}$, where $s_{\max}$ cannot be improved by another selection of such a $\hat{Z}$ or $(\Sigma_i)_i$.

Example 4.2. If Ω is a domain in $\mathbb{R}^n$, $H = L_2(\Omega)$, $V = H^m(\Omega)$ or $V = H_0^m(\Omega)$, and Σ is a standard (isotropic) wavelet basis for $Z \in \{H, V\}$ of order d, i.e., of degree $d-1$, then, similar as with finite element approximation, $s_{\max}(\Sigma; H) = \frac{d}{n}$ and $s_{\max}(\Sigma; V) = \frac{d-m}{n}$. In both cases, $\hat{Z}$ can be taken to be $H^d(\Omega) \cap V$. The fact that $s_{\max}$ decreases with n, in particular that it is proportional to $\frac{1}{n}$ is known as the "curse of dimensionality".

If, additionally, $m \in \mathbb{N}_0$, and Ω is a product domain, say $\Omega = (0,1)^n$, then the space Z itself is an (intersection of) n-fold tensor products of spaces of univariate functions, and the curse of dimensionality can be avoided. Indeed, with $\Sigma^{(1)}$ being a Riesz basis for $L_2(0,1)$ of order d, the collection Σ defined as the n-fold tensor product of $\Sigma^{(1)}$, is a Riesz basis for $L_2((0,1)^n)$. If, additionally, when normalized in $H^m(0,1)$, $\Sigma^{(1)}$ is a Riesz basis for $H^m(0,1)$ or $H_0^m(0,1)$, then, when normalized in $H^m((0,1)^n)$, Σ is a Riesz basis for $H^m((0,1)^n)$ or $H_0^m((0,1)^n)$ (cf. [DS10a]).

For this tensor product basis Σ, it holds that $s_{\max}(\Sigma, V) = d - m$, assuming $m > 0$, and $s_{\max}(\Sigma, H) = d$ up to some "log-factors". More precisely, with a suitable trial space of dimension N, the error in H in the corresponding biorthogonal projection is of order

$$N^{-d} (\log N)^{(n-1)(\frac{1}{2}+d)} \tag{4.1}$$

(cf. discussion below). The sequence of (spans of) subsets of Σ used for demonstrating these linear approximation rates are known as (optimized) sparse grid spaces (cf.

[GK00, Dij09]), and $\hat{Z}$ can be taken to be (the intersection of V with) the n-fold tensor product of $H^d(0,1)$ (actually a "slightly" larger space can be used).

For avoiding the curse of dimensionality on non-product domains, currently the use of piecewise tensor approximations is under investigation.

For the moment assuming an exact geometric convergence rate from Σ in Z, i.e., excluding the situation (4.1), similar to $s_{\max}(\Sigma, Z)$, let $s_{\max}(\Theta, L_2(\mathrm{I}))$ denote the best possible approximation rate from Θ in $L_2(\mathrm{I})$. In view of the one-dimensional domain and the $L_2(\mathrm{I})$-norm, the latter rate is equal to the order of the wavelets. Let $(\Theta_p)_{p\geq 1}$ be the corresponding nested sequence of subsets of Θ with $\#\Theta_p \eqsim \rho(\Theta)^p$ for some constant $\rho(\Theta) > 1$, and let $\widehat{L_2(\mathrm{I})}$ denote the corresponding densely embedded subspace of $L_2(\mathrm{I})$.

Let us first consider the case that $s_{\max}(\Theta, L_2(\mathrm{I})) < s_{\max}(\Sigma, Z)$. Then, with $\Sigma_0 := \Theta_0 := \emptyset$, using that $\Theta \otimes \Sigma$ is a Riesz basis for $L_2(\mathrm{I}; Z) \eqsim L_2(\mathrm{I}) \otimes Z$, one verifies that (e.g. cf. [BG04, Dij09])

$$\begin{aligned} &\|I - \sum_{p+\alpha q\leq i} (Q^{\Theta_p} - Q^{\Theta_{p-1}}) \otimes (Q^{\Sigma_q} - Q^{\Sigma_{q-1}})\|_{\widehat{L_2(\mathrm{I})}\otimes \hat{Z}\to L_2(\mathrm{I})\otimes Z} \\ &\lesssim \sqrt{\sum_{p+\alpha q>i} \|Q^{\Theta_p} - Q^{\Theta_{p-1}}\|^2_{\widehat{L_2(\mathrm{I})}\to L_2(\mathrm{I})} \|Q^{\Sigma_q} - Q^{\Sigma_{q-1}}\|^2_{\hat{Z}\to Z}} \lesssim (\#\Theta_i)^{-s_{\max}(\Theta, L_2(\mathrm{I}))} \end{aligned} \tag{4.2}$$

if $\alpha \in (0, \frac{s_{\max}(\Sigma,Z)}{s_{\max}(\Theta,L_2(\mathrm{I}))} \frac{\log(\rho(\Sigma))}{\log(\rho(\Theta))})$. The map $\sum_{p+\alpha q\leq i}(Q^{\Theta_p} - Q^{\Theta_{p-1}}) \otimes (Q^{\Sigma_q} - Q^{\Sigma_{q-1}})$ is the biorthogonal projector onto the span of $\cup_{p+\alpha q\leq i}(\Theta_p\backslash\Theta_{p-1}) \otimes (\Sigma_q\backslash\Sigma_{q-1}) \subset \Theta\otimes\Sigma$. Taking $\alpha > \frac{\log(\rho(\Sigma))}{\log(\rho(\Theta))}$, the cardinality of this collection is $\eqsim \#\Theta_i$, and we conclude that $s_{\max}(\Theta \otimes \Sigma, L_2(\mathrm{I}; Z)) = s_{\max}(\Theta, L_2(\mathrm{I}))$.

Secondly, if $s_{\max}(\Theta, L_2(\mathrm{I})) > s_{\max}(\Sigma, Z)$, then $s_{\max}(\Theta\otimes\Sigma, L_2(\mathrm{I}; Z)) = s_{\max}(\Sigma, Z)$.

Thirdly, when $s_{\max}(\Theta, L_2(\mathrm{I})) = s_{\max}(\Sigma, Z)$, by taking $\alpha = \frac{\log(\rho(\Sigma))}{\log(\rho(\Theta))}$, the right hand side of (4.2) reads as $(\log \#\Theta_i)^{\frac{1}{2}}(\#\Theta_i)^{-s_{\max}(\Theta,L_2(\mathrm{I}))}$, whereas in this case we have $\#\cup_{p+\alpha q\leq i}(\Theta_p\backslash\Theta_{p-1}) \otimes (\Sigma_q\backslash\Sigma_{q-1}) \eqsim (\log \#\Theta_i)\#\Theta_i$. That is, with a trial space of dimension N, the error in $L_2(\mathrm{I}; Z)$ in the corresponding biorthogonal projection is bounded by a multiple of

$$N^{-s_{\max}(\Theta,L_2(\mathrm{I}))}(\log N)^{\frac{1}{2}+s_{\max}(\Theta,L_2(\mathrm{I}))}. \tag{4.3}$$

Finally, we discuss the situation from Example 4.2 of Σ being the n-fold tensor product of a univariate wavelet collection $\Sigma^{(1)}$ of order d, $Z = H$ so that (4.1) applies, and $s_{\max}(\Theta, L_2(\mathrm{I})) = d$, i.e., the order of the temporal wavelets being equal to that of the univariate spatial wavelets. Then an easy generalization of the analysis that led to (4.3) shows that with a trial space of dimension N, the error in $L_2(\mathrm{I}; H)$ in the corresponding biorthogonal projection is bounded by a multiple of

$$N^{-s_{\max}(\Theta,L_2(\mathrm{I}))}(\log N)^{n(\frac{1}{2}+s_{\max}(\Theta,L_2(\mathrm{I})))}. \tag{4.4}$$

As in all previous cases, this result is generally the best possible.

Summarizing, since in all cases, $s_{\max}(\Theta \otimes \Sigma, L_2(\mathrm{I}; Z))$ is essentially equal to $\min(s_{\max}(\Theta, L_2(\mathrm{I})), s_{\max}(\Sigma, Z))$, we can say that when $s_{\max}(\Theta, L_2(\mathrm{I})) \geq s_{\max}(\Sigma, Z)$, thanks to the use of a tensor product basis of temporal and spatial wavelets, the solution of the parabolic problem can be *approximated* with essentially the same rate as the solution of the corresponding stationary elliptic problem, in any case when these

solutions are sufficiently smooth. With the use of the adaptive wavelet schemes, the same holds true for the complexity of *solving* both problems.

5. A sparse stiffness matrix and the adaptive wavelet-Galerkin method. The application of an adaptive wavelet scheme to (3.3) requires a recurrent application of $\mathbf{B}^\top\mathbf{B}$ to finitely supported vectors. Generally, each row and column of $\mathbf{B}$ has infinitely many non-zero entries, meaning that the application of $\mathbf{B}^\top\mathbf{B}$ has to be approximated. Let us assume that the wavelets from the collections $\Theta^{\mathcal{X}}$, $\Sigma^{\mathcal{X}}$, $\Theta^{\mathcal{Y}}$ and $\Sigma^{\mathcal{Y}}$ are sufficiently smooth, have sufficiently many vanishing moments, and let the bilinear form $a(t;\cdot,\cdot)$ stem from a partial differential operator with sufficiently smooth coefficients, where, in all cases, "sufficient" has to be related to the best possible rate that can be expected. It has been verified (cf. [Ste09] + references cited there) that then the sizes of the entries of $\mathbf{B}$ decay sufficiently fast away from the diagonal, so that an adaptive routine can be designed that approximates the application of $\mathbf{B}^\top\mathbf{B}$ to a finitely supported vector within a prescribed tolerance, with which the overall adaptive wavelet scheme is quasi-optimal.

Although qualitatively satisfactory, numerical experiments showed that quantitatively the application of this approximate matrix-vector routine, commonly called the **apply**-routine, is quite demanding, where, moreover, this routine is not easy to implement.

Therefore, continuing earlier investigations for elliptic problems in [DS10b], we will design wavelet bases such that for

parabolic problems of second order with constant coefficients and with spatial domains of product type, the matrix $\mathbf{B}$ *and thus* $\mathbf{B}^\top$ *are (truly) sparse,*

and thus can be applied *exactly* to any finitely supported vector in linear complexity.

Remark 5.1. For parabolic PDEs with smooth, non-constant coefficients, the additional non-zero entries outside the sparsity pattern of a constant coefficient operator will be much smaller, depending on the levels of the wavelets involved. For the residual computation inside the adaptive wavelet scheme, which is quantitatively the most demanding part, it can be envisaged that in each column additional non-zero entries corresponding to wavelets on higher levels can be ignored.

Next we briefly describe the **AWGM** for solving $\mathbf{B}^\top\mathbf{B}\mathbf{u} = \mathbf{B}^\top\mathbf{f}$. We denote the index set of the Riesz basis $\Psi^{\mathcal{X}} = \Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}$ as ∇, so that $\mathbf{B}^\top\mathbf{B}$ is a boundedly invertible mapping on $\ell_2(\nabla)$, and $\mathbf{B}^\top\mathbf{f} \in \ell_2(\nabla)$.

To relate to the results derived in the previous section, we make the obvious observation that, since $\Psi^{\mathcal{X}}$ is a Riesz basis for $\mathcal{X}$, for any $\Lambda \subset \nabla$,

$$\inf_{v\in\operatorname{span}\{\psi^{\mathcal{X}}_\lambda:\lambda\in\Lambda\}} \|u-v\|_{\mathcal{X}} \eqsim \inf_{\{\mathbf{v}\in\ell_2(\nabla):\operatorname{supp}\mathbf{v}\subset\Lambda\}} \|\mathbf{u}-\mathbf{v}\| = \|\mathbf{u}-\mathbf{u}|_\Lambda\|,$$

where $\|\cdot\| := \|\cdot\|_{\ell_2(\nabla)}$. As a consequence, a quasi-best choice for an approximation to u as a linear combination of N elements from $\Psi^{\mathcal{X}}$ is to take $\mathbf{u}_N^\top\Psi^{\mathcal{X}}$, where $\mathbf{u}_N$ is a *best N-term approximation* for $\mathbf{u}$, i.e., a vector with support length less than or equal to N that coincides to $\mathbf{u}$ on those positions where the latter has its N largest coefficients in modulus.

In the "idealized" Adaptive Wavelet-Galerkin Method, a sequence $(\Lambda_i)_i \subset \nabla$ and a sequence of approximations $(\mathbf{u}^{(i)})_i$ to $\mathbf{u}$ with $\operatorname{supp}\mathbf{u}^{(i)} \subset \Lambda_i$ are created in the following way: $\Lambda_0 := \emptyset$, and $\mathbf{u}^{(0)} := 0$; for some constant $\mu \in (0,1]$, for $i \geq 0$, $\Lambda_{i+1} \supset \Lambda_i$ is the smallest set such that, with $\mathbf{P}_\Lambda$ denoting the restriction of a vector in $\ell_2(\nabla)$ to the indices in $\Lambda \subset \nabla$,

$$\|\mathbf{P}_{\Lambda_{i+1}}\mathbf{B}^\top(\mathbf{f}-\mathbf{B}\mathbf{u}^{(i)})\|_{\ell_2(\Lambda_{i+1})} \geq \mu\|\mathbf{B}^\top(\mathbf{f}-\mathbf{B}\mathbf{u}^{(i)})\|_{\ell_2(\nabla)};$$

$\mathbf{u}^{(i+1)}$ is determined as the solution of

$$(\mathbf{B}^\top\mathbf{B})|_{\Lambda_{i+1}\times\Lambda_{i+1}}\mathbf{u}^{(i+1)} = (\mathbf{B}^\top\mathbf{f})|_{\Lambda_{i+1}}.$$

It is known ([Ste09, §4.1]) that this $(\mathbf{u}^{(i)})_i$ converges linearly to $\mathbf{u}$, and that, for $\mu < \kappa(\mathbf{B}^\top\mathbf{B})^{-\frac{1}{2}}$, if, for whatever $s > 0$,

$$\mathbf{u} \in \mathcal{A}^s := \{\mathbf{v} \in \ell_2(\nabla) : |\mathbf{v}|_{\mathcal{A}^s} := \sup_{N\in\mathbb{N}_0} N^s\|\mathbf{v}-\mathbf{v}_N\| < \infty\},$$

then $\#\operatorname{supp}\mathbf{u}^{(i+1)} \lesssim \|\mathbf{u}-\mathbf{u}^{(i)}\|^{-1/s}|\mathbf{u}|_{\mathcal{A}^s}^{1/s}$.

The above "idealized" method cannot be implemented since generally $\mathbf{f}$ has infinite support (although usually also each column of $\mathbf{B}^\top$ and $\mathbf{B}$ has infinite support, in the situation of this work, $\mathbf{B}^\top$ and $\mathbf{B}$ can be applied exactly to any finitely supported vector at linear cost). Moreover, the aim is to have a method of optimal *computational* complexity. Therefore in the practical Adaptive Wavelet-Galerkin Method **AWGM** given below, residuals $\mathbf{B}^\top(\mathbf{f}-\mathbf{B}\mathbf{u}^{(i)})$ are computed only *inexactly* up to some sufficiently small relative tolerance (parameter δ); Λ_{i+1} is determined such that $\#(\Lambda_{i+1}\backslash\Lambda_i)$ is only minimal up to some absolute multiple; the Galerkin systems are solved only *inexactly* up to some sufficiently small relative tolerance (parameter γ) .

We will assume availability of the following routine **RHS**. In our examples this assumption is verified easily.

$\mathbf{RHS}[\varepsilon] \to \mathbf{f}_\varepsilon$:
% Input: $\varepsilon > 0$. *Output: a finitely supported* $\mathbf{f}_\varepsilon$ *with*

$$\|\mathbf{f}-\mathbf{f}_\varepsilon\| \le \varepsilon \quad \textit{and} \quad \#\operatorname{supp}\mathbf{f}_\varepsilon \lesssim \min\{N : \|\mathbf{f}-\mathbf{f}_N\| \le \varepsilon\},$$

% taking a number of operations that is bounded by some absolute multiple of
% $\#\operatorname{supp}\mathbf{f}_\varepsilon + 1$.

Since a call of **RHS** yields an approximation to $\mathbf{f}$ within some absolute tolerance, whereas inside the adaptive wavelet-Galerkin method an approximation of the residual is needed within some sufficiently small relative tolerance, the **AWGM** contains an inner loop in which an absolute tolerance for $\mathbf{f}$ is determined that yields the desired relative tolerance for the residual. With a suitably chosen initial absolute tolerance (determined by parameter θ), it usually terminates after one or two iterations.

The **AWGM** reads as follows:

$\mathbf{AWGM}[\varepsilon, \varepsilon_{-1}] \to \mathbf{u}_\varepsilon$:
% Input: $\varepsilon, \varepsilon_{-1} > 0$, *with* ε *being the required upper bound for* $\|\mathbf{B}^\top(\mathbf{f}-\mathbf{B}\mathbf{u}_\varepsilon)\|$,
% and ε_{-1} *an estimate for the initial residual* $\|\mathbf{B}^\top\mathbf{f}\|$.
% Parameters: $\mu, \delta, \gamma, \theta$ *such that* $\delta \in (0,\mu)$, $\frac{\mu+\delta}{1-\delta} < \kappa(\mathbf{B}^\top\mathbf{B})^{-\frac{1}{2}}$, $\theta > 0$ *and*
% $\gamma \in \big(0, \frac{(1-\delta)(\mu-\delta)}{1+\delta}\kappa(\mathbf{B}^\top\mathbf{B})^{-1}\big)$.

```
i := 0, u^(i) := 0, Λ_i := ∅
do ζ := θε_{i-1}
    do ζ := ζ/2, r^(i) := B^⊤(RHS[ζ] − Bu^(i))
        if ε_i := ‖r^(i)‖ + ζ‖B‖ ≤ ε then u_ε := u^(i) stop endif
    until ζ‖B‖ ≤ δ‖r^(i)‖
    Λ_{i+1} := EXPAND[Λ_i, r^(i), μ‖r^(i)‖]
    u^(i+1) := GALERKIN[Λ_{i+1}, u^(i), (B^⊤RHS[γε_i/‖B‖])|_{Λ_{i+1}}, (1+γ)ε_i, γε_i]
    i := i + 1
enddo
```

By a call of **GALERKIN**, the Galerkin system corresponding to the current active wavelet index set is solved within some prescribed tolerance. The previous approximate Galerkin solution is used as a starting value for an iterative solver.

GALERKIN$[\Lambda, \mathbf{w}, \mathbf{g}, \delta, \varepsilon] \to \bar{\mathbf{w}}$:
% Input: $\varepsilon, \delta > 0$, $\Lambda \subset \nabla$, $\mathbf{w}, \mathbf{g} \in \ell_2(\Lambda)$ *with* $\|\mathbf{g} - (\mathbf{B}^\top \mathbf{B})|_{\Lambda\times\Lambda}\mathbf{w}\| \leq \delta$.
% Output: $\bar{\mathbf{w}} \in \ell_2(\Lambda)$ *with* $\|\mathbf{g} - (\mathbf{B}^\top \mathbf{B})|_{\Lambda\times\Lambda}\bar{\mathbf{w}}\| \leq \varepsilon$ *in* $\mathcal{O}(\log(\delta/\varepsilon)\#\Lambda)$ *operations.*

Using that $(\mathbf{B}^\top \mathbf{B})|_{\Lambda\times\Lambda}$ is well-conditioned, uniformly in Λ, this routine can be implemented as a Conjugate Residual iteration with starting vector $\mathbf{w}$.

By a call of **EXPAND**, the current active wavelet index set is expanded with those indices where the current residual has its largest values.

EXPAND$[\Lambda, \mathbf{g}, \sigma] \to \bar{\Lambda}$:
% Input: $\Lambda \subset \nabla$, *a finitely supported* $\mathbf{g} \in \ell_2(\nabla)$, *and a scalar* $\sigma \in [0, \|\mathbf{g}\|_{\ell_2(\nabla)}]$.
% Output: $\Lambda \subset \bar{\Lambda} \subset \nabla$ *with* $\|\mathbf{P}_{\bar{\Lambda}}\mathbf{g}\| \geq \sigma$ *and such that, up to some absolute multiple,*
% $\#(\bar{\Lambda} \setminus \Lambda)$ *is minimal over all such* $\bar{\Lambda}$, *and the cost of the call is* $\mathcal{O}(\#\Lambda \cup \operatorname{supp} \mathbf{g})$.

Noting that $\|\mathbf{P}_{\bar{\Lambda}}\mathbf{g}\| \geq \sigma$ is equivalent to $\|\mathbf{P}_{\bar{\Lambda}\setminus\Lambda}\mathbf{g}\|^2 \geq \sigma^2 - \|\mathbf{P}_\Lambda \mathbf{g}\|^2$, a set $\bar{\Lambda}$ with truly minimal $\#(\bar{\Lambda} \setminus \Lambda)$ is found by ordering $\{\mathbf{g}_\lambda : \lambda \in \bar{\Lambda} \setminus \Lambda\}$ by non-increasing modulus, and then by selecting elements from the head to the tail of this ordered sequence until the criterion is met. The loglinear complexity of such an implementation can be reduced to a linear complexity by performing an approximate sorting, at the expense of obtaining a $\bar{\Lambda}$ for which $\#(\bar{\Lambda} \setminus \Lambda)$ is only minimal up to some absolute factor.

The following result states that the **AWGM** is quasi-optimal.

THEOREM 5.2 ([Ste09, Thm.4.1]). *For* $\mathbf{u}_\varepsilon := \mathbf{AWGM}[\varepsilon, \varepsilon_{-1}]$, *it holds that* $\|\mathbf{B}^\top(\mathbf{f} - \mathbf{B}\mathbf{u}_\varepsilon)\| \leq \varepsilon$. *If for some* $s > 0$, $\mathbf{u} \in \mathcal{A}^s$, *then both* $\#\operatorname{supp} \mathbf{u}_\varepsilon$ *and, assuming* $\varepsilon \lesssim \varepsilon_{-1} \eqsim \|\mathbf{B}^\top \mathbf{f}\|$, *the number of operations used by the call are bounded by absolute multiples of* $\varepsilon^{-1/s}|\mathbf{u}|_{\mathcal{A}^s}^{1/s}$.

Remark 5.3. In Theorem 5.2, it is assumed that the best N-term approximations to $\mathbf{u}$ converge exactly algebraically. In view of (4.4), we note that if for some increasing function $G : [0, \infty) \to [0, \infty)$ with, for $\rho \in (0, 1)$, $\sum_{k=0}^\infty G^{-1}(\rho^k \delta^{-1}) \lesssim G^{-1}(\delta^{-1})$, $|\mathbf{u}|_{\mathcal{A}(G)} := \sup_{N\in\mathbb{N}_0} G(N)\|\mathbf{u} - \mathbf{u}_N\| < \infty$ (uniformly in $\delta > 0$), then a direct generalization of Theorem 5.2 shows that both $\#\operatorname{supp} \mathbf{u}_\varepsilon$ and the number of operations needed are bounded by absolute multiples of $G^{-1}(\varepsilon^{-1}|\mathbf{u}|_{\mathcal{A}(G)})$.

For *constructing* sparse $\mathbf{B}$ and thus $\mathbf{B}^\top$, recall (3.6) and (3.7), and consider $a_1(\cdot) = 1$. For $\Omega = (0, 1)^n$, $H = L_2(\Omega)$, $V = H_0^1(\Omega)$, $W = H_0^1(\Omega) \cap H^2(\Omega)$, and $a_2(w, v) = \int_\Omega \sum_{|\alpha|,|\beta|\leq 1} a_{\alpha,\beta}\partial^\alpha w \partial^\beta v$ with constant coefficients $a_{\alpha,\beta}$, in [DS10b] a univariate wavelet collection $\Sigma^{(1)}$ of order 4 has been constructed, consisting of cubic Hermite splines, such that its n-fold tensor product $\Sigma = \Sigma^{\mathcal{X}} = \Sigma^{\mathcal{Y}}$, normalized in the corresponding norms, is a Riesz basis for H and for W, and such that $a_2(\Sigma, \Sigma)$ and $\langle \Sigma, \Sigma\rangle_{L_2(\Omega)}$ are sparse.

In [CS10], along similar lines we construct such a collection $\Sigma^{(1)}$ of order 5, consisting of quartic Hermite splines, which has better quantitative properties than that of [DS10b], as well as locally supported duals.

The dual wavelets corresponding to $\Sigma^{(1)}$ from both [DS10b] and [CS10] are *discontinuous*, and so the resulting Σ does *not* generate a basis for $V = H^{-1}(\Omega)$. So to transform the parabolic problem with these wavelets into an equivalent, well-posed bi-infinite matrix-vector problem, we *cannot* apply Theorem 2.2, and therefore we apply Theorem 2.4 instead. So in the remainder of this paper, we take

$$\mathcal{X} = L_2(\mathrm{I}; H) \quad \text{and} \quad \mathcal{Y} = L_2(\mathrm{I}; W) \cap H^1_{0,\{T\}}(\mathrm{I}; H).$$

In the next section, we will construct collections $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ such that $\Theta^{\mathcal{X}}$ is a Riesz basis for $L_2(\mathrm{I})$, $\Theta^{\mathcal{Y}}$ is Riesz basis for $L_2(\mathrm{I})$ and, renormalized, for $H^1_{0,\{T\}}(\mathrm{I})$, and such that $\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ and $\langle \Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ are sparse. Combined with the sparsity of $a_2(\Sigma,\Sigma)$ and $\langle \Sigma, \Sigma\rangle_{L_2(\Omega)}$, we conclude that with respect to the test and trial wavelet collections $\Theta^{\mathcal{X}} \otimes \Sigma^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}} \otimes \Sigma^{\mathcal{Y}}$ the system matrix $\mathbf{B}$, and so also its transpose, is sparse.

Remark 5.4. Even when $T = 1$, it is not possible simply to take $\Theta^{\mathcal{X}} = \Theta^{\mathcal{Y}} = \Sigma^{(1)}$ because of the homogeneous boundary conditions incorporated in $\Sigma^{(1)}$ at *both* boundary points. Indeed, since $\Theta^{\mathcal{Y}}$ has to be a basis for $H^1_{0,\{T\}}(\mathrm{I})$, not all of its elements can vanish at 0.

In view of the order of the aforementioned spatial wavelets from [CS10], and the considerations about the best possible rates in $\mathcal{X}$ from the previous section, we will construct $\Theta^{\mathcal{X}}$ of order 5.

6. Construction of the temporal trial and test wavelets.

6.1. Necessary conditions on the trial wavelets. We will search collections of univariate wavelets $\Theta^{\mathcal{Z}} = \{\theta^{\mathcal{Z}}_\lambda \colon \lambda \in \nabla^{\mathcal{Z}}\}$ ($\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$), such that, with $|\lambda| \in \mathbb{N}_0$ denoting the level of $\theta^{\mathcal{Z}}_\lambda$ or that of λ,

($1^{\mathcal{Z}}$) $\operatorname{diam}\operatorname{supp}\theta^{\mathcal{Z}}_\lambda \lesssim 2^{-|\lambda|}$,

($2^{\mathcal{Z}}$) $\sup_{j,k\in\mathbb{N}_0} \#\{|\lambda| = j \colon [k2^{-j}, (k+1)2^{-j}] \cap \operatorname{supp}\theta^{\mathcal{Z}}_\lambda \neq \emptyset\} < \infty$,

($3^{\mathcal{Z}}$) $\Theta^{\mathcal{Z}}$ is Riesz basis for $L_2(\mathrm{I})$,

(4) $\{\theta^{\mathcal{Y}}_\lambda / \|\dot{\theta}^{\mathcal{Y}}_\lambda\|_{L_2(\mathrm{I})} \colon \lambda \in \nabla^{\mathcal{Y}}\}$ is a Riesz basis for $H^1_{0,\{T\}}(\mathrm{I})$,

(5) $\int_{\mathrm{I}} \theta^{\mathcal{X}}_\mu \dot{\theta}^{\mathcal{Y}}_\lambda = 0$ when $||\lambda| - |\mu|| > M$,

(6) $\int_{\mathrm{I}} \theta^{\mathcal{X}}_\mu \theta^{\mathcal{Y}}_\lambda = 0$ when $||\lambda| - |\mu|| > M$,

where $M \in N_0$ is some constant, that later will be chosen to be 1. As a consequence, with respect to a level-wise partition of the wavelets, $\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ and $\langle \Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ will be block tridiagonal with, because of ($1^{\mathcal{Z}}$) and ($2^{\mathcal{Z}}$) ($\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$), sparse non-zero blocks. Note that under the assumptions ($1^{\mathcal{Z}}$) and ($2^{\mathcal{Z}}$) ($\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$), $\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ and $\langle \Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}$ are sparse if and only if (5) and (6), respectively, are valid. We will refer to the combined properties ($1^{\mathcal{Z}}$) and ($2^{\mathcal{Z}}$) by saying that the wavelets $\{\theta^{\mathcal{Z}}_\lambda : \lambda \in \nabla^{\mathcal{Z}}\}$ are *(uniformly) local*.

The next proposition shows that essentially above conditions can only be fulfilled for a collection $\Theta^{\mathcal{X}}$ of functions that are continuous and vanish at zero. The proof is similar to that of [DS10b, Prop.1].

PROPOSITION 6.1. *If, in addition to* ($1^{\mathcal{Y}}$), ($2^{\mathcal{Y}}$), ($3^{\mathcal{Y}}$), (4) *and* (5), *each wavelet* $\theta^{\mathcal{X}}_\mu$ *is piecewise smooth with a bounded piecewise derivative, then necessarily* $\Theta^{\mathcal{X}} \subset C(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I})$.

In view of the fact that, for $u_0 \neq 0$, the solution of the parabolic problem does *not* identically vanish at $t = 0$, the condition that all elements of $\Theta^{\mathcal{X}}$ vanish at zero is a price that has to be paid for getting sparse matrices. We will construct $\Theta^{\mathcal{X}}$ such that also its corresponding dual collection is uniformly local. As we will see in §6.7, thanks to that, the (strongly) reduced approximation order of $\Theta^{\mathcal{X}}$ locally at zero can be compensated by adding wavelets on higher levels with supports near zero, so that the overall approximation order is not negatively affected. The adaptive algorithm will take care of this "automatically".

6.2. Biorthogonal multiresolution analyses and wavelets. In order to construct wavelet collections $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ that, properly scaled, generate Riesz bases for a range of Sobolev spaces, we will use the following well-known theorem (cf.

[Dah96, DS99, Coh03]).

THEOREM 6.2 (Biorthogonal space decompositions). *For $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$, let*

$$V_0^{\mathcal{Z}} \subset V_1^{\mathcal{Z}} \subset \cdots \subset L_2(\mathrm{I}), \qquad \tilde{V}_0^{\mathcal{Z}} \subset \tilde{V}_1^{\mathcal{Z}} \subset \cdots \subset L_2(\mathrm{I})$$

be sequences of primal and dual spaces such that

$$\dim V_j^{\mathcal{Z}} = \dim \tilde{V}_j^{\mathcal{Z}} < \infty \quad \textit{and} \quad \inf_{j\in\mathbb{N}_0} \inf_{0\neq \tilde{v}_j \in \tilde{V}_j^{\mathcal{Z}}} \sup_{0\neq v_j \in V_j^{\mathcal{Z}}} \frac{|\langle \tilde{v}_j, v_j\rangle_{L_2(\mathrm{I})}|}{\|\tilde{v}_j\|_{L_2(\mathrm{I})}\|v_j\|_{L_2(\mathrm{I})}} > 0. \quad (6.1)$$

In addition, for some $0 < \gamma_{\mathcal{Z}} < d_{\mathcal{Z}}$, let

$$\inf_{v_j\in V_j^{\mathcal{Z}}} \|v - v_j\|_{L_2(\mathrm{I})} \lesssim 2^{-jd_{\mathcal{Z}}}\|v\|_{\mathcal{H}_{\mathcal{Z}}^{d_{\mathcal{Z}}}(\mathrm{I})} \quad (v \in \mathcal{H}_{\mathcal{Z}}^{d_{\mathcal{Z}}}(\mathrm{I})) \quad \text{(Jackson estimate), } \textit{and}$$

$$\|v_j\|_{\mathcal{H}_{\mathcal{Z}}^{s}(\mathrm{I})} \lesssim 2^{js}\|v_j\|_{L_2(\mathrm{I})} \quad (v_j \in V_j^{\mathcal{Z}},\ s \in [0,\gamma_{\mathcal{Z}})) \quad \text{(Bernstein estimate),}$$

where for a Hilbert space $\mathcal{H}_{\mathcal{Z}}^{d_{\mathcal{Z}}}(\mathrm{I}) \hookrightarrow L_2(\mathrm{I})$, $\mathcal{H}_{\mathcal{Z}}^{s}(\mathrm{I}) := [L_2(\mathrm{I}), \mathcal{H}_{\mathcal{Z}}^{d_{\mathcal{Z}}}(\mathrm{I})]_{s/d_{\mathcal{Z}}}$ $(s \in [0, d_{\mathcal{Z}}])$, and let similar estimates be valid at the dual side with $((V_j^{\mathcal{Z}})_j, d_{\mathcal{Z}}, \gamma_{\mathcal{Z}}, \mathcal{H}_{\mathcal{Z}}^{s}(\mathrm{I}))$ reading as $((\tilde{V}_j^{\mathcal{Z}})_j, \tilde{d}_{\mathcal{Z}}, \tilde{\gamma}_{\mathcal{Z}}, \tilde{\mathcal{H}}_{\mathcal{Z}}^{s}(\mathrm{I}))$.

Then, with $\Phi_0^{\mathcal{Z}} = \{\phi_{0,k}^{\mathcal{Z}} \colon k \in I_0^{\mathcal{Z}}\}$ being a basis for $V_0^{\mathcal{Z}}$ (scaling functions) and $\Theta_j^{\mathcal{Z}} = \{\theta_{j,k}^{\mathcal{Z}} \colon k \in J_j^{\mathcal{Z}}\}$ being $L_2(\mathrm{I})$-Riesz bases for $W_j^{\mathcal{Z}} := V_j^{\mathcal{Z}} \cap (\tilde{V}_{j-1}^{\mathcal{Z}})^{\perp_{L_2(\mathrm{I})}}$ uniformly in $j \in \mathbb{N}$ (wavelets), for $s \in (-\tilde{\gamma}_{\mathcal{Z}}, \gamma_{\mathcal{Z}})$ the collection of properly scaled primal biorthogonal wavelets

$$\Phi_0^{\mathcal{Z}} \cup \cup_{j\in\mathbb{N}} 2^{-sj}\Theta_j^{\mathcal{Z}}$$

is a Riesz basis for $\mathcal{H}_{\mathcal{Z}}^{s}(\mathrm{I})$, where $\mathcal{H}_{\mathcal{Z}}^{s}(\mathrm{I}) := (\tilde{\mathcal{H}}_{\mathcal{Z}}^{-s}(\mathrm{I}))'$ for $s < 0$.

In view of the notations introduced earlier, we denote (j,k) also as λ, where $|\lambda| = j$, $\phi_{0,k}^{\mathcal{Z}}$ as $\theta_{0,k}^{\mathcal{Z}}$, and $I_0^{\mathcal{Z}} \cup \cup_{j\in\mathbb{N}} J_j^{\mathcal{Z}}$ as $\nabla^{\mathcal{Z}}$.

The existence of the Riesz basis

$$\Theta^{\mathcal{Z}} := \Phi_0^{\mathcal{Z}} \cup \cup_{j\in\mathbb{N}} \Theta_j^{\mathcal{Z}}$$

for $L_2(\mathrm{I})$ as in Theorem 2.2 is equivalent to the existence of another, dual Riesz basis $\tilde{\Theta}^{\mathcal{Z}}$ for $L_2(\mathrm{I})$. Writing this basis correspondingly as $\tilde{\Phi}_0^{\mathcal{Z}} \cup \cup_{j\in\mathbb{N}} \tilde{\Theta}_j^{\mathcal{Z}}$, $\tilde{\Phi}_0^{\mathcal{Z}}$ is a basis for $\tilde{V}_0^{\mathcal{Z}}$, and $\tilde{\Theta}_j^{\mathcal{Z}}$ for $\tilde{V}_j^{\mathcal{Z}} \cap (V_{j-1}^{\mathcal{Z}})^{\perp_{L_2(\mathrm{I})}}$. For $s \in (-\gamma, \tilde{\gamma})$, the collection

$$\tilde{\Phi}_0^{\mathcal{Z}} \cup \cup_{j\in\mathbb{N}} 2^{-sj}\tilde{\Theta}_j^{\mathcal{Z}}$$

is a Riesz basis for $\tilde{\mathcal{H}}_{\mathcal{Z}}^{s}(\mathrm{I})$.

We will construct $V_j^{\mathcal{Z}}$ and $\tilde{V}_j^{\mathcal{Z}}$ such that they can be equipped with biorthogonal, uniformly local, uniform $L_2(\mathrm{I})$-Riesz bases. As shown in the next proposition, this will mean that the property (6.1) is satisfied, and moreover, that under a mild additional condition, uniformly local primal *and* dual wavelets become available. For a proof we refer to [CDP96, Ste03].

PROPOSITION 6.3. *Let $(V_j^{\mathcal{Z}})_j, (\tilde{V}_j^{\mathcal{Z}})_j \subset L_2(\mathrm{I})$ be nested sequences of finite dimensional spaces, and let $\Phi_j^{\mathcal{Z}}$ and $\tilde{\Phi}_j^{\mathcal{Z}}$ be biorthogonal, uniform $L_2(\mathrm{I})$-Riesz bases for $V_j^{\mathcal{Z}}$ and $\tilde{V}_j^{\mathcal{Z}}$, respectively.* (a). *Property* (6.1) *is satisfied.*

(b). *Let $\Xi_{j+1}^{\mathcal{Z}} \subset V_{j+1}^{\mathcal{Z}}$ ("initial stable completion") be such that $\Phi_j^{\mathcal{Z}} \cup \Xi_{j+1}^{\mathcal{Z}}$ is a uniform $L_2(\mathrm{I})$-Riesz basis for $V_{j+1}^{\mathcal{Z}}$, then*

$$\Theta_{j+1}^{\mathcal{Z}} := \Xi_{j+1}^{\mathcal{Z}} - \langle \Xi_{j+1}^{\mathcal{Z}}, \tilde{\Phi}_j^{\mathcal{Z}}\rangle_{L_2(\mathrm{I})}\Phi_j^{\mathcal{Z}}$$

is a uniform $L_2(\mathrm{I})$-Riesz basis for $V_{j+1}^{\mathcal{Z}} \cap (\tilde{V}_j^{\mathcal{Z}})^{\perp_{L_2(\mathrm{I})}}$. (Here, as on more places, we view a collection of functions formally as a column vector, and for pair $(\Sigma^{(1)}, \Sigma^{(2)})$ of such collections, $\langle \Sigma^{(1)}, \Sigma^{(2)} \rangle$ denotes the matrix $[\langle \sigma^{(1)}, \sigma^{(2)} \rangle]_{\sigma^{(1)} \in \Sigma^{(1)}, \sigma^{(2)} \in \Sigma^{(2)}}$.)

(c). *Writing* $\left[(\Phi_j^{\mathcal{Z}})^\top \quad (\Xi_{j+1}^{\mathcal{Z}})^\top\right] = (\Phi_{j+1}^{\mathcal{Z}})^\top \check{\mathbf{M}}_j$, $\check{\mathbf{M}}_j^{-1} = \begin{bmatrix} \mathbf{G}_{j,0} \\ \mathbf{G}_{j,1} \end{bmatrix}$, *i.e.,* $(\Phi_{j+1}^{\mathcal{Z}})^\top = (\Phi_j^{\mathcal{Z}})^\top \mathbf{G}_{j,0} + (\Xi_{j+1}^{\mathcal{Z}})^\top \mathbf{G}_{j,1}$, *then* $\tilde{\Theta}_{j+1}^{\mathcal{Z}} = \mathbf{G}_{j,1} \tilde{\Phi}_{j+1}^{\mathcal{Z}}$.

(d). *If $\Phi_j^{\mathcal{Z}}$, $\tilde{\Phi}_j^{\mathcal{Z}}$, and $\Xi_{j+1}^{\mathcal{Z}}$ are uniformly local, then $\Theta_{j+1}^{\mathcal{Z}}$ is uniformly local. If $\tilde{\Phi}_{j+1}^{\mathcal{Z}}$ is uniformly local and $\mathbf{G}_{j,1}$ is uniformly local, with which we mean that in the expansion of $\phi_{j+1,k}^{\mathcal{Z}}$ in terms of $\Phi_j^{\mathcal{Z}} \cup \Xi_{j+1}^{\mathcal{Z}}$ the coefficient in front of $\xi_{j+1,\ell}^{\mathcal{Z}}$ vanishes whenever* $\operatorname{dist}(\operatorname{supp} \phi_{j+1,k}^{\mathcal{Z}}, \operatorname{supp} \xi_{j+1,\ell}^{\mathcal{Z}}) \gtrsim 2^{-j}$, *then $\tilde{\Theta}_{j+1}^{\mathcal{Z}}$ is uniformly local.*

6.3. Biorthogonal multiresolution analyses that lead to sparse $\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}} \rangle$ and $\langle \Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}} \rangle$. For $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$, we will select biorthogonal multiresolution analyses $((V_j^{\mathcal{Z}})_j, (\tilde{V}_j^{\mathcal{Z}})_j)$ and corresponding uniformly local primal biorthogonal wavelets $\Theta^{\mathcal{Z}}$ as in Theorem 6.2 (with $(\gamma_{\mathcal{Z}}, d_{\mathcal{Z}}, \mathcal{H}_{\mathcal{Z}}^{d_{\mathcal{Z}}}, \tilde{\gamma}_{\mathcal{Z}}, \tilde{d}_{\mathcal{Z}}, \tilde{\mathcal{H}}_{\mathcal{Z}}^{\tilde{d}_{\mathcal{Z}}})$ that will be specified in (6.9) and (6.10)), with additionally

$$(V_j^{\mathcal{X}})_j \subset H^1_{0,\{0\}}(\mathrm{I}), \quad (V_j^{\mathcal{Y}})_j \subset H^1_{0,\{T\}}(\mathrm{I}), \tag{6.2}$$

$$V_j^{\mathcal{Y}} + \dot{V}_j^{\mathcal{Y}} \subset \tilde{V}_{j+1}^{\mathcal{X}}, \tag{6.3}$$

$$V_j^{\mathcal{X}} + \dot{V}_j^{\mathcal{X}} \subset \tilde{V}_{j+1}^{\mathcal{Y}}, \tag{6.4}$$

where a "dot" on top of a linear space of functions denotes the linear space of derivative functions. Then, thanks to (6.3) and (6.4), for $|\mu| > |\lambda| + 1$ we have that

$$0 = \langle \theta_\mu^{\mathcal{X}}, \theta_\lambda^{\mathcal{Y}} \rangle_{L_2(\mathrm{I})} \text{ and } 0 = \langle \theta_\mu^{\mathcal{X}}, \dot{\theta}_\lambda^{\mathcal{Y}} \rangle_{L_2(\mathrm{I})}, \tag{6.5}$$

and for $|\lambda| > |\mu| + 1$ that

$$0 = \langle \theta_\mu^{\mathcal{X}}, \theta_\lambda^{\mathcal{Y}} \rangle_{L_2(\mathrm{I})} \text{ and } 0 = \langle \dot{\theta}_\mu^{\mathcal{X}}, \theta_\lambda^{\mathcal{Y}} \rangle_{L_2(\mathrm{I})} = -\langle \theta_\mu^{\mathcal{X}}, \dot{\theta}_\lambda^{\mathcal{Y}} \rangle_{L_2(\mathrm{I})}, \tag{6.6}$$

respectively, where the last inequality follows from (6.2). Together, (6.5) and (6.6) are equivalent to (5) and (6) with $M = 1$.

6.4. A realization. We select the primal spaces as the continuous piecewise quartics with respect to dyadically refined partitions satisfying the appropriate boundary conditions, i.e.,

$$V_j^{\mathcal{Z}} = \sum_{k=0}^{2^{j+1}-1} P_4(k2^{-(j+1)}T, (k+1)2^{-(j+1)}T) \cap C^0(\mathrm{I}) \cap \begin{cases} H^1_{0,\{0\}}(\mathrm{I}) & \text{when } \mathcal{Z} = \mathcal{X}, \\ H^1_{0,\{T\}}(\mathrm{I}) & \text{when } \mathcal{Z} = \mathcal{Y}. \end{cases}$$

For $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$, we have

$$V_j^{\mathcal{Z}} + \dot{V}_j^{\mathcal{Z}} \subset \sum_{k=0}^{2^{j+1}-1} P_4(k2^{-(j+1)}T, (k+1)2^{-(j+1)}T) \tag{6.7}$$

(actually even equality holds), the latter space being of dimension $5 \cdot 2^{j+1}$. We will select $\tilde{V}_{j+1}^{\mathcal{Z}}$ as an extension of this space to a space of dimension $8 \cdot 2^{j+1}$, being the dimension of $V_{j+1}^{\mathcal{Z}}$. Care has to be taken that the dual spaces are nested.

We start with a construction on the "reference macro element" $(-1,1)$. Let

$$V := P_4(-1,0) \times P_4(0,1) \cap C(-1,1), \quad \tilde{V}_1 := P_4(-1,1).$$

We performed the following steps:

- Determine V_1 as the orthogonal projection of $\tilde{V}_1$ onto $V \cap H^1_0(-1,1)$,
- determine $V_2 \perp \tilde{V}_1$ such that $V = V_1 \oplus V_2$,
- with $\tilde{V} := \tilde{V}_1 + V_2$, determine $\tilde{V}_2 \perp V_1$ such that $\tilde{V} = \tilde{V}_1 \oplus \tilde{V}_2$,
- equip both pairs $(V_1, \tilde{V}_1)$ and $(V_2, \tilde{V}_2)$ with biorthogonal bases.

Now the union of the bases for V_1 and V_2 and that for $\tilde{V}_1$ and $\tilde{V}_2$ are biorthogonal bases for V and $\tilde{V}$. Writing them as $\{\phi_0, \dots, \phi_8\}$ and $\{\tilde{\phi}_0, \dots, \tilde{\phi}_8\}$, respectively, by a suitable numbering we have $V_1 = \operatorname{span}\{\phi_2, \dots, \phi_6\}$ and $\tilde{V}_1 = \operatorname{span}\{\tilde{\phi}_2, \dots, \tilde{\phi}_6\}$. The bases can be organized such that for $k = 0, \dots, 8$,

$$\phi_{8-k}(x) = \phi_k(-x), \quad \tilde{\phi}_{8-k}(x) = \tilde{\phi}_k(-x),$$

and that ϕ_0 is the only primal basis function that does not vanish at -1 (and thus that ϕ_8 is the only primal basis function that does not vanish at 1).

It holds that $\tilde{V} = V$, but whereas $P_4(-1,1) = \operatorname{span}\{\tilde{\phi}_2, \dots, \tilde{\phi}_6\}$, at the primal side $P_4(-1,1) \not\subset \operatorname{span}\{\phi_1, \dots, \phi_7\}$ since the latter space is in $H^1_0(-1,1)$.

The biorthogonal basis functions $\{\phi_0, \dots, \phi_4\}$ and $\{\tilde{\phi}_0, \dots, \tilde{\phi}_4\}$ are illustrated in Figure 6.1, and their values at $\frac{1}{4}\mathbb{Z} \cap [-1,1]$ are given in Tables 6.1 and 6.2, respectively.

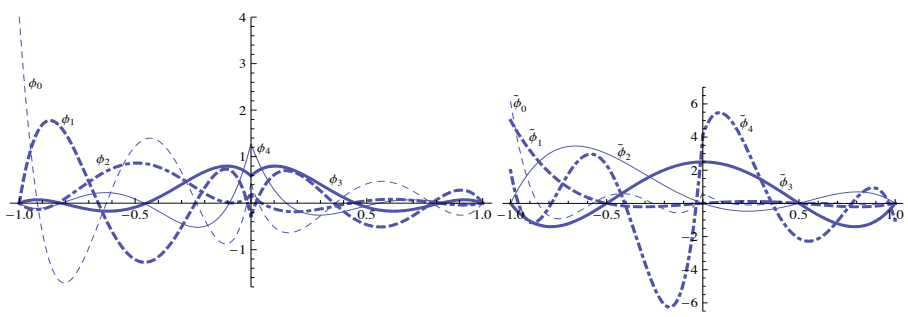

FIG. 6.1. *Biorthogonal basis functions $\{\phi_0, \dots, \phi_4\}$ and $\{\tilde{\phi}_0, \dots, \tilde{\phi}_4\}$ on the reference macro element.*

	-1	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
ϕ_0	4	$-\frac{48237}{32768}$	$\frac{7489}{6144}$	$\frac{5155}{32768}$	$\frac{55}{128}$	$-\frac{15197}{32768}$	$\frac{2561}{6144}$	$\frac{3347}{32768}$
ϕ_1	0	$\frac{8379}{8192}$	$-\frac{621}{512}$	$-\frac{693}{8192}$	$-\frac{9}{32}$	$\frac{3915}{8192}$	$-\frac{237}{512}$	$-\frac{837}{8192}$
ϕ_2	0	$\frac{294405}{1048576}$	$\frac{56525}{65536}$	$\frac{384885}{1048576}$	$\frac{705}{4096}$	$-\frac{171915}{1048576}$	$\frac{3405}{65536}$	$\frac{37125}{1048576}$
ϕ_3	0	$-\frac{2697}{32768}$	$-\frac{177}{2048}$	$\frac{17991}{32768}$	$\frac{75}{128}$	$\frac{17991}{32768}$	$-\frac{177}{2048}$	$-\frac{2697}{32768}$
ϕ_4	0	$\frac{5705}{49152}$	$\frac{91}{1024}$	$-\frac{8365}{16384}$	$\frac{245}{192}$	$-\frac{3885}{16384}$	$-\frac{175}{3072}$	$\frac{3017}{49152}$

TABLE 6.1
Values of $\{\phi_0, \dots, \phi_4\}$ at $\frac{1}{4}\mathbb{Z} \cap [-1,1)$

From the bases on the reference macro element, we construct global biorthogonal bases in the way known from finite elements. We start at the $\mathcal{X}$-side. With

$$\phi_{8,0}(x) := \tfrac{1}{2}\sqrt{2} \times \begin{cases} \phi_8(x) & \text{when } x \le 1, \\ \phi_0(x-2) & \text{when } x \ge 1, \end{cases}$$

	-1	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\tilde{\phi}_0$	$\frac{49}{8}$	$-\frac{953}{1024}$	$\frac{27}{64}$	$-\frac{73}{1024}$	$\frac{5}{8}$	$-\frac{97}{1024}$	$\frac{3}{64}$	$-\frac{17}{1024}$	$\frac{1}{8}$
$\tilde{\phi}_1$	5	$\frac{175}{128}$	0	$-\frac{25}{128}$	0	$\frac{15}{128}$	0	$-\frac{25}{128}$	0
$\tilde{\phi}_2$	0	$\frac{105}{32}$	3	$\frac{45}{32}$	0	$-\frac{15}{32}$	0	$\frac{21}{32}$	0
$\tilde{\phi}_3$	0	$-\frac{175}{128}$	0	$\frac{225}{128}$	$\frac{5}{2}$	$\frac{225}{128}$	0	$-\frac{175}{128}$	0
$\tilde{\phi}_4$	$\frac{1551}{784}$	$\frac{9651}{12544}$	$\frac{7179}{3136}$	$-\frac{4029}{784}$	$\frac{3225}{784}$	$\frac{17307}{6272}$	$-\frac{6639}{3136}$	$-\frac{4461}{12544}$	$-\frac{801}{784}$

TABLE 6.2
Values of $\{\tilde{\phi}_0, \dots, \tilde{\phi}_4\}$ *at* $\frac{1}{4}\mathbb{Z} \cap [-1, 1]$

which has support $[-1, 3]$, for $j \in \mathbb{N}_0$ we define

$$\phi_{j,k}^{(i,\mathcal{X})} := \sqrt{\tfrac{2^j}{T}} \times \begin{cases} \phi_i(\frac{2^j}{T} \cdot -k) & \text{when } i \in \{1, \dots, 7\},\ k \in \{0, \dots, 2^j - 1\}, \\ \phi_{8,0}(\frac{2^j}{T} \cdot -k) & \text{when } i = 8,\ k \in \{0, \dots, 2^j - 2\}, \\ \phi_8(\frac{2^j}{T} \cdot -k) & \text{when } i = 8 \text{ and } k = 2^j - 1. \end{cases}$$

and, by replacing ϕ by $\tilde{\phi}$ on all places, similarly at the dual side.

It holds that $V_j^{\mathcal{X}} = \operatorname{span}\{\phi_{j,k}^{(i,\mathcal{X})} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$, and we set $\tilde{V}_j^{\mathcal{X}} := \operatorname{span}\{\tilde{\phi}_{j,k}^{(i,\mathcal{X})} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$. Since none of $\tilde{\phi}_2, \dots, \tilde{\phi}_6$ is "glued" over interfaces between "macro elements" or "dropped" at the (left) boundary, from $P_4(-1, 1) = \operatorname{span}\{\tilde{\phi}_2, \dots, \tilde{\phi}_6\}$ and $\tilde{V} \subset P_4(-1, 0) \times P_4(0, 1)$, we have

$$\prod_{k=0}^{2^j} P_4(k2^{-j}T, (k+1)2^{-j}T) \subset \tilde{V}_j^{\mathcal{X}} \subset \prod_{k=0}^{2^{j+1}} P_4(k2^{-(j+1)}T, (k+1)2^{-(j+1)}T), \tag{6.8}$$

and so $\tilde{V}_j^{\mathcal{X}} \subset \tilde{V}_{j+1}^{\mathcal{X}}$ $(j \in \mathbb{N}_0)$.

The collections $\{\phi_{j,k}^{(i,\mathcal{X})} : i, k\}$ and $\{\tilde{\phi}_{j,k}^{(i,\mathcal{X})} : i, k\}$ are uniformly local, and uniform biorthogonal $L_2(\mathrm{I})$-Riesz bases for their spans, and so Proposition 6.3(a) shows that the "inf-inf-sup" condition (6.1) of Theorem 6.2 is satisfied. The Jackson and Bernstein conditions are satisfied with

$$(\gamma_{\mathcal{X}}, d_{\mathcal{X}}, \mathcal{H}_{\mathcal{X}}^{d_{\mathcal{X}}}, \tilde{\gamma}_{\mathcal{X}}, \tilde{d}_{\mathcal{X}}, \tilde{\mathcal{H}}_{\mathcal{X}}^{\tilde{d}_{\mathcal{X}}}) = (\tfrac{3}{2}, 5, H^5(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I}), \tfrac{1}{2}, 5, H^5(\mathrm{I})). \tag{6.9}$$

At the $\mathcal{Y}$-side, setting $\phi_{j,k}^{(i,\mathcal{Y})}(x) = \phi_{j,k}^{(i,\mathcal{X})}(-x)$ and $\tilde{\phi}_{j,k}^{(i,\mathcal{Y})}(x) = \tilde{\phi}_{j,k}^{(i,\mathcal{X})}(-x)$, it holds that $V_j^{\mathcal{Y}} = \operatorname{span}\{\phi_{j,k}^{(i,\mathcal{Y})} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$, and we set $\tilde{V}_j^{\mathcal{Y}} := \operatorname{span}\{\tilde{\phi}_{j,k}^{(i,\mathcal{Y})} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$. Then everything above is also valid with $\mathcal{X}$ reading as $\mathcal{Y}$, with

$$(\gamma_{\mathcal{Y}}, d_{\mathcal{Y}}, \mathcal{H}_{\mathcal{Y}}^{d_{\mathcal{Y}}}, \tilde{\gamma}_{\mathcal{Y}}, \tilde{d}_{\mathcal{Y}}, \tilde{\mathcal{H}}_{\mathcal{Y}}^{\tilde{d}_{\mathcal{Y}}}) = (\tfrac{3}{2}, 5, H^5(\mathrm{I}) \cap H^1_{0,\{T\}}(\mathrm{I}), \tfrac{1}{2}, 5, H^5(\mathrm{I})). \tag{6.10}$$

We may conclude that, as required, the wavelet collections $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ that we are going to construct in the next subsection will, properly scaled, be Riesz bases for, in particular, $L_2(\mathrm{I})$ or $L_2(\mathrm{I})$ and $H^1_{0,\{T\}}(\mathrm{I})$, respectively.

The conditions (6.2), (6.3) and (6.4) are satisfied by construction, in particular by (6.7) and the left inclusion in (6.8) (which is also valid with $\mathcal{X}$ reading as $\mathcal{Y}$).

Remark 6.4. Besides realizing nestedness of the spaces and Jackson and Bernstein estimates, as well as satisfying the conditions (6.2)-(6.4) for sparsity of $\langle \Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}} \rangle$

and $\langle\Theta^{\mathcal{X}},\Theta^{\mathcal{Y}}\rangle$, our construction of $\tilde{V}$, and so of $(\tilde{V}_j^{\mathcal{Z}})_j$ ($\mathcal{Z}\in\{\mathcal{X},\mathcal{Y}\}$), was motivated by the aim of making the angle between V and $\tilde{V}$, and so between $V_j^{\mathcal{Z}}$ and $\tilde{V}_j^{\mathcal{Z}}$ as small as possible. A small angle allows for a construction of the wavelets spanning the biorthogonal complements so that overall wavelet basis has a small condition number (in any case in $L_2(\mathrm{I})$).

6.5. Definition of the wavelets. In this subsection, we will define the basis

$$\Theta_{j+1}^{\mathcal{X}}=\big\{\theta_{j+1,k}^{(i,\mathcal{X})}:i\in\{1,\dots,8\},\,k\in\{0,\dots,2^j-1\}\big\}$$

for $V_{j+1}^{\mathcal{X}}\cap(\tilde{V}_j^{\mathcal{X}})^{\perp_{L_2(\mathrm{I})}}$. The basis $\Theta_{j+1}^{\mathcal{Y}}=\big\{\theta_{j+1,k}^{(i,\mathcal{Y})}:i\in\{1,\dots,8\},\,k\in\{0,\dots,2^j-1\}\big\}$ for $V_{j+1}^{\mathcal{Y}}\cap(\tilde{V}_j^{\mathcal{Y}})^{\perp_{L_2(\mathrm{I})}}$ can then be defined by

$$\theta_{j+1,k}^{(i,\mathcal{Y})}(x)=\theta_{j+1,2^j-1-k}^{(i,\mathcal{X})}(T-x). \tag{6.11}$$

In view of Proposition 6.3(b), in order to define these wavelets, it is sufficient to construct $\Xi_{j+1}^{\mathcal{X}}$ such that $\Phi_j^{\mathcal{X}}\cup\Xi_{j+1}^{\mathcal{X}}$ is a uniform $L_2(\mathrm{I})$-Riesz basis for $V_{j+1}^{\mathcal{X}}$. Moreover, in view of Proposition 6.3(d), in order to obtain wavelets that are uniformly local, *and* have uniformly local duals, we shall select $\Xi_{j+1}^{\mathcal{X}}$ in such a way that the basis transformation between $\Phi_j^{\mathcal{X}}\cup\Xi_{j+1}^{\mathcal{X}}$ and $\Phi_{j+1}^{\mathcal{X}}$ *and* its inverse are uniformly local. Since the basis transformations between the standard interpolatory basis of $V_\ell^{\mathcal{X}}$ and $\Phi_\ell^{\mathcal{X}}$ are uniformly local, the latter condition is equivalent to the uniform locality of the basis transformations between the union of the interpolatory basis for $V_j^{\mathcal{X}}$ and $\Xi_{j+1}^{\mathcal{X}}$ and the interpolatory basis for $V_{j+1}^{\mathcal{X}}$.

A natural choice for $\Xi_{j+1}^{\mathcal{X}}$ is the subset of interpolatory basis functions for $V_{j+1}^{\mathcal{X}}$ that correspond to the new degrees of freedom. With this choice, the last mentioned basis transformations are uniformly local. Indeed, with I_ℓ being the canonical interpolation operator onto $V_\ell^{\mathcal{X}}$, the argument is that for $u_{j+1}\in V_{j+1}^{\mathcal{X}}$ the computation of the splitting $u_{j+1}=I_ju_{j+1}+I_{j+1}(u_{j+1}-I_ju_{j+1})$ requires local quantities only.

In order to reduce the support size of most of the resulting wavelets, we do not simply take $\Xi_{j+1}^{\mathcal{X}}$ to be the above collection, but we construct it from that collection by applying a uniformly local transformation with uniformly local inverse. Our aim is to ensure that most functions in $\Xi_{j+1}^{\mathcal{X}}$ are orthogonal to those duals scaling functions on level j that correspond to primal scaling functions that have supports that extend to more than one macro-element. The wavelets resulting from such functions in $\Xi_{j+1}^{\mathcal{X}}$ will then have no components in the directions of these scaling functions, and therefore will be supported inside one macro-element.

We start with a construction on the "reference macro element" $(-1,1)$. Let

$$W:=\{w\in\prod_{k=-2}^{1}P_4(\tfrac{k}{2},\tfrac{k+1}{2})\cap C(-1,1):w=0\text{ on }\tfrac{1}{4}\mathbb{Z}\}.$$

Note that $\prod_{k=-2}^{1}P_4(\frac{k}{2},\frac{k+1}{2})\cap C(-1,1)=P_4(-1,0)\times P_4(0,1)\cap C(-1,1)\oplus W$. We performed the following steps:

- Determine the orthogonal projection of $\mathrm{span}\{\phi_0,\phi_8\}$ onto W, and equip it with a basis $\{\xi_1,\xi_8\}$ that is biorthogonal to $\{\phi_0,\phi_8\}$,
- determine $W_2\perp\mathrm{span}\{\phi_0,\phi_8\}$ such that $W=\mathrm{span}\{\xi_1,\xi_8\}\oplus W_2$, and equip W_2 with a basis $\{\xi_2,\dots,\xi_7\}$.

The basis $\{\xi_1, \dots, \xi_8\}$ for W can be organized such that $\xi_k(x) = \xi_{9-k}(x)$.

Similar to the previous subsection, we lift these functions on the reference macro element to the macro elements $[k2^{-j}T, (k+1)2^{-j}T]$ that form a partition of I. Since, all ξ_i vanish at the boundary of the reference macro element, here is no need to "glue" functions over interfaces. We define $\Xi^{\mathcal{X}}_{j+1} = \{\xi^{(i,\mathcal{X})}_{j+1,k} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$ by

$$\xi^{(i,\mathcal{X})}_{j+1,k} := \sqrt{\tfrac{2^j}{T}}\xi_i(\tfrac{2^j}{T} \cdot -k).$$

With this definition, $\Phi^{\mathcal{X}}_j \cup \Xi^{\mathcal{X}}_{j+1}$ is a uniform $L_2(\mathrm{I})$-Riesz basis for $V^{\mathcal{X}}_{j+1}$ and, in view of above comments, the basis transformations between $\Phi^{\mathcal{X}}_j \cup \Xi^{\mathcal{X}}_{j+1}$ and $\Phi^{\mathcal{X}}_{j+1}$ are uniformly local. As shown in Proposition 6.3, therefore

$$\Theta^{\mathcal{X}}_{j+1} := \Xi^{\mathcal{X}}_{j+1} - \langle \Xi^{\mathcal{X}}_{j+1}, \tilde{\Phi}^{\mathcal{X}}_j \rangle_{L_2(\mathrm{I})} \Phi^{\mathcal{X}}_j$$

is a uniformly local, uniform $L_2(\mathrm{I})$-Riesz basis for $V^{\mathcal{X}}_{j+1} \cap (\tilde{V}^{\mathcal{X}}_j)^{\perp_{L_2(\mathrm{I})}}$ with a corresponding uniform $L_2(\mathrm{I})$-Riesz dual basis $\tilde{\Theta}^{\mathcal{X}}_{j+1}$ for $\tilde{V}^{\mathcal{X}}_{j+1} \cap (V^{\mathcal{X}}_j)^{\perp_{L_2(\mathrm{I})}}$ that is also uniformly local.

In order to improve its conditioning, next we apply some uniformly local basis transformations to $\Theta^{\mathcal{X}}_{j+1}$ with uniformly local inverses, so that both $\Theta^{\mathcal{X}}_{j+1}$ and $\tilde{\Theta}^{\mathcal{X}}_{j+1}$ remain uniformly local.

The wavelets from $\Theta^{\mathcal{X}}_{j+1}$ can be subdivided into 3 categories:

- Wavelets with supports inside a macro element $(kT2^{-j}, (k+1)T2^{-j})$ $(k \in \{0, \dots, 2^j - 1\})$. Six for each macro element.
- Wavelets with supports inside the union of 2 macro elements that share an interface. Two for each interface.
- One wavelet "associated to" the right boundary, non-zero at that boundary, and one "associated to" the left boundary, zero at that boundary.

We orthogonalize each group of six "internal wavelets", creating 3 symmetric and 3 anti-symmetric wavelets; we make each group of two "interface wavelets" orthogonal to the "adjacent" internal wavelets, and after that, we make them mutual orthogonal, creating a symmetric and an anti-symmetric wavelet; we make both boundary wavelets orthogonal to the "adjacent" internal wavelets.

Exploiting dilation and translation invariance, it suffices to specify the resulting left and right boundary wavelets $\theta^{\mathcal{X},L}$ and $\theta^{\mathcal{X},R}$, the 3 symmetric and 3 anti-symmetric interior wavelets $\theta^{\mathcal{X},S,1}$, $\theta^{\mathcal{X},S,2}$, $\theta^{\mathcal{X},S,3}$ and $\theta^{\mathcal{X},A,1}$, $\theta^{\mathcal{X},A,2}$, $\theta^{\mathcal{X},A,3}$, respectively, all as functions on $C([-1,1]) \cap \prod_{k=-2}^{1} P_4(\frac{k}{2}, \frac{k+1}{2})$, and the symmetric and anti-symmetric interface wavelets $\theta^{\mathcal{X},S}$ and $\theta^{\mathcal{X},A}$, both as functions on $C([-1,3]) \cap \prod_{k=-2}^{5} P_4(\frac{k}{2}, \frac{k+1}{2})$, i.e., with double support length. These "mother wavelets" are illustrated in Figures 6.2 and 6.3.

Values of these "mother wavelets" in $\frac{1}{8}\mathbb{Z} \cap (-1,1]$ (boundary and interface wavelets) or in $\frac{1}{8}\mathbb{Z} \cap (-1,0]$ (interior wavelets) are given in Table 6.3. Values of the interface wavelets at $\frac{1}{8}\mathbb{Z} \cap [1,3)$ or of the interior wavelets at $\frac{1}{8}\mathbb{Z} \cap [0,1)$ can be found using the (anti-) symmetry of these functions.

When computing the corresponding collection of dual wavelets $\tilde{\Theta}^{\mathcal{X}}_{j+1} = \{\tilde{\theta}^{(i,\mathcal{X})}_{j+1,k} : i \in \{1, \dots, 8\}, k \in \{0, \dots, 2^j - 1\}\}$, it turns out that for each j, the *seven* dual wavelets that have their supports nearest to the left boundary, i.e., $\tilde{\theta}^{(1,\mathcal{X})}_{j+1,0}, \dots, \tilde{\theta}^{(7,\mathcal{X})}_{j+1,0}$, have no vanishing first moment (all dual wavelets have vanishing second to fifth moments).

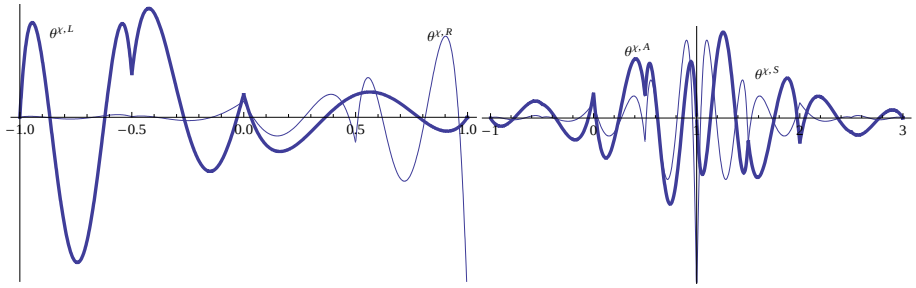

FIG. 6.2. *Boundary wavelets $\theta^{\mathcal{X},L}$, $\theta^{\mathcal{X},R}$ (left picture), and interface wavelets $\theta^{\mathcal{X},S}$, $\theta^{\mathcal{X},A}$ (right picture).*

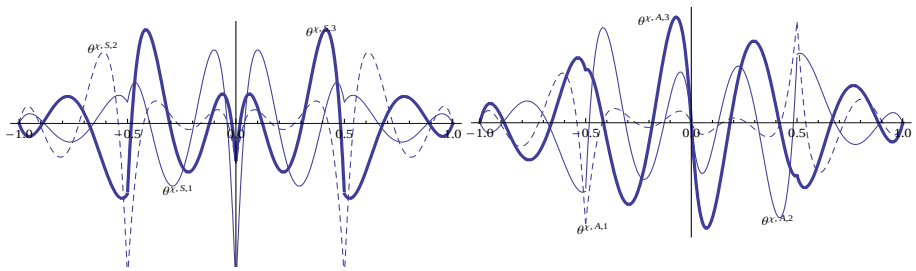

FIG. 6.3. *Symmetric interior wavelets $\theta^{\mathcal{X},S,1}$, $\theta^{\mathcal{X},S,2}$, $\theta^{\mathcal{X},S,3}$ (left picture), and anti-symmetric interior wavelets $\theta^{\mathcal{X},A,1}$, $\theta^{\mathcal{X},A,2}$, $\theta^{\mathcal{X},A,3}$ (right picture).*

The existence of such dual wavelets reflects the fact that, due to the fact that all primal wavelets vanish at the left boundary, the primal collection has a locally strongly reduced approximation order near this boundary. This was a price that we had to pay for obtaining sparse system matrices.

For $i = 2, \ldots, 7$, and with $\beta_i := \int_{\mathrm{I}} \tilde{\theta}^{(i,\mathcal{X})}_{j+1,0}(x)dx / \int_{\mathrm{I}} \tilde{\theta}^{(1,\mathcal{X})}_{j+1,0}(x)dx$, which scalars are independent of j, let us now *redefine* the left boundary primal wavelet $\theta^{(1,\mathcal{X})}_{j+1,0}$, i.e., the wavelet resulting from the mother wavelet $\theta^{\mathcal{X},L}$, by

$$\theta^{(1,\mathcal{X})}_{j+1,0} \leftarrow \theta^{(1,\mathcal{X})}_{j+1,0} + \sum_{i=2}^{7} \beta_i \theta^{(i,\mathcal{X})}_{j+1,0}.$$

Clearly, this transformation does not change the span of the wavelets on level $j+1$, or the (qualitative) stability properties of the basis $\Theta^{\mathcal{X}}$. Since the corresponding transformation at the dual side is given by $\tilde{\theta}^{(i,\mathcal{X})}_{j+1,0} \leftarrow \tilde{\theta}^{(i,\mathcal{X})}_{j+1,0} - \beta_i \tilde{\theta}^{(1,\mathcal{X})}_{j+1,0}$, $(i = 2, \ldots, 7)$, we conclude, however, that after this transformation, on each level there remains only *one* dual wavelet without vanishing moment. This allows for a quantitatively more efficient representation of a function that is non-zero at 0 in the wavelet basis $\Theta^{\mathcal{X}}$, and so improves the quantitative properties of the resulting adaptive wavelet scheme. For $i = 2, \ldots, 7$, $\beta_i = -\frac{1}{32}, -\frac{173}{10608}, \frac{20659}{376064}, \frac{8001065}{87548032}, \frac{12452429}{69354240}, \frac{162901153}{762799872}$.

Finally, in order to improve its conditioning in Sobolev norms other than the $L_2(\mathrm{I})$-norm, in the definition of $\Theta^{\mathcal{X}}$ we replace the single scale basis $\Phi^{\mathcal{X}}_0$ on the lowest level by an orthogonal four-scale basis. Recalling that $V^{\mathcal{X}}_0 = P_4(0, T/2) \times P_4(T/2, T) \cap$

	$\theta^{\mathcal{X},L}$	$\theta^{\mathcal{X},R}$	$\theta^{\mathcal{X},S,1}$	$\theta^{\mathcal{X},S,2}$	$\theta^{\mathcal{X},S,3}$	$\theta^{\mathcal{X},A,1}$	$\theta^{\mathcal{X},A,2}$	$\theta^{\mathcal{X},A,3}$	$\theta^{\mathcal{X},S}$	$\theta^{\mathcal{X},A}$
$-\frac{7}{8}$	$\frac{2698131}{35651584}$	$\frac{-2585}{1671168}$	$\frac{-3269185}{619753472}$	$\frac{-96479}{2311808}$	$\frac{21275}{1265664}$	$\frac{-5607303}{127133312}$	$\frac{10697}{749568}$	$\frac{-15019819}{1400768512}$	$\frac{-2585}{1671168}$	$\frac{-6434887}{106954752}$
$-\frac{6}{8}$	$\frac{-1465853}{2228224}$	$\frac{-905}{104448}$	$\frac{-1868545}{38734592}$	$\frac{-6559}{144488}$	$\frac{5947}{79104}$	$\frac{-409435}{7945832}$	$\frac{959}{15616}$	$\frac{-7473487}{87548032}$	$\frac{-905}{104448}$	$\frac{112777}{6684672}$
$-\frac{5}{8}$	$\frac{-1264957}{35651584}$	$\frac{4317}{557056}$	$\frac{16606575}{619753472}$	$\frac{428337}{2311808}$	$\frac{-40615}{421888}$	$\frac{25038585}{127133312}$	$\frac{-20925}{249856}$	$\frac{80992245}{1400768512}$	$\frac{4317}{557056}$	$\frac{4050883}{35651584}$
$-\frac{4}{8}$	$\frac{23899}{139264}$	$\frac{19}{6528}$	$\frac{134975}{2420912}$	$\frac{-7510}{18061}$	$\frac{-1045}{4944}$	$\frac{-419094}{993229}$	$\frac{-613}{2928}$	$\frac{680411}{5471752}$	$\frac{19}{6528}$	$\frac{44753}{417792}$
$-\frac{3}{8}$	$\frac{45595177}{106954752}$	$\frac{-3977}{1671168}$	$\frac{-33051825}{619753472}$	$\frac{138385}{2311808}$	$\frac{280555}{1265664}$	$\frac{7229001}{127133312}$	$\frac{54643}{249856}$	$\frac{-135830043}{1400768512}$	$\frac{-3977}{1671168}$	$\frac{893737}{106954752}$
$-\frac{2}{8}$	$\frac{89785}{6684672}$	$\frac{-1817}{104448}$	$\frac{-5136465}{38734592}$	$\frac{-1391}{144488}$	$\frac{-3335}{26368}$	$\frac{-55731}{7945832}$	$\frac{-2441}{15616}$	$\frac{-13807799}{87548032}$	$\frac{-1817}{104448}$	$\frac{-891335}{6684672}$
$-\frac{1}{8}$	$\frac{-10916749}{35651584}$	$\frac{11143}{1671168}$	$\frac{107443295}{619753472}$	$\frac{53569}{2311808}$	$\frac{-7237}{1265664}$	$\frac{1656681}{127133312}$	$\frac{6137}{749568}$	$\frac{237582629}{1400768512}$	$\frac{11143}{1671168}$	$\frac{-15206311}{106954752}$
0	$\frac{293}{1536}$	$\frac{1}{12}$	$\frac{-67875}{151307}$	$\frac{-1984}{18061}$	$\frac{-35}{309}$	0	0	0	$\frac{1}{12}$	$\frac{293}{1536}$
$\frac{1}{8}$	$\frac{-15206311}{106954752}$	$\frac{-154385}{1671168}$	–	–	–	–	–	–	$\frac{-154385}{1671168}$	$\frac{-10916749}{35651584}$
$\frac{2}{8}$	$\frac{-891335}{6684672}$	$\frac{-2705}{104448}$	–	–	–	–	–	–	$\frac{-2705}{104448}$	$\frac{89785}{6684672}$
$\frac{3}{8}$	$\frac{893737}{106954752}$	$\frac{195295}{1671168}$	–	–	–	–	–	–	$\frac{195295}{1671168}$	$\frac{45595177}{106954752}$
$\frac{4}{8}$	$\frac{44753}{417792}$	$\frac{-869}{6528}$	–	–	–	–	–	–	$\frac{-869}{6528}$	$\frac{23899}{139264}$
$\frac{5}{8}$	$\frac{4050883}{35651584}$	$\frac{-19243}{557056}$	–	–	–	–	–	–	$\frac{-19243}{557056}$	$\frac{-1264957}{35651584}$
$\frac{6}{8}$	$\frac{112777}{6684672}$	$\frac{-10091}{34816}$	–	–	–	–	–	–	$\frac{-10091}{34816}$	$\frac{-1465853}{2228224}$
$\frac{7}{8}$	$\frac{-6434887}{106954752}$	$\frac{613327}{1671168}$	–	–	–	–	–	–	$\frac{613327}{1671168}$	$\frac{2698131}{35651584}$
1	0	-1	–	–	–	–	–	–	-1	0

TABLE 6.3
Values of the "mother wavelets at $\frac{1}{8}\mathbb{Z}$.

$C(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I})$, we set $V^{\mathcal{X}}_{-1} = P_4(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I})$, $V^{\mathcal{X}}_{-2} = P_2(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I})$, and $V^{\mathcal{X}}_{-3} = P_1(\mathrm{I}) \cap H^1_{0,\{0\}}(\mathrm{I})$, and replace the single scale basis for $V^{\mathcal{X}}_0$ by a basis that is the union of orthogonal bases for $V^{\mathcal{X}}_{-3}$, $V^{\mathcal{X}}_{-2} \cap (V^{\mathcal{X}}_{-3})^{\perp_{L_2(\mathrm{I})}}$, $V^{\mathcal{X}}_{-1} \cap (V^{\mathcal{X}}_{-2})^{\perp_{L_2(\mathrm{I})}}$, and $V^{\mathcal{X}}_0 \cap (V^{\mathcal{X}}_{-1})^{\perp_{L_2(\mathrm{I})}}$.

6.6. Condition numbers. As shown by Theorem 6.2, the collection $\Theta^{\mathcal{X}}$, renormalized in the corresponding norm, is a Riesz basis for range of Sobolev spaces, in particular for $L_2(\mathrm{I})$ and $H^1_{0,\{0\}}(\mathrm{I})$. In various estimates, the values of (the quotients of) the corresponding Riesz constants play a role. To provide an estimate for these values, in Figure 6.6, we present numerically computed condition numbers of

$$\mathbf{M}^{\mathcal{X}}_J := \Big[\frac{\langle \theta^{\mathcal{X}}_\mu, \theta^{\mathcal{X}}_\lambda\rangle_{L_2(\mathrm{I})}}{\|\theta^{\mathcal{X}}_\mu\|_{L_2(\mathrm{I})}\|\theta^{\mathcal{X}}_\lambda\|_{L_2(\mathrm{I})}}\Big]_{|\lambda|,|\mu|\le J} \text{ and } \mathbf{A}^{\mathcal{X}}_J := \Big[\frac{\langle \dot\theta^{\mathcal{X}}_\mu, \dot\theta^{\mathcal{X}}_\lambda\rangle_{L_2(\mathrm{I})}}{\|\dot\theta^{\mathcal{X}}_\mu\|_{L_2(\mathrm{I})}\|\dot\theta^{\mathcal{X}}_\lambda\|_{L_2(\mathrm{I})}}\Big]_{|\lambda|,|\mu|\le J}.$$

For $J \to \infty$, these condition numbers converge to $(\Lambda_{L_2(\mathrm{I})}(|\Theta^{\mathcal{X}}|)/\lambda_{L_2(\mathrm{I})}(|\Theta^{\mathcal{X}}|))^2$ and $(\Lambda_{H^1_{0,\{0\}}(\mathrm{I})}(|\Theta^{\mathcal{X}}|)/\lambda_{H^1_{0,\{0\}}(\mathrm{I})}(|\Theta^{\mathcal{X}}|))^2$, respectively, where $H^1_{0,\{0\}}(\mathrm{I})$ is equipped with $|\cdot|_{H^1(\mathrm{I})}$. The computed condition numbers compare very favorable to other wavelet constructions of order 5.

6.7. Approximation order. We have constructed a Riesz basis $\Theta^{\mathcal{X}}$ for $L_2(\mathrm{I})$ consisting of continuous wavelets that all vanish at 0. This basis is going to be used for the (adaptive) approximation in $L_2(\mathrm{I})$ of a function that not necessarily vanishes at 0. In this subsection, we investigate whether there is a reduction in approximation order due to a possible mismatch of boundary conditions.

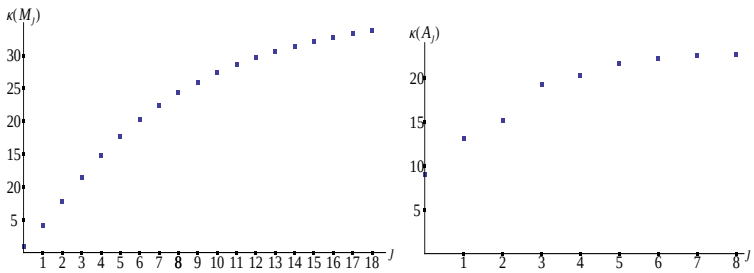

FIG. 6.4. *Condition number of* $\mathbf{M}_J^{\mathcal{X}}$ *(left) and* $\mathbf{A}_J^{\mathcal{X}}$ *(right).*

With, for $u \in L_2(\mathrm{I})$, $u_\lambda := \langle u, \tilde{\theta}_\lambda^{\mathcal{X}} \rangle_{L_2(\mathrm{I})}$, we have $u = \sum_{\lambda \in \nabla^{\mathcal{X}}} u_\lambda \theta_\lambda^{\mathcal{X}}$ and $\|u\|_{L_2(\mathrm{I})}^2 \eqsim \sum_{\lambda \in \nabla^{\mathcal{X}}} |u_\lambda|^2$, and so $\|u - \sum_{\lambda \in \Lambda} u_\lambda \theta_\lambda^{\mathcal{X}}\|_{L_2(\mathrm{I})}^2 \eqsim \sum_{\lambda \in \nabla^{\mathcal{X}} \setminus \Lambda} |u_\lambda|^2$ for any $\Lambda \subset \nabla^{\mathcal{X}}$.

We subdivide the dual wavelets in those that have $d_{\mathcal{X}} = 5$ vanishing moments, and those that have less vanishing moments, which in the current situation actually means that they have no vanishing moments. On each level there is a uniformly bounded number – being one – of dual wavelets without vanishing moments, viz., the dual wavelet that has its support closest to the left boundary. We call indices λ that correspond to the latter dual wavelets left boundary indices, and the other ones regular indices.

For the regular indices, we have

$$|u_\lambda| \lesssim 2^{-d_{\mathcal{X}}|\lambda|} |u|_{H^{d_{\mathcal{X}}}(\operatorname{supp} \tilde{\theta}_\lambda^{\mathcal{X}})},$$

and for the left boundary ones,

$$|u_\lambda| \lesssim \|u\|_{L_2(\operatorname{supp} \tilde{\theta}_\lambda^{\mathcal{X}})} \lesssim 2^{-\frac{1}{2}|\lambda|} \|u\|_{L_\infty(\mathrm{I})} \lesssim 2^{-\frac{1}{2}|\lambda|} \|u\|_{H^{d_{\mathcal{X}}}(\mathrm{I})}.$$

By defining Λ as the union of all regular indices up to some level J together with all left boundary indices up to level $2d_{\mathcal{X}}J$, we have $\#\Lambda \eqsim 2^J + 2d_{\mathcal{X}}J \eqsim 2^J$, and by estimating $\sum_{\lambda \in \nabla^{\mathcal{X}} \setminus \Lambda} |u_\lambda|^2$, we conclude that $\|u - \sum_{\lambda \in \Lambda} u_\lambda \theta_\lambda^{\mathcal{X}}\|_{L_2(\mathrm{I})} \lesssim 2^{-d_{\mathcal{X}} J}$. So despite of the fact that all primal wavelets vanish at 0, the order is $d_{\mathcal{X}}$.

The above result deals with linear approximation, and requires a sufficiently smooth u. The smoothness conditions can be relaxed largely by considering non-linear approximation. For $s \in [0, d_{\mathcal{X}})$, $\frac{1}{\tau} = \frac{1}{2} + s$ and $u \in B_{\tau,\tau}^s(\mathrm{I})$, it is known, see e.g. [DeV98, Coh00], that

$$\Big(\sum_{\{\lambda \in \nabla^{\mathcal{X}} : \lambda \text{ is regular }\}} |u_\lambda|^\tau \Big)^{1/\tau} \lesssim \|u\|_{B_{\tau,\tau}^s(\mathrm{I})}.$$

Making a non-increasing arrangement of the coefficients corresponding to all regular indices, i.e., $|u_{\lambda_1}| \geq |u_{\lambda_2}| \geq \ldots$, we have

$$n |u_{\lambda_n}|^\tau \leq \sum_{k=1}^n |u_{\lambda_k}|^\tau \lesssim \|u\|_{B_{\tau,\tau}^s(\mathrm{I})}^\tau,$$

and so $\sqrt{\sum_{n > N} |u_{\lambda_n}|^2} \lesssim N^{-s} \|u\|_{B_{\tau,\tau}^s(\mathrm{I})}$.

To estimate the coefficients corresponding to the left boundary indices, let us additionally assume that $u \in B_{\tau',\tau'}^s(\mathrm{I})$ for some $\tau' > \tau$. Taking, w.l.o.g., $\frac{1}{\tau'} \geq \frac{1}{\tau} - \frac{1}{2}$,

and with $\frac{1}{p} := \frac{1}{2} + \frac{1}{\tau'} - \frac{1}{\tau}$, Hölder's inequality and Sobolev's embedding theorem show that

$$|u_\lambda| \lesssim \|u\|_{L_2(\operatorname{supp}\tilde{\theta}^{\mathcal{X}}_\lambda)} \lesssim 2^{(\frac{1}{\tau'}-\frac{1}{\tau})|\lambda|}\|u\|_{L_p(\operatorname{supp}\tilde{\theta}^{\mathcal{X}}_\lambda)} \lesssim 2^{(\frac{1}{\tau'}-\frac{1}{\tau})|\lambda|}\|u\|_{B^s_{\tau',\tau'}(\mathrm{I})}.$$

By taking $J = \lceil \frac{s\log_2 N}{1/\tau - 1/\tau'} \rceil$, and defining Λ as the union of $\{\lambda_1, \ldots, \lambda_N\}$ and all left boundary indices with levels $\leq J$, we conclude that $\#\Lambda \eqsim N + \log_2 N \eqsim N$ and $\|u - \sum_{\lambda\in\Lambda} u_\lambda \theta^{\mathcal{X}}_\lambda\|_{L_2(\mathrm{I})} \lesssim N^{-s}\|u\|_{B^s_{\tau',\tau'}(\mathrm{I})}$. So under an only marginal stronger smoothness requirement, we obtain the same qualitative result concerning best N-term approximation as for a common wavelet basis of order $d_{\mathcal{X}}$, i.e., one without the condition that all primal wavelets vanish at zero.

We can rephrase our finding by saying that for $s \in [0, d_{\mathcal{X}})$, and $\frac{1}{\tau'} < \frac{1}{2} + s$, $B^s_{\tau',\tau'}(\mathrm{I})$ contains the nonlinear approximation class $\mathcal{A}^s_\infty(L_2(\mathrm{I}))$ corresponding to $\Theta^{\mathcal{X}}$.

Remark 6.5. For approximating the solution of the parabolic problem using the tensor product wavelet basis $\Theta^{\mathcal{X}} \otimes \Sigma^{(1)} \otimes \cdots \otimes \Sigma^{(1)}$, a (near) characterization of the corresponding approximation class $\mathcal{A}^s_\infty(L_2(\mathrm{I}; L_2((0,1)^n)))$ is relevant. By a combination of results from [Nit06] and [Coh03, §3.10] (the latter for dealing with the boundary conditions incorporated in $\Sigma^{(1)}$), one can deduce that *without* the incorporation of the zero boundary condition in $\Theta^{\mathcal{X}}$, for $s \in [0, d_{\mathcal{X}})$ and $\frac{1}{\tau} = \frac{1}{2} + s$,

$$\begin{aligned} u \in \mathcal{A}^s_\infty(L_2(\mathrm{I}; L_2((0,1)^n))) \Longleftarrow u \in \mathcal{A}^s_\tau(L_2(\mathrm{I}; L_2((0,1)^n))) \Longleftrightarrow \\ u \in B^s_{\tau,\tau}(\mathrm{I}) \otimes_\tau \overset{\circ}{B}{}^s_{\tau,\tau}(0,1) \otimes_\tau \cdots \otimes_\tau \overset{\circ}{B}{}^s_{\tau,\tau}(0,1), \end{aligned}$$

where $\overset{\circ}{B}{}^s_{\tau,\tau}(0,1)$ is the closure in $B^s_{\tau,\tau}(0,1)$ of the smooth functions on $[0,1]$ that vanish at $\{0,1\}$; whereas *with* the incorporation of the zero boundary condition in $\Theta^{\mathcal{X}}$, for $s \in [0, d_{\mathcal{X}})$ and $\frac{1}{\tau'} > \frac{1}{\tau} = \frac{1}{2} + s$,

$$u \in \mathcal{A}^s_\infty(L_2(\mathrm{I}; L_2((0,1)^n))) \Longleftarrow u \in B^s_{\tau',\tau'}(\mathrm{I}) \otimes_\tau \overset{\circ}{B}{}^s_{\tau,\tau}(0,1) \otimes_\tau \cdots \otimes_\tau \overset{\circ}{B}{}^s_{\tau,\tau}(0,1).$$

Again, the required smoothness condition is only marginally stronger.

7. First test on an ODE. As a first test of our temporal wavelets $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ and of the adaptive wavelet-Galerkin scheme, we consider the linear ODE

$$\begin{cases} \dot{u}(t) + \nu u(t) = g(t) & (t \in \mathrm{I}), \\ u(0) = u_0, \end{cases} \tag{7.1}$$

where $\nu \geq 0$. The corresponding variational formulation reads as

$$b(u, v) = f(v), \tag{7.2}$$

where

$$b(w, v) := \int_{\mathrm{I}} -w(t)\dot{v}(t) + \nu w(t)v(t)dt, \quad f(v) := \int_{\mathrm{I}} g(t)v(t)dt + u_0 v(0).$$

As a special case of Theorem 2.2 (or 2.4) we have the following result. For the purpose of illustrating the technique, we included the proof for this elementary case.

THEOREM 7.1. *With $\mathcal{X} := L_2(\mathrm{I})$ and $\mathcal{Y}(\nu) := H^1_{0,\{T\}}(\mathrm{I})$, equipped with $\|\cdot\|_{\mathcal{Y}(\nu)} := \sqrt{\nu^2\|\cdot\|^2_{L_2(\mathrm{I})} + |\cdot|^2_{H^1(\mathrm{I})}}$, the operator $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}(\nu)')$ defined by $(Bw)(v) = b(w, v)$ is boundedly invertible, with $\|B\| \leq \sqrt{2}$, $\|B^{-1}\| \leq \sqrt{2}$.*

For say $g \in L_2(\mathrm{I})$, $\|f\|_{\mathcal{Y}(\nu)'} \lesssim \frac{1}{\max(1,\nu)}(\|g\|_{L_2(\mathrm{I})} + |u_0|)$.

Proof. We verify the sufficient (and necessary) continuity, inf-sup and non-degeneracy conditions (2.5), (2.6) and (2.7).

The continuity follows from

$$|b(w,v)| \leq \|w\|_{L_2(\mathrm{I})}(\|\dot{v}\|_{L_2(\mathrm{I})} + \nu\|v\|_{L_2(\mathrm{I})}) \leq \|w\|_{\mathcal{X}}\sqrt{2}\,\|v\|_{\mathcal{Y}(\nu)}. \tag{7.3}$$

To show the inf-sup condition, given $v \in \mathcal{Y}(\nu)$, define $w = -\dot{v} + \nu v$. Then

$$b(w,v) = \int_{\mathrm{I}} \dot{v}(t)^2 - 2\nu v(t)\dot{v}(t) + \nu^2\, v(t)^2 dt \geq \|v\|^2_{\mathcal{Y}(\nu)} \geq \tfrac{1}{2}\sqrt{2}\,\|v\|_{\mathcal{Y}(\nu)}\|w\|_{\mathcal{X}}, \tag{7.4}$$

where we used that $\int_{\mathrm{I}} -2v(t)\dot{v}(t)dt = -\int_{\mathrm{I}} \frac{d}{dt}v(t)^2 dt = -v(T)^2 + v(0)^2 = v(0)^2 \geq 0$ and $\|w\|_{\mathcal{X}} \leq \sqrt{2}\,\|v\|_{\mathcal{Y}(\nu)}$.

For given $w \in \mathcal{X}$, define v by

$$\begin{cases} -\dot{v}(t) + \nu v(t) = w & (t \in \mathrm{I}), \\ v(T) = 0. \end{cases}$$

From

$$\int_{\mathrm{I}} w(t)v(t)dt = \int_0^T -\dot{v}(t)v(t) + \nu v(t)^2 dt = \tfrac{1}{2}v(0)^2 + \nu\int_{\mathrm{I}} v(t)^2 dt,$$

and the application of the Cauchy-Schwarz inequality to the left-hand side, we have $\nu\|v\|_{L_2(\mathrm{I})} \leq \|w\|_{L_2(\mathrm{I})}$ and so $\|\dot{v}\|_{L_2(\mathrm{I})} = \|w - \nu v\|_{L_2(\mathrm{I})} \leq 2\|w\|_{L_2(\mathrm{I})}$ or $\|v\|_{\mathcal{Y}(\nu)} \leq \sqrt{5}\,\|w\|_{\mathcal{X}}$, in particular meaning that $v \in \mathcal{Y}(\nu)$. From

$$b(w,v) = \int_{\mathrm{I}} -w(t)\dot{v}(t) + \nu w(t)v(t)dt = \int_{\mathrm{I}} w(t)^2 dt$$

we conclude (2.7).

The estimates (7.3) and (7.4) also give the bounds on the norms of B and B^{-1}.

The last statement follows from $|v(0)|^2 \lesssim \nu\|v\|^2_{L_2(\mathrm{I})} + \frac{1}{\nu}|v|^2_{H^1(\mathrm{I})}$ on $H^1(\mathrm{I})$, uniformly in $\nu > 0$, and for $\nu \in [0,1]$, from $\|\cdot\|_{L_2(\mathrm{I})} \lesssim |\cdot|_{H^1(\mathrm{I})}$ on $H^1_{0,\{T\}}(\mathrm{I})$. □

Let $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$ be the wavelet collections constructed in Sect. 6 normalized in $L_2(\mathrm{I})$. By equipping $\mathcal{X} = L_2(\mathrm{I})$ by $\Theta^{\mathcal{X}}$ and $\mathcal{Y}(\nu) = H^1_{0,\{T\}}(\mathrm{I})$ by $\Theta^{\mathcal{Y}}$, an equivalent matrix vector formulation of (7.2) reads as $\mathbf{Bu} = \mathbf{f}$, where

$$\mathbf{B} = \mathbf{D}_{\mathcal{Y}}^{-1}\big[-\langle\Theta^{\mathcal{X}}, \dot{\Theta}^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})} + \nu\langle\Theta^{\mathcal{X}}, \Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}\big], \tag{7.5}$$

$$\mathbf{f} = \mathbf{D}_{\mathcal{Y}}^{-1}\big[\int_{\mathrm{I}} g(t)\theta^{\mathcal{Y}}(t)dt + u_0\theta^{\mathcal{Y}}(0)\big]_{\theta^{\mathcal{Y}}\in\nabla^{\mathcal{Y}}}, \mathbf{D}_{\mathcal{Y}} = \mathrm{diag}\big\{\sqrt{\nu^2 + \|\dot{\theta}^{\mathcal{Y}}\|^2_{L_2(\mathrm{I})}} : \theta^{\mathcal{Y}} \in \nabla^{\mathcal{Y}}\big\}.$$

In our numerical experiments, we took $\nu = 1$ and $T = 1$. First we computed $\kappa((\mathbf{B}^\top\mathbf{B})_J)$ for levels $J \leq 8$, where $(\mathbf{B}^\top\mathbf{B})_J := [(\mathbf{B}^\top\mathbf{B})_{\lambda,\mu}]_{|\lambda|,|\mu|\leq J}$. The results are given in Table 7.1. For comparison, we also included condition numbers of $\mathbf{B}_J^\top\mathbf{B}_J$, where $\mathbf{B}_J = [\mathbf{B}_{\lambda,\mu}]_{|\lambda|,|\mu|\leq J}$ is the Galerkin matrix. Our results show that here the Galerkin approach, i.e., without forming the normal equations, is not applicable.

Remark 7.2. In any case in the non-adaptive setting, a valid alternative is to solve an overdetermined least squares problem $\mathrm{argmin}_{\mathbf{v}\in\ell_2(\Lambda^{\mathcal{X}})}\|(\mathbf{f} - \mathbf{Bv})|_{\Lambda^{\mathcal{Y}}}\|_{\ell_2(\Lambda^{\mathcal{Y}})}$

J	1	2	3	4	5	6	7	8
$\kappa((\mathbf{B}^\top\mathbf{B})_J)$	8.8	10.0	11.6	12.8	13.3	13.6	13.9	14.0
$\kappa(\mathbf{B}_J^\top\mathbf{B}_J)$	47	175	688	2719	10955	43807	175109	730250

TABLE 7.1
Condition numbers of $(\mathbf{B}^\top\mathbf{B})_J$ *and* $\mathbf{B}_J^\top\mathbf{B}_J$ *with* $\mathbf{B}$ *from* (7.5).

for some suitably chosen $\Lambda^{\mathcal{Y}} \subset \nabla^{\mathcal{Y}}$ with $\#\Lambda^{\mathcal{Y}} > \#\Lambda^{\mathcal{X}}$ (see [BJ89, And10]). This approach generalizes to the parabolic problem.

To obtain a solution that is not smooth, or piecewise smooth with respect to a dyadic refinement of $\mathrm{I} = (0,1)$, we took

$$g(t) = \begin{cases} 1 & \text{when } t \in (0, \frac{1}{3}), \\ 2 & \text{when } t \in (\frac{1}{3}, 1), \end{cases} \tag{7.6}$$

together with either $u_0 = 0$ or $u_0 = 1$.

With this choice, $\mathbf{f}_\lambda \neq 0$ only if $\frac{1}{3} \in \operatorname{supp} \theta_\lambda^{\mathcal{Y}}$ or $\theta_\lambda^{\mathcal{Y}}(0) \neq 0$, and, given an $\varepsilon > 0$, in the routine **RHS** we collect an, up to an absolute multiple, minimal number of such λ such that $\|\mathbf{f} - \mathbf{f}_\varepsilon\| \leq \varepsilon$.

We applied the adaptive wavelet-Galerkin method **AWGM** with parameters $\mu = 0.6$, $\gamma = 0.05$, $\theta = 0.5$, $\delta = 0.25$. These values do not respect the theoretical bounds we have derived, but turn out to be close to the optimal values for practical computations.

In Figure 7, for both choices $u_0 = 0$ and $u_0 = 1$, we illustrate the relative energy-norm errors $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ as function of the support length $\#\operatorname{supp} \mathbf{u}_\varepsilon$ of the computed approximate solution $\mathbf{u}_\varepsilon$. One may verify that

$$\frac{\|u - \mathbf{u}_\varepsilon^\top \Theta^{\mathcal{X}}\|_{L_2(\mathrm{I})}}{\|u\|_{L_2(\mathrm{I})}} \Big/ \frac{\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|}{\|\mathbf{f}\|} \in [\alpha^{-1}, \alpha], \text{ where } \alpha = \|B\|\|B^{-1}\| \frac{\Lambda_{\mathcal{Y}(1)}(|\Theta^{\mathcal{Y}}|)}{\lambda_{\mathcal{Y}(1)}(|\Theta^{\mathcal{Y}}|)}$$

and where α can be further estimated similarly to (3.4). The results illustrate that for both initial values, the practical convergence rates approach the theoretical asymptotic rate -5 indicated by the slope of the hypotenuse of the triangle.

In Figure 7, for the case of $u_0 = 1$ and $\#\mathbf{u}_\varepsilon = 202$, the location as well as the modulus of the non-zero wavelet coefficients are indicated.

Finally in this section, using the result of Theorem 7.1, for time-independent, symmetric and coercive spatial operators, we show that the results of Theorems 2.2 and 2.4 about boundedly invertibility of the parabolic problem can be supplemented with quantitative bounds. A similar analysis was carried out in [BJ89].

THEOREM 7.3. *In the setting of Sect. 2, let* $a(t;\eta,\zeta) = a(\eta,\zeta) = a(\zeta,\eta)$, $\lambda_0 = 0$ *and let* $V \hookrightarrow H$ *be compact. Then with either*

$$\mathcal{X} = L_2(\mathrm{I}; V), \quad \mathcal{Y} = L_2(\mathrm{I}; V) \cap H^1_{0,\{T\}}(\mathrm{I}; V'),$$

where V *is equipped with norm* $\sqrt{a(\cdot,\cdot)}$ *and* $H^1_{0,\{T\}}(\mathrm{I})$ *with* $|\cdot|_{H^1(\mathrm{I})}$, *or*

$$\mathcal{X} = L_2(\mathrm{I}; H), \quad \mathcal{Y} = L_2(\mathrm{I}; D(A)) \cap H^1_{0,\{T\}}(\mathrm{I}; H),$$

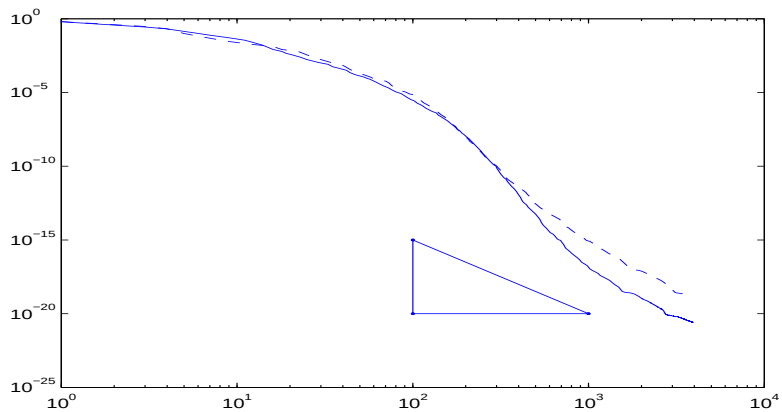

FIG. 7.1. *The* **AWGM** *applied to the ODE from* (7.1) *with right-hand side* g *from* (7.6). $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ *vs.* $\#\operatorname{supp}\mathbf{u}_\varepsilon$ *for* $u_0 = 1$ *(solid lines) and* $u_0 = 0$ *(dashed lines).*

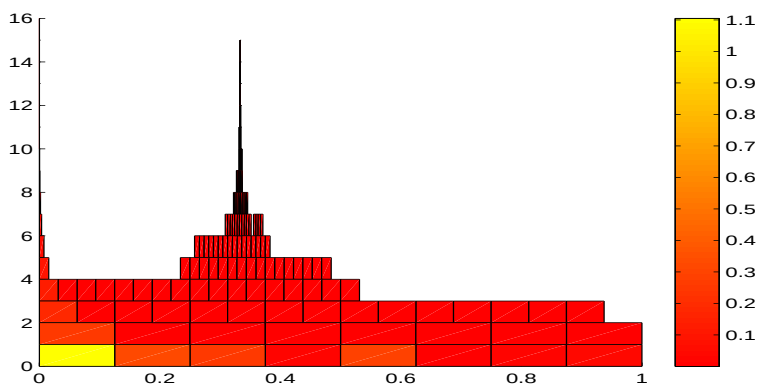

FIG. 7.2. *For* $u_0 = 1$ *and* $\#\mathbf{u}_\varepsilon = 202$*, the non-zero coefficients of* $\mathbf{u}_\varepsilon$*. Non-zero coefficients corresponding to the left boundary wavelet run to level* 51*, so far outside the scale of this picture.*

where $D(A) = \{w \in H : Aw \in H\}$ *is equipped with* $\|A\cdot\|_H$ *and, again,* $H^1_{0,\{T\}}(\mathrm{I})$ *with* $|\cdot|_{H^1(\mathrm{I})}$*,* $B : \mathcal{X} \to \mathcal{Y}'$ *is boundedly invertible with* $\|B\| \le \sqrt{2}$ *and* $\|B^{-1}\| \le \sqrt{2}$*.*

Proof. Let $\{\varphi\}$ be an orthonormal basis for H of eigenfunctions of A with eigenvalues λ_φ (e.g. see [DL90, Ch.VIII, §2.6, Th.7]). Then $\{\frac{\varphi}{\sqrt{\lambda_\varphi}}\}$ and $\{\frac{\varphi}{\lambda_\varphi}\}$ are orthonormal bases for V and $D(A)$ equipped with $\sqrt{a(\cdot,\cdot)}$ and $\|A\cdot\|$, respectively.

Writing $u = \sum_\varphi u_\varphi(t) \otimes \varphi$, $f = \sum_\varphi f_\varphi(t) \otimes \varphi$, $v = \sum_\varphi v_\varphi(t) \otimes \varphi$, finding $u \in \mathcal{X}$ such that for $f \in \mathcal{Y}'$,

$$b(u, v) = f(v) \quad (v \in \mathcal{Y}),$$

is equivalent to

$$\int_{\mathrm{I}} -u_\varphi \dot{v}_\varphi + \lambda u_\varphi v_\varphi \, dt = \int_{\mathrm{I}} f_\varphi v_\varphi \, dt \quad (\varphi \in \{\varphi\},\ v_\varphi \in L_2(\mathrm{I})).$$

We have $\|u\|^2_{L_2(I;H)} = \sum_\varphi \|u_\varphi\|^2_{L_2(\mathrm{I})}$, $\|u\|^2_{L_2(I;V)} = \sum_\varphi \lambda_\varphi \|u_\varphi\|^2_{L_2(\mathrm{I})}$, $\|u\|^2_{L_2(I;D(A))} =$

$\sum_\varphi \lambda_\varphi^2 \|u_\varphi\|^2_{L_2(\mathrm{I})}$,

$$\|v\|^2_{L_2(\mathrm{I};D(A))\cap H^1_{0,\{T\}}(\mathrm{I};H)} = \sum_\varphi \lambda_\varphi^2 \|v_\varphi\|^2_{L_2(\mathrm{I})} + |v_\varphi|^2_{H^1(\mathrm{I})} = \sum_\varphi \|v_\varphi\|^2_{\mathcal{Y}(\lambda_\varphi)},$$

where $\mathcal{Y}(\nu)$ is as Theorem 7.1,

$$\|v\|^2_{L_2(\mathrm{I};V)\cap H^1_{0,\{T\}}(\mathrm{I};V')} = \sum_\varphi \lambda_\varphi \|v_\varphi\|^2_{L_2(\mathrm{I})} + \lambda_\varphi^{-1}|v_\varphi|^2_{H^1(\mathrm{I})} = \sum_\varphi \lambda_\varphi^{-1}\|v_\varphi\|^2_{\mathcal{Y}(\lambda_\varphi)}$$

and so

$$\|f\|_{(L_2(\mathrm{I};V)\cap H^1_{0,\{T\}}(\mathrm{I};V'))'} = \sum_\varphi \lambda_\varphi \|f_\varphi\|^2_{(\mathcal{Y}(\lambda_\varphi))'},$$
$$\|f\|_{(L_2(\mathrm{I};D(A))\cap H^1_{0,\{T\}}(\mathrm{I};H))'} = \sum_\varphi \|f_\varphi\|^2_{(\mathcal{Y}(\lambda_\varphi))'}.$$

Theorem 7.1 shows that $\frac{1}{2}\sqrt{2}\|f_\varphi\|_{(\mathcal{Y}(\lambda_\varphi))'} \le \|u_\varphi\|_{L_2(\mathrm{I})} \le \sqrt{2}\|f_\varphi\|_{(\mathcal{Y}(\lambda_\varphi))'}$. By summing these inequalities over φ, the proof is completed. □

8. Numerical results for the heat equation. With $\Omega := (0,1)^n$, we consider the heat equation

$$\tfrac{\partial}{\partial t}u - \Delta_{\mathbf{x}} u = g \text{ on } \mathrm{I}\times\Omega,\ u = 0 \text{ on } \mathrm{I}\times\partial\Omega,\ u(0,\cdot) = u_0.$$

Since Ω is convex with Lipschitz continuous boundary, it is well-known that $-\Delta_{\mathbf{x}} : H^2(\Omega)\cap H^1_0(\Omega) = D(-\Delta_{\mathbf{x}}) \to L_2(\Omega)$ is boundedly invertible. As exposed in Sect. 2, in particular in Theorem 2.4, a well-posed space-time variational formulation of this equation is to find $u \in \mathcal{X} = L_2(\mathrm{I}; L_2(\Omega))$ such that

$$(Bu)(v) = b(u,v) = f(v) \quad (v \in \mathcal{Y} = L_2(\mathrm{I}; H^2(\Omega)\cap H^1_0(\Omega)) \cap H^1_{0,\{T\}}(\mathrm{I}; L_2(\Omega))),$$

where

$$b(w,v) = \int_{\mathrm{I}}\int_\Omega -w(x,t)\tfrac{\partial v}{\partial t}(t,x) + \nabla_x w(x,t)\cdot\nabla_x v(x,t)dxdt,$$
$$f(v) = \int_{\mathrm{I}}\int_\Omega g(t,x)v(t,x)dxdt + \int_\Omega u_0(x)v(0,x)dx,$$

assuming that $f \in \mathcal{Y}'$.

Concerning the latter, since $[L_2(\Omega), H^2(\Omega)\cap H^1_0(\Omega)]_{1/2} = H^1_0(\Omega)$, for $v \in \mathcal{Y}$, $v(0) \in H^1_0(\Omega)$ with $\|v(0)\|_{H^1(\Omega)} \lesssim \|v\|_{\mathcal{Y}}$. So similar to Remark 2.3, if $u_0 \in H^{-1}(\Omega)$, and, say, $g \in L_2(\mathrm{I}; (H^2(\Omega)\cap H^1_0(\Omega))')$, then $f \in \mathcal{Y}'$ with $\|f\|_{\mathcal{Y}'} \lesssim \|u_0\|_{H^{-1}(\Omega)} + \|g\|_{L_2(\mathrm{I};(H^2(\Omega)\cap H^1_0(\Omega))')}$.

By Theorem 7.3, for $H^2(\Omega)\cap H^1_0(\Omega)$ being equipped with $\|-\Delta_{\mathbf{x}}\cdot\|_{L_2(\Omega)}$, it holds that $\|B\|_{\mathcal{X}\to\mathcal{Y}'} \le \sqrt{2}$ and $\|B^{-1}\|_{\mathcal{Y}'\to\mathcal{X}} \le \sqrt{2}$. Below, we will equip $H^2(\Omega)\cap H^1_0(\Omega)$ with norm $(\sum_{i=1}^n \|\partial_i^2\cdot\|^2_{L_2(\Omega)})^{\frac{1}{2}}$, being the norm on the intersection space

$$(H^2\cap H^1_0)\otimes L_2\otimes\cdots\otimes L_2 \cap \ \cdots\ \cap L_2\otimes\cdots\otimes L_2\otimes(H^2\cap H^1_0)$$

where $L_2 = L_2(0,1)$ and $H^2\cap H^1_0 = H^2(0,1)\cap H^1_0(0,1)$ equipped with $|\cdot|_{H^2(0,1)}$. Using the eigenvector basis of ∂_i, a straightforward calculation shows that $\sum_{i=1}^n \|\partial_i^2\cdot\|^2_{L_2(\Omega)} \le \|-\Delta_{\mathbf{x}}\cdot\|^2_{L_2(\Omega)} \le n\sum_{i=1}^n \|\partial_i^2\cdot\|^2_{L_2(\Omega)}$ on $H^2(\Omega)\cap H^1_0(\Omega)$, so that with the new norm,

$$\|B\|_{\mathcal{X}\to\mathcal{Y}'} \le \sqrt{2n} \quad\text{and}\quad \|B^{-1}\|_{\mathcal{Y}'\to\mathcal{X}} \le \sqrt{2}. \tag{8.1}$$

Let $\Sigma^{(1)}$ be the collection of quartic Hermite wavelets from [CS10]. Normalized in the corresponding norms, it is a Riesz basis for $[L_2(0,1), H^5(0,1)\cap H^1_0(0,1)]_{s/5}$ ($s\in[0,\frac{5}{2})$), and for $H^{-s}(0,1)'$ ($s\in(-\frac{1}{2},0]$), and so in particular for $L_2(0,1)$ and $H^2(0,1)\cap H^1_0(0,1)$. This collection was designed such that the bi-infinite matrices $\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}$, $\langle\dot\Sigma^{(1)},\dot\Sigma^{(1)}\rangle_{L_2(0,1)}$, $\langle\dot\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}$, and $\langle\Sigma^{(1)},\dot\Sigma^{(1)}\rangle_{L_2(0,1)}$ are all truly sparse.

With this collection $\Sigma^{(1)}$, and $\Theta^{\mathcal{X}}$, $\Theta^{\mathcal{Y}}$ from Sect. 6, all L_2-normalized, we equip $\mathcal{X}=L_2(\mathrm{I};L_2(\Omega))$ with Riesz basis $\Theta^{\mathcal{X}}\otimes\Sigma^{(1)}\otimes\cdots\otimes\Sigma^{(1)}$, and $\mathcal{Y}=L_2(\mathrm{I};H^2(\Omega)\cap H^1_0(\Omega))\cap H^1_{0,\{T\}}(\mathrm{I};L_2(\Omega))$ with $\Theta^{\mathcal{Y}}\otimes\Sigma^{(1)}\otimes\cdots\otimes\Sigma^{(1)}$ being, normalized in $\mathcal{Y}$, a Riesz basis for that space.

By combining (3.2), (8.1), and estimates analogous to (3.4), (3.5), one may verify that

$$\begin{aligned}\|\mathbf{B}\| &\le \sqrt{2n}\,\Lambda_{L_2(\mathrm{I})}(\Theta^{\mathcal{X}})[\Lambda_{L_2(0,1)}(\Sigma^{(1)})]^{2n-1}\times\\ &\qquad \max\Big(\Lambda_{L_2(\mathrm{I})}(\Theta^{\mathcal{Y}})\Lambda_{H^2(0,1)\cap H^1_0(0,1)}(|\Sigma^{(1)}|),\Lambda_{H^1_{0,\{T\}}(\mathrm{I})}(|\Theta^{\mathcal{Y}}|)\Lambda_{L_2(0,1)}(\Sigma^{(1)})\Big),\\ \|\mathbf{B}^{-1}\|^{-1} &\ge \tfrac12\sqrt{2}\,\lambda_{L_2(\mathrm{I})}(\Theta^{\mathcal{X}})[\lambda_{L_2(0,1)}(\Sigma^{(1)})]^{2n-1}\times\\ &\qquad \min\Big(\lambda_{L_2(\mathrm{I})}(\Theta^{\mathcal{Y}})\lambda_{H^2(0,1)\cap H^1_0(0,1)}(|\Sigma^{(1)}|),\lambda_{H^1_{0,\{T\}}(\mathrm{I})}(|\Theta^{\mathcal{Y}}|)\lambda_{L_2(0,1)}(\Sigma^{(1)})\Big),\end{aligned}$$

In view of these estimates, it is very favourable to use $L_2(0,1)$-orthonormal wavelets. In the present paper, however, we did not follow this approach because we preferred to have truly sparse stiffness matrices.

With $\mathbf{u}$ denoting the coefficient vector of u with respect to $\Theta^{\mathcal{X}}\otimes\Sigma^{(1)}\otimes\cdots\otimes\Sigma^{(1)}$, the matrix-vector representation of the variational problem reads as $\mathbf{Bu}=\mathbf{f}$, where

$$\begin{aligned}\mathbf{B} &= \mathbf{D}_{\mathcal{Y}}^{-1}\Big[-\langle\Theta^{\mathcal{X}},\dot\Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}\otimes\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}\otimes\cdots\otimes\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}+\\ &\quad\langle\Theta^{\mathcal{X}},\Theta^{\mathcal{Y}}\rangle_{L_2(\mathrm{I})}\otimes\Big\{\langle\dot\Sigma^{(1)},\dot\Sigma^{(1)}\rangle_{L_2(0,1)}\otimes\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}\otimes\cdots\otimes\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}+\\ &\qquad\qquad\vdots\\ &\qquad\qquad\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}\otimes\cdots\otimes\langle\Sigma^{(1)},\Sigma^{(1)}\rangle_{L_2(0,1)}\otimes\langle\dot\Sigma^{(1)},\dot\Sigma^{(1)}\rangle_{L_2(0,1)}\Big\}\Big],\\ \mathbf{f} &= \mathbf{D}_{\mathcal{Y}}^{-1}\Big[\int_{\mathrm{I}}\int_{\Omega}g(t,x)\theta^{\mathcal{Y}}(t)\sigma_1^{(1)}(x_1)\cdots\sigma_n^{(1)}(x_n)dtdx+\\ &\qquad\qquad\int_{\Omega}u_0(x)\theta^{\mathcal{Y}}(0)\sigma_1^{(1)}(x_1)\cdots\sigma_n^{(1)}(x_n)dx\Big]_{\theta^{\mathcal{Y}}\in\Theta^{\mathcal{Y}},\sigma_1^{(1)},\ldots,\sigma_n^{(1)}\in\Sigma^{(1)}},\\ \mathbf{D}_{\mathcal{Y}} &= \mathrm{diag}\Big\{\sqrt{\sum_{i=1}^n\|\ddot\sigma_i^{(1)}\|^2_{L_2(0,1)}+\|\dot\theta^{\mathcal{Y}}\|^2_{L_2(\mathrm{I})}}:\sigma_1^{(1)},\ldots,\sigma_n^{(1)}\in\Sigma^{(1)},\theta^{\mathcal{Y}}\in\Theta^{\mathcal{Y}}\Big\}.\end{aligned}$$

In Figure 8.1, we give the numerical results obtained with the **AWGM** for $n=1$, $g=1$ and $u_0=0$. As parameters for the heat equation in one and two spatial dimensions, we took $\mu=\frac12$, $\gamma=\frac{1}{64}$, $\theta=\frac12$, $\delta=\frac14$. For comparison, we included corresponding results obtained with the non-adaptive full and sparse-grids methods, i.e., Galerkin approximations to $\mathbf{B}^\top\mathbf{Bu}=\mathbf{B}^\top\mathbf{f}$ from the span of sets of the form $\{\theta^{\mathcal{X}}_\lambda\sigma^{(1)}_{\mu^{(1)}}\cdots\sigma^{(1)}_{\mu^{(n)}}:|\lambda|,|\mu^{(1)}|,\ldots,|\mu^{(n)}|\le J\}$ or $\{\theta^{\mathcal{X}}_\lambda\sigma^{(1)}_{\mu^{(1)}}\cdots\sigma^{(1)}_{\mu^{(n)}}:|\lambda|+|\mu^{(1)}|+\ldots+|\mu^{(n)}|\le J\}$, respectively. The results show that the adaptive method converges with the best possible rate $N^{-5}(\log N)^{5\frac12}$, cf. (4.3), and that it outperforms the non-adaptive methods with orders of magnitude.

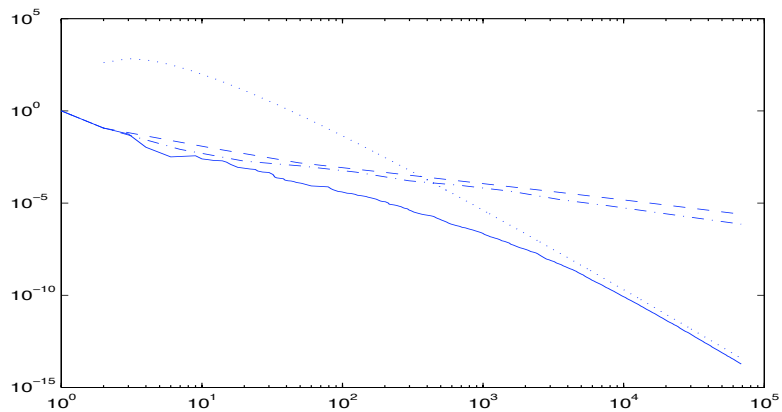

FIG. 8.1. *Heat eqn. in $n = 1$ spatial dimension, right-hand side $g = 1$ and initial condition $u_0 = 0$. $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ vs. $N = \#\operatorname{supp}\mathbf{u}_\varepsilon$ for the* **AWGM** *(solid), full-grid (dashed) and sparse-grid method (dashed-dotted). The dotted line is a multiple of $N^{-5}(\log N)^{5\frac{1}{2}}$.*

In Figure 8.2, we give the numerical results obtained with the **AWGM** for $n = 1$, $g = 1$ and $u_0 = 1$. Due to the fact that the primal temporal wavelets all vanish at 0, a very strong local refinement near $t = 0$ is necessary. For this reason, in this case we did not make a comparison with the non-adaptive methods.

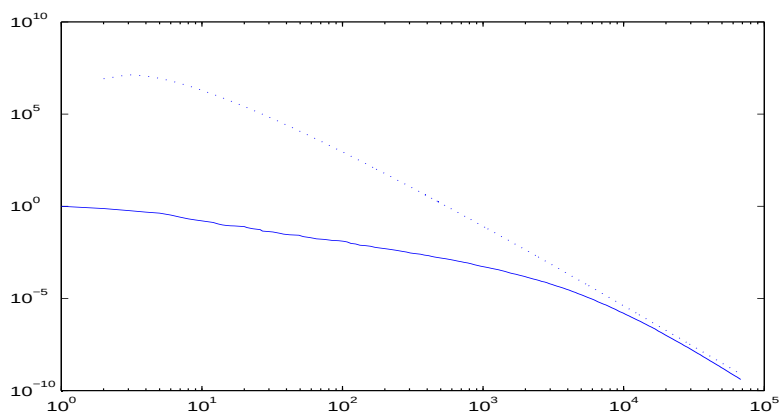

FIG. 8.2. **AWGM** *applied to heat eqn. in $n = 1$ spatial dimension, right-hand side $g = 1$ and initial condition $u_0 = 1$. $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ vs. $\#\operatorname{supp}\mathbf{u}_\varepsilon$. The dotted line is a multiple of $N^{-5}(\log N)^{5\frac{1}{2}}$.*

Figure 8.3 illustrates which tensor product wavelets were selected by the **AWGM**. Note the strong refinement near $t = 0$, in particular near the corners of the space-time cylinder.

In Figure 8.4, we give the numerical results obtained with the **AWGM** for $n = 2$, $g = 1$ and $u_0 = 0$. Although also here the **AWGM** outperforms the sparse grid and full grid methods (the final computed error is approximately a factor 300 or 2500

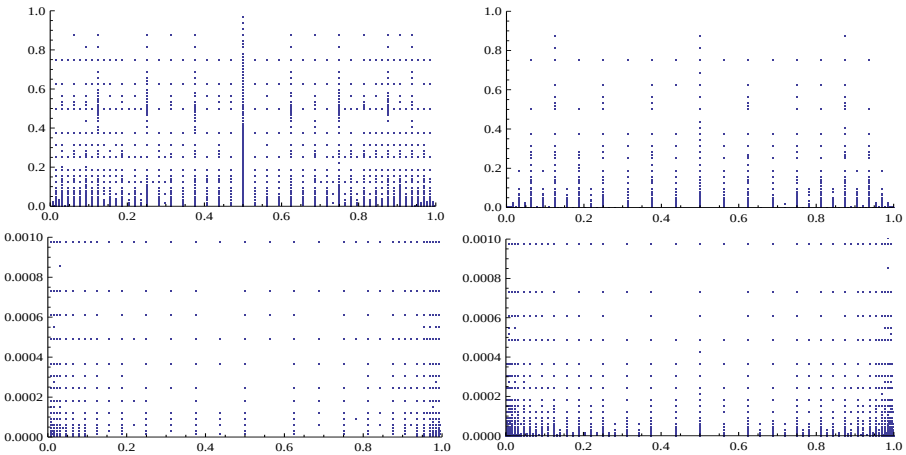

FIG. 8.3. *Heat eqn. in $n = 1$ spatial dimension and right-hand side $g = 1$. Centers of the supports of the wavelets selected by the* **AWGM**. *Left $u_0 = 0$ and $\#\mathbf{u}_\varepsilon = 13420$. Right $u_0 = 1$ and $\#\mathbf{u}_\varepsilon = 13917$. A zoom in near $t = 0$ is given at the bottom row.*

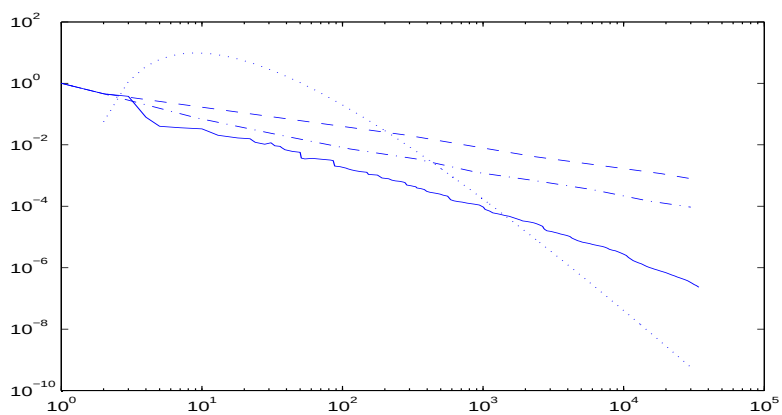

FIG. 8.4. *Heat eqn. in $n = 2$ spatial dimensions, right-hand side $g = 1$ and initial condition $u_0 = 0$. $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ vs. $N = \#\operatorname{supp}\mathbf{u}_\varepsilon$ for the* **AWGM** *(solid), full-grid (dashed) and sparse-grid method (dashed-dotted). The dotted line is a multiple of $N^{-5}(\log N)^{11}$.*

smaller), we do not observe the maximal possible rate $N^{-5}(\log N)^{11}$, cf. (4.3). This could mean either that for this problem, this rate does not hold because the solution has not enough smoothness in the scale of relevant tensor product Besov spaces (see [Nit06, SU09]), or that this rate shows up only with larger support sizes that, with our current implementation, cannot be reached on a PC. Indeed, it is generally observed that for adaptive (wavelet) approximations of higher order, as with our approximations of order 5, the asymptotic rate shows up later than with lower order approximations. For a solution that has singularities, often initially the errors are even worse than with lower order approximations. A possible explanation is that with higher order approximations, the dual wavelets have larger supports. Furthermore, although (adaptive) tensor product approximations are shown to converge with dimension independent asymptotic rates, with an increasing dimension it takes longer

before these rates become visible. Likely, the combination of both effect is the reason that in our results in two spatial dimensions, the maximal possible rate is yet not visible.

Finally, we tested the **AWGM** on the heat equation in $n = 2$ spatial dimensions with right-hand side $g(t, \mathbf{x}) = t^4 x_1(x_1 - 1)(x_1^2 - x_1 - 1)x_2(x_2 - 1)(x_2^2 - x_2 - 1)$. Using that g satisfies homogeneous boundary conditions of order 5 at $t = 0$, both $\partial_t^5 g(0, \cdot)$ and $\Delta_{\mathbf{x}} \partial_t^5 g(0, \cdot)$ are in $H_0^1(\Omega) \cap H^2(\Omega)$, and g is symmetric in $x = y$ and $y = 1 - x$, one can verify that the solution $u \in H^5(0, 1) \otimes H^5(0, 1) \otimes H^5(0, 1)$. As a consequence, the error in the sparse grid approximation and thus also in the adaptive wavelet approximation of length N is of the best possible order $N^{-5}(\log N)^{11}$, and that in the full grid approximation of length N is of the best possible order $N^{-5/3}$. As shown in Figure 8.5, these asymptotic results are confirmed by the numerical results. The oscillations in the curve for the **AWGM** and the fact that initially it lies above the one for the sparse grid method can be cured by solving the Galerkin systems more accurately.

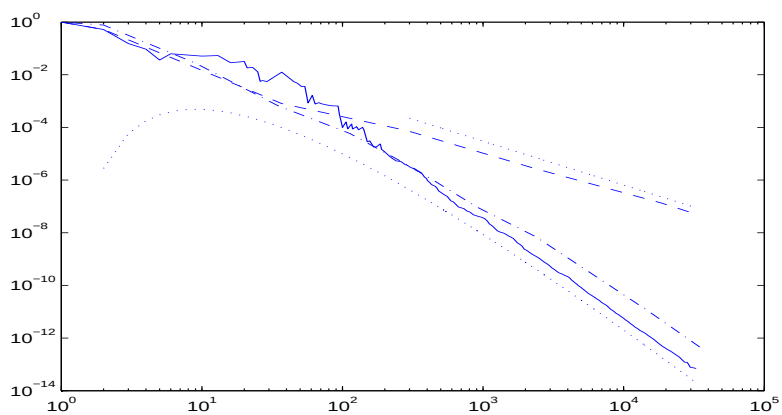

FIG. 8.5. *Heat eqn. in $n = 2$ spatial dimensions, right-hand side $g(t, \mathbf{x}) = t^4 x_1(x_1 - 1)(x_1^2 - x_1 - 1)x_2(x_2 - 1)(x_2^2 - x_2 - 1)$ and initial condition $u_0 = 0$. $\|\mathbf{B}\mathbf{u}_\varepsilon - \mathbf{f}\|/\|\mathbf{f}\|$ vs. $N = \#\operatorname{supp} \mathbf{u}_\varepsilon$ for the* **AWGM** *(solid), full-grid (dashed) and sparse-grid method (dashed-dotted). The dotted lines are multiples of $N^{-5}(\log N)^{11}$ and $N^{-5/3}$, respectively.*

Acknowledgments. The authors like to thank Professor Christoph Schwab (Zürich) for useful discussions on various topics related to this paper, and Professor Jan Prüss (Halle) for his help with Theorem 2.4.

REFERENCES

[And10] R. Andreev. Space-time wavelet fem for parabolic equations. SAM report 2010-20, ETH Zürich, 2010.

[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.

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

[CDD02] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[CDP96] J.M. Carnicer, W. Dahmen, and J.M. Peña. Local decomposition of refinable spaces and wavelets. *Appl. and Comp. Harm. Anal.*, 3:127–153, 1996.

[Coh00] A. Cohen. Wavelet methods in numerical analysis. In P.G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis. Vol. VII.*, pages 417–711. North-Holland, Amsterdam, 2000.

[Coh03] A. Cohen. *Numerical Analysis of Wavelet Methods*. Elsevier, Amsterdam, 2003.

[CS10] N.G. Chegini and R.P. Stevenson. The adaptive tensor product wavelet scheme: Sparse matrices and the application to singularly perturbed problems. Technical report, Korteweg - de Vries Institute, University of Amsterdam, 2010. Submitted.

[Dah96] W. Dahmen. Stability of multiscale transformations. *J. Fourier Anal. Appl.*, 2(4):341–362, 1996.

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

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

[DL90] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Vol. 3*. Springer-Verlag, Berlin, 1990. Spectral theory and applications, With the collaboration of Michel Artola and Michel Cessenat, Translated from the French by John C. Amson.

[dS64] L. de Simon. Un'applicazione della teoria degli integrali singolari allo studio delle equazioni differenziali lineari astratte del primo ordine. *Rend. Sem. Mat. Univ. Padova*, 34:205–223, 1964.

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

[DS10a] M. Dauge and R.P. Stevenson. Sparse tensor product wavelet approximation of singular functions. *SIAM J. Math. Anal.*, 42(5):2203–2228, 2010.

[DS10b] T.J. Dijkema and R.P. Stevenson. A sparse Laplacian in tensor product wavelet coordinates. *Numer. Math.*, 115(3):433–449, 2010.

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

[GK00] M. Griebel and S. Knapek. Optimized tensor-product approximation spaces. *Constr. Approx.*, 16(4):525–540, 2000.

[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.

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

[PS01] J. Prüss and R. Schnaubelt. Solvability and maximal regularity of parabolic evolution equations with coefficients continuous in time. *J. Math. Anal. Appl.*, 256(2):405–430, 2001.

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

[Ste03] R.P. Stevenson. Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes. *Constr. Approx.*, 19(4):477–508, 2003.

[Ste09] R.P. Stevenson. Adaptive wavelet methods for solving operator equations: An overview. In R.A. DeVore and A. Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*, pages 543–598. Springer, 2009.

[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.

[Woo07] I. Wood. Maximal L^p-regularity for the Laplacian on Lipschitz domains. *Math. Z.*, 255(4):855–875, 2007.