Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations

R. Andreev

RICAM-Report 2014-27

WAVELET-IN-TIME MULTIGRID-IN-SPACE PRECONDITIONING OF PARABOLIC EVOLUTION EQUATIONS

ROMAN ANDREEV*

Abstract. Two different space-time variational formulations of linear parabolic evolution equations are considered, one is symmetric and elliptic on the trial space while the other is not. In each case, space-time Petrov–Galerkin discretization using suitable tensor product trial and test functions leads to a large linear system of equations. Well-posedness of this system with respect to space-time norms induces a canonical preconditioner for the algebraic equations that arise after a choice of basis. For the iterative resolution of this algebraic system with parallelization in the temporal direction we propose a sparse algebraic wavelet-in-time transformation on possibly nonuniform temporal meshes. This transformation block-diagonalizes the preconditioner, and the individual spatial blocks can then be inverted approximately by standard spatial multigrid methods in parallel. We document the performance of the preconditioner in a series of numerical experiments.

Key words. space-time, parabolic, discretization, preconditioning, wavelet, multigrid

AMS subject classifications. 35K90, 65M12, 65M20, 65M55, 65M60, 65Y05

1. Introduction and model problem.

1.1. Introduction. We present a preconditioner for space-time simultaneous Petrov–Galerkin discretizations of linear parabolic evolution equations that is suitable for large scale parallel-in-time computation. The basic rationale for the design of the preconditioner is the ability to use existing spatial finite element codes and preconditioners for elliptic problems, such as multigrid. In the context of large scale parallel computation we therefore stipulate that the space-time solution vector is distributed across computing nodes along the temporal dimension so that a node, or a group of nodes, each hold only a few temporal snapshots of the solution out of many. By many we mean hundreds or thousands of snapshots. Then, space-time simultaneous iterative computation asks for the reduction of communication between these groups of nodes as far as possible, in particular when it comes to preconditioning. This will be achieved by a transformation to a temporal multiscale basis.

We consider two different space-time variational formulations of the linear parabolic evolution equation (Section 1.3). The key difference between them is that one is symmetric and elliptic while the other is not. The lack of symmetry has implications on the choice of discrete trial and test spaces and on the choice of the iterative solver (Sections 2.1–2.4). The symmetric formulation, on the other hand, involves the inverse of the spatial operator, which renders its numerical implementation more challenging. However, we will show that both formulations admit the same canonical preconditioner induced by the continuous norm on the trial space (Section 2.4). Throughout, we assume the regime of small Péclet numbers. That is we assume that the symmetric part of the generator is elliptic, and the asymmetric part is relatively small.

As we shall explain (Section 3.2.2), the space-time preconditioner involves the construction of a temporal piecewise polynomial multiscale basis with compact support which can be rescaled to a Riesz basis both in the Lebesgue space L_2 and in the Sobolev space H^1. Many such bases of wavelet type are known, but their construction is typically based on dyadic refinement, posing restrictions on the temporal mesh. The

*RICAM, Altenberger Str. 69, 4040 Linz, Austria (roman.andreev@oeaw.ac.at)

classical fast Fourier transform suffers from the same drawback. For piecewise linear spline spaces we provide an algebraic algorithm that works on arbitrary temporal meshes (Section 3.3). The algorithm proceeds as follows: a) identify basis functions in the top frequency band; b) approximately orthogonalize to those of lower frequency; c) split off the resulting detail space; d) repeat. Approximate iterative orthogonalization in the second step allows to produce compactly supported wavelet-like functions. Unlike often the case with wavelet transforms, due to the purely algebraic approach there is no difficulty associated with the domain being bounded. Extensions to splines of higher polynomial degree are of interest but will not be discussed here.

The sparsity of the multiscale transformation to the wavelet-like basis in the temporal variable is *crucial* for reducing the inter-process communication cost. Moreover, the Riesz basis property simultaneously in L_2 and H^1 allow to transform and approximate the nonlocal canonical preconditioner to a block-diagonal one, where each block corresponds to a spatial Helmholtz problem with an imaginary frequency. For the purpose of preconditioning, this block can be replaced by multigrid versions of positive definite Helmholtz problems with real frequency while maintaining control over the quality of the preconditioner (Section 3.2). This block-diagonal preconditioner can then be applied in parallel. We emphasize that the parallelization is performed in the temporal direction, but the effect of the multiscale wavelet-like transformation in time is such that the individual spatial blocks of the preconditioner do not correspond to small temporal subintervals, rather to wave packets of different temporal scale.

In Section 4 we present our numerical experiments and conclude in Section 5.

1.2. Problem class. Let V and H be real separable Hilbert spaces with continuous and dense embedding $V \hookrightarrow H$. The Hilbert space H is identified with its (continuous) dual H' by the Riesz isomorphism. This results in the Gelfand triple $V \hookrightarrow H \cong H' \hookrightarrow V'$ with continuous and dense embeddings. We write $\|\cdot\|_V$ for the norm on V. The scalar product on H and (its continuous extension to) the duality pairing between V and V' is denoted by $(\cdot,\cdot)$. For the duality pairing on other Banach space we write $\langle\cdot,\cdot\rangle$. We write $L_2(J;V')$ for the Bochner space of V'-measurable functions on J, and $H^1(J;V')$ for the Bochner–Sobolev subspace of weakly differentiable functions w on J with derivative $d_t w$ in $L_2(J;V')$, see [11] or [14, Chapter 1]. Similar notation is employed for other instances of Hilbert spaces.

Let $J = (0,T)$ be a nonempty bounded interval. Let $A(t) : V \to V'$, (a.e.) $t \in J$, be a family of bounded linear operators. We assume that, for some fixed constants $C > 0$, $\alpha > 0$, and $\gamma_0 \geq 0$, the family $A(t)$, $t \in J$, satisfies the conditions

$$\text{a) For all } \chi,\tilde{\chi} \in V, \text{ the mapping } t \mapsto (A(t)\chi,\tilde{\chi}) \text{ is Lebesgue measurable.} \tag{1}$$

$$\text{b) } |(A(t)\chi,\tilde{\chi})| \leq C\|\chi\|_V\|\tilde{\chi}\|_V \text{ for all } \chi,\tilde{\chi} \in V \text{ and (a.e.) } t \in J. \tag{2}$$

$$\text{c) } (A(t)\chi,\chi) \geq \alpha^2\|\chi\|_V^2 - \gamma_0^2\|\chi\|_H^2 \text{ for all } \chi \in V \text{ and (a.e.) } t \in J. \tag{3}$$

The conditions imply that $A+\gamma_0^2\iota$ can be interpreted as an isomorphism from $L_2(J;V)$ onto its dual (identified with) $L_2(J;V')$, where ι is the embedding $V \hookrightarrow V'$.

Let $f \in L_2(J;V')$ and $g \in H$ be given. The subject of this paper is the abstract linear parabolic evolution equation

$$u(0) = g, \quad d_t u(t) + A(t)u(t) = f(t) \quad \text{(a.e.) } t \in J, \tag{4}$$

where the equalities are understood in H and V', respectively. Equivalently, one could write $d_t u + Au = f$ with equality in $L_2(J;V')$, where Au stands for the map $t \mapsto A(t)u(t)$, $t \in J$.

In order to make the concept of a solution precise, we introduce the Sobolev–Bochner space

$$X := H^1(J;V') \cap L_2(J;V). \tag{5}$$

A norm on X is given by

$$|||w|||_X := \left[\|d_t w\|^2_{H^1(J;V')} + \|w\|^2_{L_2(J;V')}\right]^{1/2}, \quad w \in X. \tag{6}$$

We recall the continuous embedding $X \hookrightarrow C^0([0,T];H)$. In particular, $w(t_0)$ is well-defined in H for any $w \in X$ and $0 \le t_0 \le T$, and $w \mapsto w(t_0)$ is continuous, see [11, Section 5.9.2] or [14, Chapter 1]. An equivalent norm (with constants depending on T) is therefore given by

$$\|w\|^2_X := |||w|||^2_X + \|w(T)\|^2_H, \quad w \in X, \tag{7}$$

and henceforth, X is understood to be equipped with this norm. With the obvious corresponding inner product, it is a Hilbert space. The choice of the norm can be motivated as follows. Suppose A is the embedding $V \hookrightarrow V'$. Then $\|d_t u + Au\|^2_{L_2(J;V')} = \|u\|^2_X - \|u(0)\|^2_H$ can be easily verified by expanding the square and using integration by parts. Therefore $\|u\|^2_X = \|f\|^2_{L_2(J;V')} + \|g\|^2_H$ if u satisfies (4).

Under the stated assumptions on A and the data f and g, there exists a unique solution $u \in X$ to (4), and it depends continuously on the data [14, Chapter 3, Section 4.7].

In view of the usual transformation $u \mapsto v := [t \mapsto u(t)e^{-\gamma_0^2 t}]$, which is an automorphism on X, we assume from now on without loss of generality that $\gamma_0 = 0$.

1.3. Space-time variational formulations. Our first space-time variational is from [18]. It is based on testing $d_t u + Au = f$ by test functions $v_1 \in L_2(J;V')$ and appending the initial condition $u(0) = g$ using a Lagrange multiplier $v_0 \in H$. Let us set

$$Y := H \times L_2(J;V). \tag{8}$$

Define the bounded linear operator $B : X \to Y'$ by

$$\langle Bw, v\rangle := (w(0), v_0) + \int_J (d_t w + Aw, v_1)dt, \quad (w,v) \in X \times Y, \tag{9}$$

and the bounded linear functional $F \in Y'$ by

$$Fv := (g, v_0) + \int_J (f, v_1)dt, \quad v \in Y. \tag{10}$$

Here and in the following we omit the dependence of the integrands on t. The space-time variational formulation then reads

$$\text{Find} \quad u \in X: \quad \langle Bu, v\rangle = Fv \quad \forall v \in Y. \tag{11}$$

Applicability of adaptive wavelet methods to this variational formulation was shown in [18]. Stable non-adaptive space-time sparse Petrov–Galerkin discretizations were derived and implemented in [2, 1]. As an interesting special case, collocation Runge–Kutta time-stepping schemes applied to the linear parabolic evolution equation (4) can

be interpreted as stable Petrov–Galerkin discretizations of the space-time variational formulation (11), see [4] and references therein. Issues of implementation are discussed in detail in [3].

Our second variational formulation is an instance of the Brézis–Ekeland–Nayroles variational principle [8, 16]. To state it, let

$$\widehat{A} := \tfrac{1}{2}(A + A') \quad \text{and} \quad \widetilde{A} := \tfrac{1}{2}(A - A') \tag{12}$$

denote the symmetric and the anti-symmetric part of A, respectively. Set

$$\widetilde{C}w := d_t w + \widetilde{A}w, \quad w \in X. \tag{13}$$

Define the bounded linear operator $\widehat{B} : X \to X'$ by

$$\langle \widehat{B}w, v\rangle := \int_J \Big\{ (\widehat{A}^{-1}\widetilde{C}w, \widetilde{C}v) + (\widehat{A}w, v) \Big\}\, dt + (w(T), v(T)), \tag{14}$$

for $w, v \in X$, and the bounded linear functional $\widehat{F} \in X'$ by

$$\widehat{F}v := \int_J (f, \widehat{A}^{-1}\widetilde{C}v + v)dt + (g, v(0)), \quad v \in X. \tag{15}$$

The second space-time variational formulation then reads

$$\text{Find} \quad \widehat{u} \in X : \quad \langle \widehat{B}\widehat{u}, w\rangle = \widehat{F}w \quad \forall w \in X. \tag{16}$$

The essential property of this space-time variational formulation to (11) is that the trial space and the test space coincide, and the operator $\widehat{B}$ is symmetric, that is $\langle \widehat{B}w, v\rangle = \langle \widehat{B}v, w\rangle$ for all $w, v \in X$. Moreover, owing to X-ellipticity of $\widehat{B}$,

$$\exists \widehat{\gamma} > 0 : \quad \langle \widehat{B}w, w\rangle \geq \widehat{\gamma}^2 \|w\|_X^2 \quad \forall w \in X, \tag{17}$$

see [1, Proposition 3.2.26], the Lax–Milgram lemma shows that there is a unique solution $\widehat{u}$ to (16), and it depends continuously on the data. It is instructive and straightforward to verify that u from (4) satisfies (16).

For completeness, we mention another space-time variational formulation that was investigated for example in [6, 7, 12, 9, 15]. It is based on testing $d_t u + Au = f$ by $v \in X$ with $v(T) = 0$ and performing integration by parts in time, and putting the adjoint of A onto the test function. This exposes $u(0)$ which is replaced by the known initial datum g. The result is the space-time variational formulation: Find $u \in L_2(J; V)$ such that

$$\int_J (u, -d_t v + A'v)dt = (g, v(0)) + \int_J (f, v)dt \quad \forall v \in X, \quad v(T) = 0. \tag{18}$$

This space-time variational formulation can be treated analogously to the first one. Finally, time-periodic problems with periodicity condition $u(0) = u(T)$ require only minor modifications, and will not be discussed here.

2. Space-time tensor product discretization.

2.1. Space-time tensor product subspaces. The discrete space-time trial and test spaces are built from nested finite-dimensional univariate temporal subspaces $E_L \subset H^1(J)$, $F_L \subset L_2(J)$, and spatial subspaces $V_L \subset V$ parameterized by an integer $L \geq 0$. We then define the discrete trial and test spaces (recall X and Y from (5)–(8))

$$X_L := E_L \otimes V_L \subset X \quad \text{and} \quad Y_L := V_L \times (F_L \otimes V_L) \subset Y, \tag{19}$$

where $\otimes$ denotes the algebraic tensor product. If $\bigcup_{L\geq 0} E_L$ is dense in $H^1(J)$ and $\bigcup_{L\geq 0} V_L$ is dense in V then $\bigcup_{L\geq 0} X_L$ is dense in X.

We assume henceforth that all subspaces are nontrivial and satisfy the discrete inf-sup condition

$$\forall L \geq 0: \quad \gamma_L := \inf_{w\in X_L\setminus\{0\}} \sup_{v\in Y_L\setminus\{0\}} \frac{\langle Bw, v\rangle}{\|w\|_X \|v\|_Y} > 0, \tag{20}$$

where B is as in (9). Note that at this stage, $\gamma_L > 0$ may non be bounded away from zero as $L \to \infty$.

By ellipticity (17) of the symmetric operator $\widehat{B}$, its discrete inf-sup constant $\widehat{\gamma}_L$ automatically satisfies $\widehat{\gamma}_L \geq \widehat{\gamma} > 0$ when X_L is used simultaneously as the trial and as the test space.

In the quantification of the well-posedness of the discrete problems the quantity

$$\Gamma_L := \sup_{w\in X_L\setminus\{0\}} \sup_{v\in Y_L\setminus\{0\}} \frac{\langle Bw, v\rangle}{\|w\|_X \|v\|_Y}, \quad L \geq 0, \tag{21}$$

will also play a role. It is clear that $\Gamma_L \leq \|B\|$. Similarly, the corresponding sup sup quantity $\widehat{\Gamma}_L$ for $\widehat{B}$ satisfies $\widehat{\Gamma}_L \leq \|\widehat{B}\|$.

The discrete inf-sup condition (20) necessitates $\dim X_L \leq \dim Y_L$. We discuss the cases $\dim X_L = \dim Y_L$ and $\dim X_L \leq \dim Y_L$ in Sections 2.2 and 2.3, respectively.

2.2. Discrete variational formulation. In this section we assume that $X_L \subset X$ and $Y_L \subset Y$ have the same dimension. The existence of such subspaces satisfying the discrete inf-sup condition (20) is guaranteed by bounded invertibility of $B : X \to Y'$. Practical instances are more subtle due to the non-symmetric contribution of the temporal derivative; however, in [4] it was shown that collocation Runge–Kutta time stepping schemes applied to temporal semi-discretizations admit an interpretation as a discrete space-time variational formulation, see (22) below, with discrete trial and test spaces of the form (19), and, in addition, discrete stability $\gamma_L \geq \gamma > 0$ may be achieved for all $L \geq 0$ in (20).

Given such X_L and Y_L, it is straightforward to define the discrete solution by the discrete space-time variational formulation

$$\text{Find} \quad u_L \in X_L: \quad \langle Bu_L, v\rangle = Fv \quad \forall v \in Y_L. \tag{22}$$

With the assumption that the discrete inf-sup condition (20) holds for X_L and Y_L, a unique solution $u_L \in X_L$ to (22) exists by the standard finite element method theory.

For any nontrivial subspace $X_L \subset X$, symmetry and ellipticity (17) of $\widehat{B}$ guarantee its ellipticity on X_L. The Lax–Milgram theorem provides a unique discrete solution to:

$$\text{Find} \quad \widehat{u}_L \in X_L: \quad \langle \widehat{B}\widehat{u}_L, w\rangle = \widehat{F}w \quad \forall w \in X_L. \tag{23}$$

Again, we emphasize that X_L is used as the trial and as the test space simultaneously. We will content ourselves with the full tensor product subspace X_L as in (19), but we mention that space-time compressive discretizations X_L and Y_L could be used for (22) and for (23), see [2].

Let Φ be a basis for X_L, and Ψ a basis for Y_L. Here, and where appropriate, we omit the subscript L for readability. Then we write $u_L = \Phi^{\mathsf{T}}\mathbf{u}$ with a vector of coefficients $\mathbf{u} \in \mathbb{R}^{\Phi}$ (indexed by the elements of Φ), and define the system matrix $\mathbf{B} \in \mathbb{R}^{\Psi\times\Phi}$ with the components $\mathbf{B}_{\psi\phi} = \langle B\phi, \psi\rangle$, as well as the load vector $\mathbf{F} \in \mathbb{R}^{\Psi}$ with the components $\mathbf{F}_{\psi} = F\psi$, where $(\phi, \psi) \in \Phi \times \Psi$. Then the discrete variational formulation (22) is equivalent to the algebraic equation

$$\mathbf{B}\mathbf{u} = \mathbf{F}. \tag{24}$$

The discrete inf-sup condition (20) implies that $\mathbf{B}$ is injective, so that (24), and also (27) below, is uniquely solvable.

Defining $\widehat{\mathbf{B}} \in \mathbb{R}^{\Phi\times\Phi}$ and $\widehat{\mathbf{F}} \in \mathbb{R}^{\Phi}$ analogously, and writing $\widehat{u}_L := \Phi^{\mathsf{T}}\widehat{\mathbf{u}}$, the symmetric discrete variational formulation (23) is equivalent to the algebraic equation

$$\widehat{\mathbf{B}}\widehat{\mathbf{u}} = \widehat{\mathbf{F}}. \tag{25}$$

2.3. Minimal residual variational formulation. For details and proofs for this section we refer to [2, 1]. Contrary to the aforegoing section we now assume that $\dim Y_L \geq \dim X_L$, both finite but possibly unequal. The motivation to consider this case is the fact that the discrete inf-sup condition (20) and indeed, discrete stability $\gamma_L \geq \gamma > 0$, is then easier to achieve, for one is allowed to choose any discrete test space Y_L that is "large enough". In this case the discrete variational formulation (22) is meaningless. Instead, we define the discrete solution as the minimizer of the functional residual by

$$u_L := \operatorname*{argmin}_{w_L\in X_L} R(w_L), \quad R(w_L) := \sup_{v\in Y_L\setminus\{0\}} \frac{|\langle F - Bw_L, v\rangle|}{\|v\|_Y}. \tag{26}$$

It is obvious that in the particular case $\dim X_L = \dim Y_L$ this formulation reduces to (22), so it is appropriate to use the same symbol u_L here. Under the inf-sup condition (20) there exists a unique solution $u_L \in X_L$ to (26), the solution u_L depends linearly and continuously on the load functional F, and the operator norm of the discrete solution mapping $F \mapsto u_L$ is bounded by $1/\gamma_L$. Defining $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{u}$ as in the previous section, the functional residual minimization (26) is equivalent to the algebraic equation

$$\mathbf{B}^{\top}\mathbf{N}^{-1}\mathbf{B}\mathbf{u} = \mathbf{B}^{\top}\mathbf{N}^{-1}\mathbf{F}, \tag{27}$$

where $\mathbf{N}$ is the symmetric positive definite matrix such that $\mathbf{v}^{\top}\mathbf{N}\mathbf{v} = \|\Psi^{\mathsf{T}}\mathbf{v}\|_Y^2$ for all coefficient vectors $\mathbf{v} \in \mathbb{R}^{\Psi}$. Note that (27) has the form of generalized Gauss normal equations. If $\dim X_L = \dim Y_L$ then it is equivalent to (24), and therefore henceforth we consider (24) as a special case of (27), and do not discuss it separately.

Instead of the exact matrix $\mathbf{N}$ in (27) one may use an approximation of it, say $\widetilde{\mathbf{N}}$ with

$$c_{\mathbf{N}}^2\mathbf{N} \leq \widetilde{\mathbf{N}} \leq C_{\mathbf{N}}^2\mathbf{N}, \tag{28}$$

for some constants $0 \le c_{\mathbf{N}} \le C_{\mathbf{N}} < \infty$. If $\widetilde{\mathbf{u}}$ denotes the corresponding solution, and $\widetilde{u}_L := \Phi^\top \widetilde{\mathbf{u}}$ then the quasi-optimality estimate

$$\|u - \widetilde{u}_L\|_X \le C_L \inf_{w_L \in X_L} \|u - w_L\|_X, \quad C_L := \frac{C_{\mathbf{N}}}{c_{\mathbf{N}}} \frac{\Gamma_L}{\gamma_L}, \tag{29}$$

can be shown.

Our objective now is to devise a preconditioner for the symmetric algebraic equations (25) and (27). This is the subject of the next section.

2.4. Parabolic space-time preconditioners. As in the aforegoing Sections (2.2)–(2.3), assume $X_L \subset X$ and $Y_L \subset Y$ are nontrivial subspaces that satisfy the discrete inf-sup condition (20), and $\dim X_L \le \dim Y_L$ are possibly unequal. The algebraic equations (25) and (27) stem from a space-time Petrov–Galerkin discretization of a parabolic evolution equation and therefore direct solution is likely impossible due to the large size of the system. Two basic methods for their approximate iterative solution are therefore

- the preconditioned Richardson iteration, and
- the preconditioned conjugate gradients method.

It may be preferable to apply the Richardson iteration to the symmetric formulation (25), due to the fact that even the forward application on $\widehat{\mathbf{B}}$ requires solutions of elliptic problems. Generally, this inverse may only be available approximately, and the accuracy of this approximation should be increased as the Richardson iteration progresses.

Both iterative schemes necessitate a good preconditioner, that is a matrix $\mathbf{M}$ such that the condition numbers of $\mathbf{M}^{-1}\widehat{\mathbf{B}}$ and $\mathbf{M}^{-1}(\mathbf{B}^\mathsf{T}\mathbf{N}^{-1}\mathbf{B})$ are small. To exhibit such a preconditioner we exploit the mapping properties of the operators B and $\widehat{B}$. Given a basis $\Phi \subset X_L$ we define $\mathbf{M} \in \mathbb{R}^{\Phi\times\Phi}$ as the matrix such that $\mathbf{w}^\mathsf{T}\mathbf{M}\mathbf{w} = \|\Phi^\mathsf{T}\mathbf{w}\|_X^2$ for all coefficient vectors $\mathbf{w}$. Its components are then

$$\mathbf{M}_{\varphi\phi} = \int_J \{(d_t\phi(t), d_t\varphi(t))_{V'} + (\phi(t), \varphi(t))_V\}dt + (\phi(T), \varphi(T)). \tag{30}$$

It can be shown that the spectrum σ of the preconditioned matrices satisfies

$$\sigma(\mathbf{M}^{-1}\widehat{\mathbf{B}}) \subset [\widehat{\gamma}_L, \widehat{\Gamma}_L] \quad \text{and} \quad \sigma(\mathbf{M}^{-1}(\mathbf{B}^\mathsf{T}\mathbf{N}^{-1}\mathbf{B})) \subset [\gamma_L^2, \Gamma_L^2], \tag{31}$$

and the intervals are the smallest possible. See Section 2.2 for the definition of these quantities. From this, one obtains

$$\kappa_2(\mathbf{M}^{-1}\widehat{\mathbf{B}}) = \widehat{\Gamma}_L/\widehat{\gamma}_L \quad \text{and} \quad \kappa_2(\mathbf{M}^{-1}(\mathbf{B}^\mathsf{T}\mathbf{N}^{-1}\mathbf{B})) = \Gamma_L^2/\gamma_L^2 \tag{32}$$

for the condition numbers with respect to the operator 2-norm. Owing to the norm equivalence $|||\cdot|||_X \sim \|\cdot\|_X$, see Section 1.2, omitting the last term $(\phi(T), \phi'(T))$ in (30) yields an equivalent preconditioner, see Section 2.5 for more details.

In Section 3 below we discuss the Kronecker structure of the system matrices $\mathbf{B}$ and $\widehat{\mathbf{B}}$, the norm matrix $\mathbf{N}$, and the preconditioner $\mathbf{M}$. We then describe the structure of the inverses $\mathbf{N}^{-1}$ and $\mathbf{M}^{-1}$, and their approximations based on a wavelet-in-time multigrid-in-space transform.

2.5. Norm equivalence on the trial space. As noted above, the two norms $\|\cdot\|_X$ and $|||\cdot|||_X$, related by $\|w\|_X^2 = |||w|||_X^2 + \|w(T)\|_H^2$, are equivalent. Since the norm $|||\cdot|||_X$ gives rise to a simpler preconditioner than (30), it is of interest to quantify

this norm equivalence. Obviously, $|||\cdot|||_X \le \|\cdot\|_X$. For the reverse comparison we need to find the constant C_X in the estimate $\|w(T)\|_H \le C_X |||w|||_X$, from which $\|w\|_X^2 \le (1+C_X^2)\,|||w|||_X^2$ follows. We shall do this only in the (relevant) situation that H is infinite-dimensional and the embedding $V \hookrightarrow H$ is compact. In this case there exists a countable orthonormal basis $\{\sigma_k\}_{k\in\mathbb{N}}$ for H, which is also an orthogonal basis for V and V'. Set $\lambda_k := \|\sigma_k\|_V$. The indexing is assumed to be such that $0 < \lambda_1 \le \lambda_k$ for all $k \in \mathbb{N}$. Then $1/\lambda_1$ is the norm of the embedding $V \hookrightarrow H$, appearing in the Friedrichs inequality when V and H are Sobolev spaces. Further, by orthonormality of the basis in H there follows $\lambda_k^{-1} = \|\sigma_k\|_{V'}$. Expanding $w \in X$ into this basis we can write $w = \sum_{k\in\mathbb{N}} w_k \otimes \sigma_k$, which implies $\|w(0)\|_H^2 = \sum_{k\in\mathbb{N}} |w_k(0)|^2$, $\|w\|_{L_2(J;V)}^2 = \sum_{k\in\mathbb{N}} \|\lambda_k w_k\|_{L_2(J)}^2$ and $\|\partial_t w\|_{L_2(J;V')}^2 = \sum_{k\in\mathbb{N}} \|\lambda_k^{-1} w_k'\|_{L_2(J)}^2$.

This motivates the following reduced question: Given $\lambda > 0$, find the smallest $C > 0$ such that for all $f \in H^1(J)$ there holds $C^{-2}|f(0)|^2 \le \|\lambda f\|_{L_2(J)}^2 + \|\lambda^{-1} f'\|_{L_2(J)}^2$. Herein, we can without loss of generality suppose that $f(0) = 1$. Then the question becomes that of minimizing the right hand side. Applying the variational principle this leads to the boundary value problem $-\lambda^{-2} f'' + \lambda^2 f = 0$ with the boundary conditions $f'(0) = 0$ and $f(T) = 1$. This can be solved explicitly, yielding the catenary as the solution. A direct computation then shows $C = 1/\sqrt{\tanh(\lambda^2 T)}$. This is a decreasing function of λ. Therefore, the best constant in $\|w(T)\| \le C_X |||w|||_X$ is given by $C_X^2 := 1/\tanh(\lambda_1^2 T)$. As $T \to \infty$, we have $C_X \searrow 1$. For instance, if $\lambda_1^2 T \ge 1$ then $C_X \le 1.1459$. On the other hand, if T is small, we can approximate $\tanh(\lambda_1^2 T) \sim \lambda_1^2 T$, so that $C_X \sim 1/\sqrt{\lambda_1^2 T}$ diverges as $T \searrow 0$.

In summary, if T is large enough, then the last term in (30) can be omitted without sacrificing the quality of the preconditioner $\mathbf{M}$.

3. Wavelet-in-time multigrid-in-space approximation. In order to simplify the further exposition of the main ideas relating to preconditioning, we shall assume from now on that in the parabolic evolution equation (4), the generator $t \mapsto A(t)$ is time-independent. Hence we write A instead of $A(t)$. While the results are stable with respect to slight perturbations of these assumptions, the discretization of the operators (9) and (14) becomes technical, and the quality of the proposed preconditioner may deteriorate for large variations of $t \mapsto A(t)$.

The assumption that A is time-independent allows us to equip the space V with the norm $\|\chi\|_V := (\widehat{A}\chi, \chi)^{1/2}$, which is equivalent to the original norm due to (1)–(3).

3.1. Kronecker product structure of discretized operators. Recall from (19) that the discrete trial and test spaces $X_L \subset X$ and $Y_L \subset Y$ are of the tensor product form $X_L = E_L \otimes V_L$ and $Y_L = V_L \times (F_L \otimes V_L)$. This has implications on the structure of the system matrices $\mathbf{B}$ and $\widehat{\mathbf{B}}$ introduced in Section 2.2.

Let now $\Theta \subset E_L$, $\Xi \subset F_L$, and $\Sigma \subset V_L$, be bases for the respective spaces. Set $\Phi := \{\theta \otimes \sigma : (\theta, \sigma) \in \Theta \times \Sigma\}$ and $\Psi_1 := \{\xi \otimes \sigma : (\xi, \sigma) \in \Xi \times \Sigma\}$. Then

$$\Phi \subset X_L \quad \text{and} \quad \Psi := (\Sigma \times \{0\}) \cup (\{0\} \times \Psi_1) \subset Y_L \tag{33}$$

are bases for X_L and Y_L, respectively.

The system matrix $\mathbf{B} \in \mathbb{R}^{\Psi\times\Phi}$ from Section 2.2, whose components are $\mathbf{B}_{\psi\phi} = \langle B\phi, \psi\rangle$, then has the form

$$\mathbf{B} = \begin{pmatrix} \mathbf{C}_t^{FE} \otimes \mathbf{M}_x + \mathbf{M}_t^{FE} \otimes \mathbf{A}_x \\ \mathbf{e}_0 \otimes \mathbf{M}_x \end{pmatrix}, \tag{34}$$

where **a)** "temporal FEM" matrices $\mathbf{C}_t^{FE}, \mathbf{M}_t^{FE} \in \mathbb{R}^{\Xi\times\Theta}$ and the row vector $\mathbf{e}_0 \in \mathbb{R}^{1\times\Theta}$ have the components

$$[\mathbf{C}_t^{FE}]_{\xi\theta} = (\theta', \xi)_{L_2(J)}, \quad [\mathbf{M}_t^{FE}]_{\xi\theta} = (\theta, \xi)_{L_2(J)}, \quad [\mathbf{e}_s]_{1\theta} = \theta(s), \tag{35}$$

and **b)** the usual "spatial FEM" mass and stiffness matrices $\mathbf{M}_x, \mathbf{A}_x \in \mathbb{R}^{\Sigma\times\Sigma}$ have the components

$$[\mathbf{M}_x]_{\tilde{\sigma}\sigma} = (\tilde{\sigma}, \sigma), \quad [\mathbf{A}_x]_{\tilde{\sigma}\sigma} = (A\tilde{\sigma}, \sigma). \tag{36}$$

Let $\mathbf{M}_t^E$ and $\mathbf{M}_t^F$ denote the temporal mass matrices for $\Theta \subset E_L$ and $\Xi \subset F_L$, respectively, and $\mathbf{A}_t^E$ the temporal stiffness matrix for $\Theta \subset E_L$. Let $\widehat{\mathbf{A}}_x$ be the symmetric part of $\mathbf{A}_x$. Then the preconditioner $\mathbf{M}$ from Section 2.4 can be expressed as

$$\mathbf{M} = \overline{\mathbf{M}} + (\mathbf{e}_T^\mathsf{T} \otimes \mathbf{e}_T) \otimes \mathbf{M}_x, \tag{37}$$

where

$$\overline{\mathbf{M}} := \mathbf{M}_t^E \otimes \widehat{\mathbf{A}}_x + \mathbf{A}_t^E \otimes (\mathbf{M}_x \widehat{\mathbf{A}}_x^{-1} \mathbf{M}_x). \tag{38}$$

Note that $\mathbf{e}_T^\mathsf{T} \otimes \mathbf{e}_T$ is a rank-one matrix in $\mathbb{R}^{\Theta\times\Theta}$ of the same size as $\mathbf{M}_t^E$ and $\mathbf{A}_t^E$. The norm-measuring matrix $\mathbf{N}$ from Section 2.3 is the block matrix

$$\mathbf{N} = \begin{pmatrix} \mathbf{M}_t^F \otimes \widehat{\mathbf{A}}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_x \end{pmatrix}. \tag{39}$$

Finally, the matrix $\widehat{\mathbf{B}}$ from Section 2.2 has the form

$$\widehat{\mathbf{B}} = \mathbf{Z} + \mathbf{M}_t^E \otimes \widehat{\mathbf{A}}_x + (\mathbf{e}_T^\mathsf{T} \otimes \mathbf{e}_T) \otimes \mathbf{M}_x, \tag{40}$$

where $\mathbf{Z} \in \mathbb{R}^{\Phi\times\Phi}$ is a symmetric matrix with the components $\mathbf{Z}_{\varphi\phi} := \langle \widehat{A}^{-1}\widetilde{C}\varphi, \widetilde{C}\phi\rangle$. Details on the computation of the matrix $\mathbf{Z}$ are discussed for a concrete example in Section 4.3.2.

3.2. Multigrid-in-space. As explained in Section 2.4, we require an efficient way to apply the inverses of $\mathbf{N}$ and of the preconditioner $\mathbf{M}$.

3.2.1. Inverse of N. As to the inverse of $\mathbf{N}$, from (39) we have the representation

$$\mathbf{N}^{-1} = \begin{pmatrix} (\mathbf{M}_t^F)^{-1} \otimes \widehat{\mathbf{A}}_x^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_x^{-1} \end{pmatrix}. \tag{41}$$

Using an L_2-orthogonal basis for the temporal test space $F_L \subset L_2(J)$, see Section 4.4.1, renders $\mathbf{M}_t^F$ a diagonal matrix, so that $\mathbf{N}^{-1}$ is block-diagonal. The inverse $\widehat{\mathbf{A}}_x^{-1}$ can be replaced by a multigrid method.

3.2.2. Inverse of M. Based on the discussion in Section 2.5, instead of $\mathbf{M}$ we consider the simpler matrix $\overline{\mathbf{M}}$ as the preconditioner, see (37), for which the spectral estimate $\overline{\mathbf{M}} \le \mathbf{M} \le (1 + C_X^2)\overline{\mathbf{M}}$ holds. Recall that $C_X \le 2$ for sufficiently large T.

For the sake of readability, we now omit the superscript E in the notation for the temporal FEM matrices $\mathbf{M}_t^E$ and $\mathbf{A}_t^E$. Let $\mathbf{V}_t$ be a transformation matrix of the same

size as $\mathbf{A}_t$ and $\mathbf{M}_t$ such that $\mathbf{V}_t^\mathsf{T}\mathbf{M}_t\mathbf{V}_t$ and $\mathbf{V}_t^\mathsf{T}\mathbf{A}_t\mathbf{V}_t$ are spectrally equivalent to some diagonal matrices $\mathbf{J}_t$ and $\mathbf{D}_t$, respectively, where the constants of the equivalence should be close to one. We refer to the ratio of those constants as the condition number of the transformation. The canonical choice of $\mathbf{V}_t$ is the matrix collecting in its rows the $\mathbf{M}_t$-orthonormal eigenvectors of the generalized eigenvalue problem $\mathbf{A}_t\mathbf{v} = \lambda\mathbf{M}_t\mathbf{v}$, in which case $\mathbf{J}_t := \mathbf{V}_t^\mathsf{T}\mathbf{M}_t\mathbf{V}_t$ is the identity and $\mathbf{D}_t := \mathbf{V}_t^\mathsf{T}\mathbf{A}_t\mathbf{V}_t$ has the corresponding eigenvalues on the diagonal. In this case, the condition number of the transformation is one, but the drawback is that $\mathbf{V}_t$ is a dense matrix (see the discussion in the Introduction). A sparse alternative $\mathbf{V}_t := \mathbf{T}_t^\mathsf{T}$ is given by the inverse wavelet(-like) transformation introduced in Section 3.3 below. In that case we take $\mathbf{J}_t$ as the diagonal of $\mathbf{V}_t^\mathsf{T}\mathbf{M}_t\mathbf{V}_t$ and $\mathbf{D}_t$ as the diagonal of $\mathbf{V}_t^\mathsf{T}\mathbf{A}_t\mathbf{V}_t$. The condition numbers are investigated in Section 3.3.2.

Hence, we suppose that $\mathbf{V}_t$ is such that

$$\mathbf{V}_t^\mathsf{T}\mathbf{M}_t\mathbf{V}_t \sim \mathbf{J}_t \quad \text{and} \quad \mathbf{V}_t^\mathsf{T}\mathbf{A}_t\mathbf{V}_t \sim \mathbf{D}_t \tag{42}$$

with some nonnegative diagonal matrices $\mathbf{J}_t$ and $\mathbf{D}_t$, where $\mathbf{J}_t$ is positive definite (specifically, we use (53)). Let $\mathbf{I}_x$ denote the identity matrix of the same size as $\mathbf{M}_x$ and $\mathbf{A}_x$. We set $\mathbf{T} := \mathbf{V}_t^\mathsf{T} \otimes \mathbf{I}_x$. Then $\mathbf{T}^\mathsf{T} = \mathbf{V}_t \otimes \mathbf{I}_x$.

To obtain a computationally accessible approximation of $\mathbf{M}^{-1}$, we begin with the observation that

$$\mathbf{M}^{-1} \sim \overline{\mathbf{M}}^{-1} = \mathbf{T}^\mathsf{T}(\mathbf{T}\overline{\mathbf{M}}\mathbf{T}^\mathsf{T})^{-1}\mathbf{T} \sim \mathbf{T}^\mathsf{T}(\mathbf{J}_t \otimes \widehat{\mathbf{A}}_x + \mathbf{D}_t \otimes (\mathbf{M}_x\widehat{\mathbf{A}}_x^{-1}\mathbf{M}_x))^{-1}\mathbf{T}. \tag{43}$$

This requires the approximation of the inverse of the block-diagonal matrix

$$\mathbf{J}_t \otimes \widehat{\mathbf{A}}_x + \mathbf{D}_t \otimes (\mathbf{M}_x\widehat{\mathbf{A}}_x^{-1}\mathbf{M}_x), \tag{44}$$

where each block has the form $j^2\widehat{\mathbf{A}}_x + d^2(\mathbf{M}_x\widehat{\mathbf{A}}_x^{-1}\mathbf{M}_x)$ with some nonnegative reals j and d (recall that $\widehat{\mathbf{A}}_x$ denotes the symmetric part of the spatial FEM stiffness matrix $\mathbf{A}_x$). Thus, the inversion of the space-time matrix (44) is equivalent to a sequence of *independent* spatial problems of the form

$$j^2\mathbf{H}_x^\gamma\mathbf{p} = \mathbf{q} \quad \text{with} \quad \mathbf{H}_x^\gamma := \widehat{\mathbf{A}}_x + \gamma^2(\mathbf{M}_x\widehat{\mathbf{A}}_x^{-1}\mathbf{M}_x), \tag{45}$$

where $\gamma = d/j$. Now, the identity

$$\mathbf{H}_x^\gamma = (\widehat{\mathbf{A}}_x + i\gamma\mathbf{M}_x)\widehat{\mathbf{A}}_x^{-1}(\widehat{\mathbf{A}}_x - i\gamma\mathbf{M}_x), \tag{46}$$

where i is the imaginary unit, is easily verified. Thus, solving (45) is equivalent to solving two Helmholtz problems with the imaginary parameter $\pm i\gamma$, with one forward application of $\widehat{\mathbf{A}}_x$ between them. In order to avoid complex numbers arithmetics, one can observe that the left-hand side of (46) is the Schur complement of a 2×2 block matrix that admits iterative solvers that are robust in γ [19]. However, recalling that $\mathbf{M}$ is merely a preconditioner, we note that the left-hand side of (46) is estimated by the composition of two Helmholtz problems

$$\mathbf{H}_x^\gamma \le \overline{\mathbf{H}}_x^\gamma := (\widehat{\mathbf{A}}_x + \gamma\mathbf{M}_x)\widehat{\mathbf{A}}_x^{-1}(\widehat{\mathbf{A}}_x + \gamma\mathbf{M}_x) \le 2\mathbf{H}_x^\gamma \tag{47}$$

with, this time, real parameter γ. Indeed, one only has to verify $2\gamma\mathbf{M}_x \le \mathbf{H}_x^\gamma$, but this follows by changing to the eigenvalue basis of the generalized eigenvalue problem $\widehat{\mathbf{A}}_x\mathbf{p} = \mu\mathbf{M}_x\mathbf{p}$ together with the estimate $2\gamma \le \mu + \gamma^2/\mu$. Therefore, in the inverse of

(44), which is again block-diagonal, we replace each spatial block by the correspondent real double Helmholtz problem, that is, by the application of the mapping

$$\mathbf{q} \mapsto j^{-2}(\widehat{\mathbf{A}}_x + \gamma \mathbf{M}_x)^{-1} \widehat{\mathbf{A}}_x (\widehat{\mathbf{A}}_x + \gamma \mathbf{M}_x)^{-1} \mathbf{q}. \tag{48}$$

The exact solution of these positive definite Helmholtz problems can be replaced by the multigrid method, which is robust in $\gamma > 0$ [17]. In view of parallelization we reiterate here that (44) is a block-diagonal matrix.

3.3. Construction of the inverse wavelet transform. In this section the inverse wavelet transform $\mathbf{T}_t$ is constructed such that the sparse matrix $\mathbf{V}_t := \mathbf{T}_t^{\mathsf{T}}$ has the properties required in Section 3.2.2.

3.3.1. Construction. Let $\overline{\mathcal{T}}$ denote the collection of finite temporal meshes of the form $\mathcal{T} = \{0 =: t_0 < t_1 < \ldots < t_N := T\}$. Recall that $J = (0, T)$. For any $\mathcal{T} \in \overline{\mathcal{T}}$ let $\Theta(\mathcal{T})$ denote the set of hat functions (hats) $\mathcal{T}$, that is the set of piecewise linear splines θ_t on $\mathcal{T}$ that evaluate to one on some $t \in \mathcal{T}$ and to zero on all other nodes $\mathcal{T} \setminus \{t\}$. We let $s(\mathcal{T})$ denote the left-most node $s \in \mathcal{T} \setminus \{0, T\}$ for which the energy $\epsilon(\mathcal{T}) := \|\theta_s\|^2_{H^1(J)} / \|\theta_s\|^2_{L_2(J)}$ is maximized among the non-boundary hats $\theta_t \in \Theta(\mathcal{T}) \setminus \{\theta_0, \theta_T\}$.

We now describe the construction of a wavelet-like basis for the span of $\Theta(\mathcal{T})$.

Assume given:

1. $\mathcal{T} \in \overline{\mathcal{T}}$ with $\#\mathcal{T} \geq 3$.
2. Relative energy bandwidth parameter $\eta > 1$ (we will use $\eta := 1.9$). It is employed for the identification of fine and coarse nodes in Step 1 below.
3. Number of orthogonalization steps $\nu \in \mathbb{N}$. It is employed in Step 2.

Step 0. Set $\Theta_0(\mathcal{T}) := \Theta(\mathcal{T})$. Initialize $K := 0$.

Step 1. Given $\mathcal{T} \in \overline{\mathcal{T}}$, identify fine $\mathcal{T}_+ \subset \mathcal{T}$ and coarse nodes $\mathcal{T}_- \subset \mathcal{T}$, $\mathcal{T}_- \in \overline{\mathcal{T}}$, by:

1. Set $\mathcal{T}_0 := \mathcal{T}$ and $n := 0$.
2. While $\#\mathcal{T}_n \geq 3$ and $\epsilon(\mathcal{T}_n) \geq \eta^{-1} \epsilon(\mathcal{T}_0)$ do:
   Set $\mathcal{T}_{n+1} := \mathcal{T}_n \setminus s(\mathcal{T}_n)$ and increase n by 1.
3. Define $\mathcal{T}_- := \mathcal{T}_n$ and $\mathcal{T}_+ := \mathcal{T} \setminus \mathcal{T}_-$.

Please note that, after the n-th iteration, $\Theta(\mathcal{T}_{n+1})$ does not equal $\Theta(\mathcal{T}_n) \setminus \theta_{s(\mathcal{T}_n)}$ because removing a node from $\mathcal{T}_n$ modifies the hats associated with the two neighboring nodes of $s(\mathcal{T}_n) \in \mathcal{T}_n$ by widening their support. Yet, obviously $\Theta(\mathcal{T}_n) = \operatorname{span}(\theta_{s(\mathcal{T}_n)} \cup \Theta(\mathcal{T}_{n+1}))$. On a uniform grid $\mathcal{T}$, in particular, the algorithm returns every second node of $\mathcal{T} \setminus \{0, T\}$ collected in $\mathcal{T}_-$ and every other in $\mathcal{T}_+$. Also note that each $\Theta(\mathcal{T}_n)$ constitutes a partition of unity on the interval J, and so do the coarse hats $\Theta(\mathcal{T}_-)$.

Step 2. Let $\Theta_+(\mathcal{T})$ denote the set of hats that are removed from $\Theta(\mathcal{T}_n)$ over all iterations n during Step 1. In general, the obtained fine hats $\Theta_+(\mathcal{T})$ are not the same as $\{\theta_t \in \Theta(\mathcal{T}) : t \in \mathcal{T}_+\}$ because neighboring nodes of $\mathcal{T}$ may have been selected into $\mathcal{T}_+$. However, the basis $\Theta_+(\mathcal{T}) \cup \Theta(\mathcal{T}_-)$ is now a two-scale basis for the space spanned by $\Theta(\mathcal{T})$. To obtain a wavelet-like basis $\Psi_+(\mathcal{T})$ for the detail space $\operatorname{span}(\Theta_+(\mathcal{T}))$ we perform an approximate orthogonalization of the fine hats $\Theta_+(\mathcal{T})$ against the coarse hats $\Theta(\mathcal{T}_-)$ using the mapping

$$P_{\mathcal{T}_-} : L_2(J) \to L_2(J), \quad \theta \mapsto P_{\mathcal{T}_-}\theta := \theta - \sum_{\zeta \in \Theta(\mathcal{T}_-)} \frac{(\zeta, \theta)_{L_2(J)}}{(\zeta, 1)_{L_2(J)}} \zeta. \tag{49}$$

The wavelet-like basis $\Psi_+(\mathcal{T})$ is defined by

$$\Psi_+(\mathcal{T}) := \{[P_{\mathcal{T}_-}]^\nu \theta : \theta \in \Theta_+(\mathcal{T})\}, \tag{50}$$

where $\nu \in \mathbb{N}$ was assumed given. Note that an application of $P_{\mathcal{T}_-}$ to a compactly supported θ increases its support by at most the support of the few coarse hats that are nonzero there.

Step 3. Call $\Psi_K := \Psi_+(\mathcal{T})$ and $\Theta_{K+1} := \Theta(\mathcal{T}_-)$. If $\#\mathcal{T}_- = 2$ then terminate. If $\#\mathcal{T}_- \geq 3$, increase K by one, and continue with Step 1 for $\mathcal{T} := \mathcal{T}_-$.

Once the algorithm terminates, we have constructed a multiscale basis

$$\Psi(\mathcal{T}) := \Theta_{K+1} \cup \Psi_K \cup \ldots \cup \Psi_1 \cup \Psi_0 \tag{51}$$

for the span of the original hats $\Theta(\mathcal{T})$. Please note that since the number of levels is not known in advance, $K+1$ denotes the coarsest while 0 denotes the finest level.

Clearly, the mapping $P_{\mathcal{T}_-}$ in (49) operates only in the subspace spanned by $\Theta(\mathcal{T}_-)$ leaving its orthogonal complement in $L_2(J)$ unchanged. Since $\Theta(\mathcal{T}_-)$ constitute a partition of unity on J we have $P_{\mathcal{T}_-} 1 = 1 - \sum_{\zeta \in \Theta(\mathcal{T}_-)} \zeta = 0$. Therefore, $P_{\mathcal{T}_-}\theta$ has at least one vanishing moment, so it is indeed a wavelet-like oscillating function. In the limit $\nu \to \infty$, the function $[P_{\mathcal{T}}]^\nu \theta$ is orthogonal to the span of $\Theta(\mathcal{T}_-)$.

Some typical wavelet-like basis functions are shown in Figure 3. On the uniform mesh, there is a striking similarity to the ${}_{2,2\nu}\psi$ family of wavelets of [10]. But unlike those, our wavelet-like functions do not have 2ν vanishing moments, hence are not exactly the same.

In Table 1 we provide a Matlab code that constructs the coefficient transformation matrix $\mathbf{T}_t$ such that the mass matrix with respect to the multiscale basis is $\mathbf{T}_t \mathbf{M}_t \mathbf{T}_t^\mathsf{T}$ where $\mathbf{M}_t$ is the mass matrix with respect to the original hat basis $\Theta(\mathcal{T})$. The mapping $\mathbf{w} \mapsto \mathbf{T}_t^\mathsf{T}\mathbf{w}$ is the inverse wavelet transformation, as it takes a vector of coefficients $\mathbf{w}$ with respect to the multiscale wavelet-like basis $\Psi(\mathcal{T})$ and returns a vector of coefficients with respect to the hat basis $\Theta(\mathcal{T})$.

The transformation can be performed in a matrix-free fashion with linear complexity known from wavelet pyramid schemes. However, we are thinking of the situation where it is applied (from the right) to a distributed matrix that stores spatial vectors in its columns. This is because $(\mathbf{T}_t \otimes \mathbf{I}_x)\mathrm{Vec}(\mathbf{U}) = \mathrm{Vec}(\mathbf{U}\mathbf{T}_t^\mathsf{T})$ is precisely the type of operation required in (43), where $\mathbf{U}$ is a matrix of suitable size and $\mathrm{Vec}(\cdot)$ stacks the columns of its argument one after the other into a long vector. We therefore believe that it is beneficial, at least at first, to formulate and test the transformation in matrix form, in particular to be able to take advantage of distributed matrix storage and manipulation packages.

3.3.2. Condition numbers of the basis. We compute the condition number of the wavelet-like basis $\Psi(\mathcal{T}^k)$ constructed in Section 3.3.1 for a test sequence of temporal meshes $\mathcal{T}^k$, $k = 1, \ldots, 11$. Each $\mathcal{T}^k$ contains a uniform temporal mesh on $[0, T]$ with $2^k + 1$ nodes and additionally the nodes $1/2 \pm 2^{-r}$, $r = 1, \ldots, 20$, that provide geometric refinement towards $t = 1/2$. The condition number in $L_2(J)$ and $H^1(J)$ is the ratio of the Riesz basis constants of the normalized wavelet-like basis in $L_2(J)$ and in $H^1(J)$, respectively. These constants are precisely the extremal eigenvalues of $\{\mathbf{T}_t \mathbf{M}_t \mathbf{T}_t^\mathsf{T}\}$ and $\{\mathbf{T}_t \mathbf{A}_t \mathbf{T}_t^\mathsf{T}\}$, where $\{\mathbf{X}\}$ denotes the Jacobi-preconditioned matrix

$$(\mathrm{diag}\mathbf{X})^{-1/2}\mathbf{X}(\mathrm{diag}\mathbf{X})^{-1/2}, \tag{52}$$

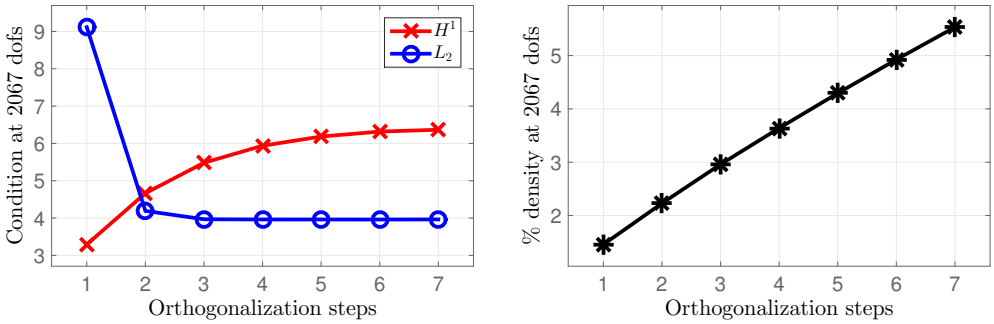

FIG. 1. *Condition number of the wavelet-like basis (left) and percentage of nonzeros in the transformation matrix (right) as a function of the number of orthogonalization steps, see Section 3.3.2. Without orthogonalization, the condition number in L_2 is very large.*

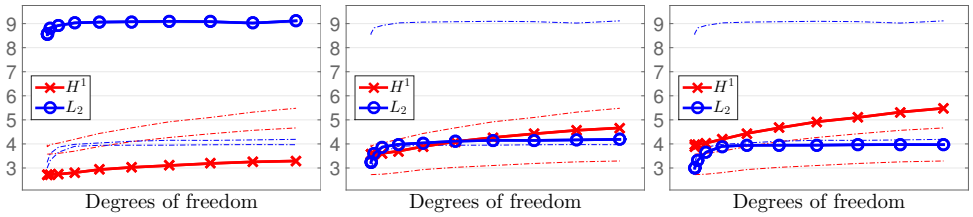

FIG. 2. *Condition number of the wavelet-like basis as a function of the size of the test mesh for $\nu = 1, 2, 3$ (left to right) orthogonalization steps.*

and $\mathbf{M}_t$ and $\mathbf{A}_t$ are the mass and the stiffness matrix with respect to the original hat basis $\Theta(\mathcal{T}^k)$, cf. Section 3.3. Setting

$$\mathbf{J}_t := \operatorname{diag}(\mathbf{V}_t^\mathsf{T}\mathbf{M}_t\mathbf{V}_t) \quad \text{and} \quad \mathbf{D}_t := \operatorname{diag}(\mathbf{V}_t^\mathsf{T}\mathbf{A}_t\mathbf{V}_t) \tag{53}$$

for $\mathbf{V}_t := \mathbf{T}_t^\mathsf{T}$, equivalence (42) is satisfied with those constants. Additionally, we measure the percentage of nonzeros in the transformation matrix $\mathbf{T}_t$.

The results depending on the number ν of orthogonalization steps are shown in Figure 1 for $\mathcal{T}^k$ with $k = 11$. It seems that $\nu = 2$ is the appropriate choice because for larger ν the condition in $L_2(J)$ remains stable while that in $H^1(J)$ increases slightly. Moreover, the percentage of nonzeros increases roughly linearly with ν, as long as ν is small.

The dependence of the condition on the number of nodes in $\mathcal{T}^k$, $k = 1, \dots, 11$, is shown in Figure 2 for $\nu = 1, 2, 3$, orthogonalization steps. The condition in $L_2(J)$ is robust as the number of nodes increases, and that in $H^1(J)$ exhibits a slight growth. Again, $\nu = 2$ seems to be a good choice.

4. Performance of the space-time preconditioner.

4.1. Model problem. In order to investigate the performance of the space-time preconditioners, we look at the model problem of convection-diffusion on the interval $D := (0, 1)$. We fix the end-time $T := 1$. Specifically, we assume for (1)–(3) that $H := L_2(D)$, $V := H_0^1(D)$, and

$$A := -\partial_x^2 + \beta\partial_x \tag{54}$$

TABLE 1
Matlab code for the construction of the inverse wavelet transform, see Section 3.3.

```
% Given: Real vector TE = [ a < b < ... ], real nu > 1
% Computes: Inverse wavelet transformation matrix Tt

N = length(TE); h = diff(TE); g = 1./h;

% Temporal mass and stiffness matrix on the mesh TE
Mt = spdiags([h 0; 0 h]' * [1 2 0; 0 2 1]/6, -1:1, N, N);
At = spdiags([g 0; 0 g]' * [-1 1 0; 0 1 -1], -1:1, N, N);

% Rows of Tt are coefficients of wavelets wrt finest hats
Tt = speye(N);

% Auxiliary variables corresponding to current coarse level
mc = Mt; ac = At; IC = 1:N;
while (IC(end) >= 3)
  % Find most energetic hats
  eta = 1.9; % Relative energy bandwidth
  IF = []; e0 = 0;
  while (length(IC) >= 3)
    e1 = diag(ac) ./ diag(mc); e1([1 end]) = 0;
    [e1, j] = max(e1);
    if (e1 <= e0 / eta); break; end;
    if (e0 == 0); e0 = e1; end

    % Neighbors of the fine hat j get coarsened:
    J = j + (-1:1);
    dt = speye(3); dt([1;3], 2) = -ac(j+[-1;1], j) / ac(j, j);
    ac(J, :) = dt * ac(J, :); ac(:, J) = ac(:, J) * dt';
    mc(J, :) = dt * mc(J, :); mc(:, J) = mc(:, J) * dt';
    ac(j, :) = []; ac(:, j) = []; mc(j, :) = []; mc(:, j) = [];

    Tt(IC(J), :) = dt * Tt(IC(J), :);

    IF = [IF, IC(j)]; IC(j) = [];
  end

  % Approximate nu-fold orthogonalization
  P = @(X) X - (X * Mt * Tt(IC,:)') * diag(1./sum(mc)) * Tt(IC,:);
  for k = 1:nu; Tt(IF,:) = P(Tt(IF,:)); end

  % Reorder rows (first coarse then fine basis functions)
  Tt(1:IC(end), :) = Tt([IC, IF], :); IC = 1:length(IC);
end
```

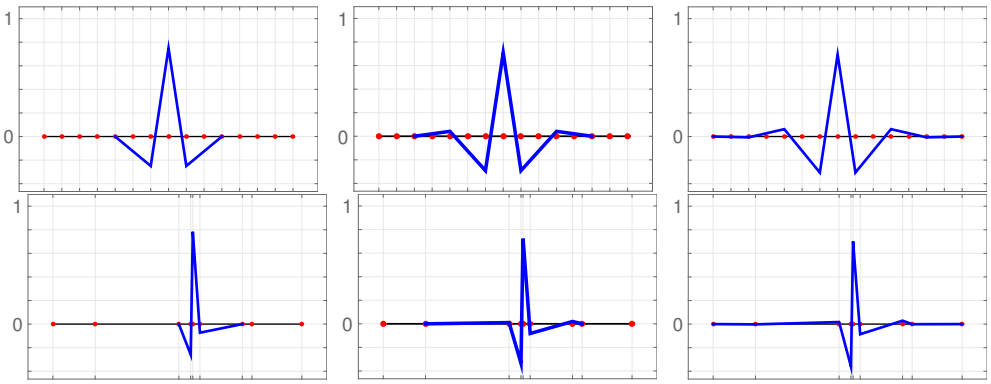

FIG. 3. *Example of a wavelet-like basis function for $\nu = 1, 2, 3$ (left to right). Top: uniform mesh, bottom: highly non uniform mesh. The mesh is indicated by the vertical lines. Note the increasing support with ν.*

with some constant $\beta \in \mathbb{R}$. The symmetric part of A is then $\widehat{A} = -\partial_x^2$, and the asymmetric part is $\widetilde{A} = \beta\partial_x$. Of particular interest is the performance of the preconditioner with respect to drift velocity β. Further numerical experiments on space-time discretization and preconditioning based on the space-time variational formulation (11) with symmetric generator A and in several spatial dimensions can be found in [1, 2, 3, 4, 5].

We recall from the introduction of Section 3 that the norm on $V = H_0^1(D)$ is taken as $\|\chi\|_V := \|\partial_x\chi\|_{L_2(D)}$. Then inequality (3) holds with $\alpha = 1$ and $\gamma_0 = 0$. Indeed, if $\chi \in H_0^1(D)$ then $(\widetilde{A}\chi, \chi) = 0$, so that $(A\chi, \chi) = (\widehat{A}\chi, \chi) = \|\chi\|_V^2$.

4.2. Computation of the condition numbers. The condition number κ_2 of the preconditioned systems $\mathbf{M}^{-1}\widehat{\mathbf{B}}$ and $\mathbf{M}^{-1}\mathbf{B}^{\mathsf{T}}\mathbf{N}^{-1}\mathbf{B}$ with different approximate variants of the inverses is computed by means of a power iteration for the maximal and minimal eigenvalue respectively. For instance, in order to approximate the maximal eigenvalue of $\mathbf{M}^{-1}\widehat{\mathbf{B}}$ we iterate $\tilde{\mathbf{x}}_n := \mathbf{M}^{-1}\widehat{\mathbf{B}}\mathbf{x}_{n-1}$, $e_n^{\max} := |\tilde{\mathbf{x}}_n|$, $\mathbf{x}_n = \tilde{\mathbf{x}}_n/e_n^{\max}$, starting with an all-ones vector. Then $e_n^{\max}$ converges to the maximum eigenvalue, and we stop the iteration when the relative improvement is less than 10^{-4}. In order to compute the minimal eigenvalue, more work is required. Here, we iterate $\tilde{\mathbf{x}}_n := \mathbf{M}\widehat{\mathbf{B}}^{-1}\mathbf{x}_{n-1}$, $e_n^{\min} := 1/|\tilde{\mathbf{x}}_n|$, $\mathbf{x}_n := \tilde{\mathbf{x}}_n e_n^{\min}$, starting with an all-ones vector, upon which $e_n^{\min}$ converges to the minimum eigenvalue. The application of the inverse of $\widehat{\mathbf{B}}$ is replaced by a preconditioned conjugate gradient iteration with preconditioner $\mathbf{M}^{-1}$ and tolerance 10^{-6}. We point out that in the case of multigrid approximation of the blocks of $\mathbf{M}^{-1}$, say $\mathbf{W} \approx \mathbf{M}^{-1}$, one needs the inverse $\mathbf{W}^{-1} \approx \mathbf{M}$ in the inverse power iteration. To that end we compute the matrix representation of the multigrid method and compute its inverse directly. We use a symmetric version of the multigrid where possible. The overall computational effort to determine the condition numbers limits our numerical experiments to one-dimensional spatial domains. All computations are done in MATLAB R2014b.

4.3. Symmetric formulation.

4.3.1. Discretization. For the sake of simplicity, for the symmetric space-time variational formulation (16) we assume in the numerical experiments that the spatial subspace $V_L \subset V$ are given as the L-dimensional space of polynomials of degree at most $(L+1)$ satisfying the boundary conditions of $H_0^1(D)$. A suitable basis $\Sigma \subset V_L$ is given by the integrated Legendre polynomials $IP_k : t \mapsto \int_0^t P_k(s)ds$, $k = 1, \ldots, L$, where P_k is the shifted Legendre polynomial on the interval $(0, 1)$ of degree k normalized in $L_2(D)$. The temporal discretization is achieved by taking $E_L \subset H^1(J)$ as the space of continuous piecewise affine functions on J with respect to a uniform partition of J. The number of the intervals will be specified below. This then defines the tensor product trial and test space X_L by (19).

4.3.2. Computation of the space-time matrix. Recall from Section 3.1 that the space-time system matrix $\widehat{\mathbf{B}}$ contains the matrix $\mathbf{Z}$. It has the components $\mathbf{Z}_{\varphi\phi} = \langle\widehat{A}^{-1}\widetilde{C}\varphi, \widetilde{C}\phi\rangle$, where $\varphi, \phi \in \Phi$ are the space-time tensor product basis functions for X_L from (33), and $\widetilde{C} = d_t + \widetilde{A}$. The computation of this matrix is delicate due to the inverse of $\widehat{A}$. Our choice of V_L is motivated by the fact that the action of the operator $\widetilde{A}$ and of the inverse $\widehat{A}^{-1}$ on functions in V_L can be computed exactly; thus we can compute the spatial matrices $[\mathbf{A}_x^{11}]_{\tilde\sigma\sigma} := (\widehat{A}^{-1}\widetilde{A}\tilde\sigma, \widetilde{A}\sigma)$, $[\mathbf{A}_x^{01}]_{\tilde\sigma\sigma} := (\widehat{A}^{-1}\tilde\sigma, \widetilde{A}\sigma)$, $[\mathbf{A}_x^{00}]_{\tilde\sigma\sigma} := (\widehat{A}^{-1}\tilde\sigma, \sigma)$, and $\mathbf{A}_x^{10} := (\mathbf{A}_x^{01})^{\mathsf{T}}$, where as before, the basis $\tilde\sigma, \sigma \in \Sigma$ is used to index the components. In addition to the temporal FEM matrices from Section 3.1

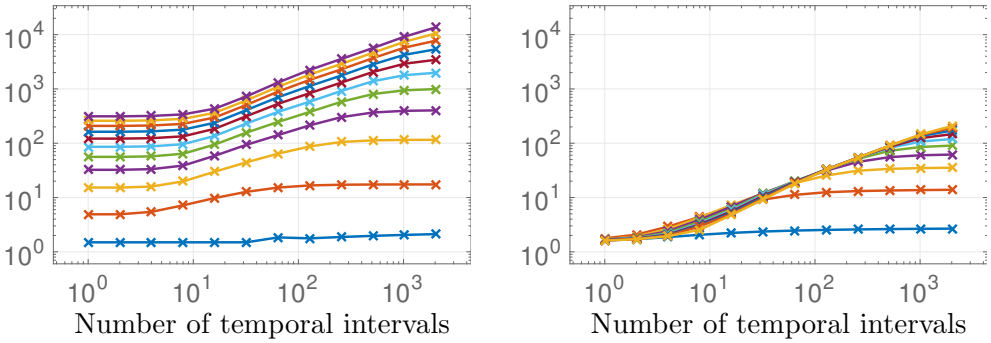

FIG. 4. *Condition number of the preconditioned system matrix for the symmetric space-time variational formulation as a function of the number of temporal intervals, see Section 4.3.3. Each line from bottom to top corresponds to a value of drift velocity $\beta = 0, 10, \ldots, 100$. Left: Wavelet-in-time and exact computation in space. Right: Same preconditioner with an additional β-dependent term.*

we need the temporal advection matrix $[\mathbf{C}_t^{01}]_{\tilde{\theta},\theta} := (\tilde{\theta}, \theta')_{L_2(J)}$, where $\tilde{\theta}, \theta$ are elements of the temporal basis Θ, as well as its transpose $\mathbf{C}_t^{10} := (\mathbf{C}_t^{01})^\mathsf{T}$. The matrix $\mathbf{Z}$ can now be written as

$$\mathbf{Z} = (\mathbf{A}_t^E \otimes \mathbf{A}_x^{00}) + (\mathbf{C}_t^{01} \otimes \mathbf{A}_x^{10}) + (\mathbf{C}_t^{10} \otimes \mathbf{A}_x^{01}) + (\mathbf{M}_t^E \otimes \mathbf{A}_x^{11}). \tag{55}$$

4.3.3. Condition number. We compute the condition number κ_2 of the preconditioned system $\mathbf{M}^{-1}\widehat{\mathbf{B}}$, where the inverse $\mathbf{M}^{-1}$ is approximated as described in Section 3.2.2. For the required temporal transformation $\mathbf{V}_t$ we use the inverse wavelet transform from Section 3.3 with $\nu = 2$ orthogonalization steps. The subspace $V_L \subset V$ is spanned by $L = 20$ integrated Legendre polynomials (the results do not essentially depend on L). All spatial inverses, such as in (48) and (55) are computed exactly. For the values $\beta = 0, 10, 20, \ldots, 100$, of the drift velocity in (54), the condition number of the preconditioned system matrix as a function of $N = 2^0, 2^1, 2^2, \ldots, 2^{11}$, equidistant temporal intervals is shown in Figure 4 (left). For each value of β, the condition number increases with N up to a certain value, which seems to scale like β^2. For $\beta \lesssim 10$ the condition number remains very small as expected, but deteriorates for larger β especially with increasing number of temporal intervals. In Figure 4 (right) we show the condition number where, motivated by the structure of $\widetilde{\mathbf{B}}$, we have added the β-dependent term $(\mathbf{M}_t^E \otimes \mathbf{A}_x^{11})$ to each block of the preconditioner (44); each block is inverted exactly. This results in significantly lower condition numbers, which are however, still not robust in β.

4.3.4. Application to optimal control. Optimal control problems contrained by evolution equations like (4) provide a major motivation for space-time simultaneous discretizations due to the coupling of the forward-in-time state equation and the backward-in-time adjoint equation. Consider for example the minimization of the standard tracking functional

$$\mathcal{J}(y,u) := \frac{1}{2}\int_J \|y(t) - y_\star(t)\|^2_{L_2(D)} dt + \frac{\lambda}{2}\int_J \|u(t)\|^2_{L_2(D)} dt, \tag{56}$$

where u is the control variable, $y_\star$ is the desired state, y is the state variable constrained by the parabolic evolution equation

$$u(0) = 0, \quad d_t y(t) + Ay(t) = f(t) + u(t) \quad \text{(a.e.) } t \in J, \tag{57}$$

and $\lambda > 0$ is a regularization parameter. Here A is given by (54) and we assume zero drift $\beta = 0$. Introducing the dual variable p for the constraint, forming the first order optimality conditions, eliminating the control variable and discretizing the resulting linear system leads to the saddle-point problem

$$\mathbf{A}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} = \mathbf{rhs} \quad \text{where} \quad \mathbf{A} := \begin{pmatrix}\mathbf{M}_0 & \widehat{\mathbf{B}}\\ \widehat{\mathbf{B}} & -\lambda^{-1}\mathbf{M}_0\end{pmatrix}, \tag{58}$$

for some vector **rhs** which is of no importance here. Here, $\mathbf{M}_0 := \mathbf{M}_t^E \otimes \mathbf{M}_x$ is the matrix corresponding to the space-time L_2 inner product. The large block matrix $\mathbf{A}$ is symmetric but indefinite, and therefore the system is typically solved using the preconditioned MinRes method. It is of particular interest to obtain computationally accessible preconditioners for (58) that are robust simultaneously in the discretization parameters and in the regularization parameter $\lambda > 0$. Having used the symmetric space-time variational formulation, we are in the situation that each block of the matrix is *symmetric*, so that the block-diagonal matrix with blocks $\mathbf{M}_0 + \sqrt{\lambda}\widehat{\mathbf{B}}$ and $\lambda^{-1}(\mathbf{M}_0 + \sqrt{\lambda}\widehat{\mathbf{B}})$) is a good preconditioner for $\mathbf{A}$, see [19, Section 4.1]. Replacing $\widehat{\mathbf{B}}$ by the spectrally equivalent $\mathbf{M}$ we obtain the preconditioner

$$\mathbf{P} := \begin{pmatrix}\mathbf{M}_0 + \sqrt{\lambda}\mathbf{M} & 0\\ 0 & \lambda^{-1}(\mathbf{M}_0 + \sqrt{\lambda}\mathbf{M})\end{pmatrix}. \tag{59}$$

Now, the preconditioned system $\mathbf{P}^{-1}\mathbf{A}$ is well-conditioned. As in Section 3.2.2 we transform this preconditioner (with the inverse wavelet transform as in Section 4.3.3) to a block-diagonal one such that the application of $\mathbf{P}^{-1}$ involves a series of independent spatial systems of the form

$$(\mathbf{M}_x + \sqrt{\lambda}\mathbf{H}_x^\gamma)\mathbf{p} = \mathbf{q} \tag{60}$$

with $\mathbf{H}_x^\gamma$ from (45). These systems are small in our case and we solve them directly. The resulting condition numbers κ_2 for $\mathbf{P}^{-1}\mathbf{A}$ with this approximation of $\mathbf{P}^{-1}$ are shown in Figure 5 as a function of the number of temporal intervals $N = 2^0, 2^1, \ldots, 2^{11}$, for the values $\lambda = 10^{-5}, 10^{-4}, \ldots, 10^5$ of the regularization parameter in (56). The condition number stays bounded by *three* across the whole range of N and λ.

4.4. Nonsymmetric formulation.

4.4.1. Discretization. The nonsymmetric space-time variational form (11) is discretized using space-time tensor product subspaces (19) as described in Section 2.3. As $E_L \subset H^1(J)$ we take the space of continuous piecewise affine functions on J with respect to a uniform partition. The number of the intervals will be specified below. The test space $F_L \subset L_2(J)$ is taken as the space of piecewise constants with respect to a single uniform refinement of the same partition. The dimension of F_L is then roughly twice that of E_L. With the basis for F_L consisting of piecewise constant functions nonzero on exactly one of the subintervals, the matrix $\mathbf{N}$ defined in (39) and its inverse (41) are block-diagonal. The spatial discretization $V_L \subset H_0^1(D)$ is also given by continuous piecewise affine functions with respect to a uniform partition. The number of elements will be specified below. For symmetric A, the resulting full tensor product discretization (19) satisfies the inf-sup condition (20) and the boundedness condition (21) with constants γ_L and Γ_L that depend only on the constants in the

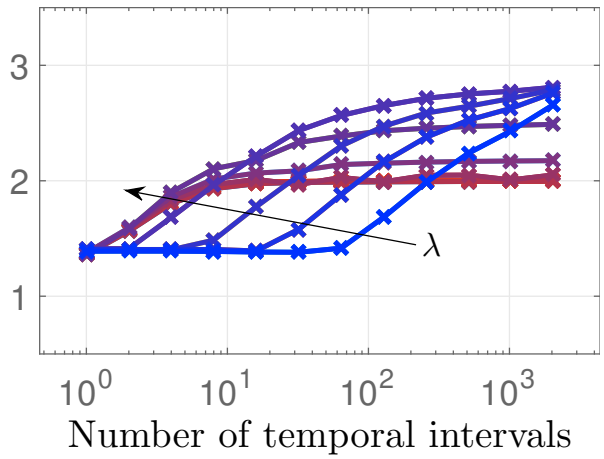

FIG. 5. *Condition number of the preconditioned optimality system as a function of the number of temporal intervals. Each line corresponds to a value* $\lambda = 10^{-5}, 10^{-4}, \dots, 10^{5}$ *of the regularization parameter, as indicated by the arrow. See Section 4.3.4.*

(1)–(3). In particular they can be chosen independently of the temporal and spatial resolution. This is *not* the case if F_L equals the space E_L or $d_t E_L$, see [4] for the proof of those facts. For nonsymmetric A this property is again lost, and other trial and test spaces, and also other continuous norms are required to recover unconditional stability [1, Theorem 5.2.6].

4.4.2. Condition number. We compute the condition number κ_2 for the preconditioned system $\mathbf{M}^{-1}\mathbf{B}^{\mathsf{T}}\mathbf{N}^{-1}\mathbf{B}$, where the inverse $\mathbf{M}^{-1}$ is approximated as described in Section 3.2.2. The temporal transformation $\mathbf{V}_t$ is given by the inverse wavelet transform from Section 3.3 with $\nu = 2$ orthogonalization steps. Since the main difficulty is in the application of the inverse of $\mathbf{M}$, we apply the block-diagonal $\mathbf{N}^{-1}$ by direct inversion in each case. The spatial discretization $V_L \subset H_0^1(D)$ is fixed as the space of continuous piecewise affine functions with respect to the uniform partition of the interval D into 2^9 equal elements.

First, no multigrid approximation is employed, and all spatial matrices are inverted directly. The number of temporal intervals is varied as $N = 2^0, 2^1, 2^2, \dots, 2^{11}$. The dependence of the condition number on the number of temporal intervals N for different the drift velocities $\beta = 0, 0.01, \dots, 0.1$, is documented in Figure 6 (left). The condition number increases dramatically with N even for these small values of the drift velocity. Note that the drift velocity is here two orders of magnitude smaller than in Section 4.3.3.

Now we set $\beta = 0$ for the drift velocity and switch on the multigrid-in-space approximation in the application of the inverse of $\mathbf{M}$. We use the multigrid as described in [13, Section 4.1]. The pre-smoother is defined by the Gauss–Seidel iteration $\mathbf{q} \mapsto (\mathbf{L}+\mathbf{D})^{-1}(-\mathbf{U}\mathbf{q}+\mathbf{f})$ and the post-smoother by $\mathbf{q} \mapsto (\mathbf{U}+\mathbf{D})^{-1}(-\mathbf{L}\mathbf{q}+\mathbf{f})$ for the equation $(\mathbf{L}+\mathbf{D}+\mathbf{U})\mathbf{q} = \mathbf{f}$ with strictly lower/upper triangular $\mathbf{L}/\mathbf{U}$ and diagonal $\mathbf{D}$. This ensures symmetry of the multrigrid procedure when the number of pre- and post-smoothing steps is the same. As the prolongation we use the canonical embedding of V_k into V_{k+1}, the restriction is its transpose. In order to approximate the action of (48), the multigrid procedure with one pre-smoothing step is applied with the matrix $\widehat{\mathbf{A}}_x + \gamma\mathbf{M}_x$, then $\widehat{\mathbf{A}}_x$ is applied, then again the multigrid procedure.

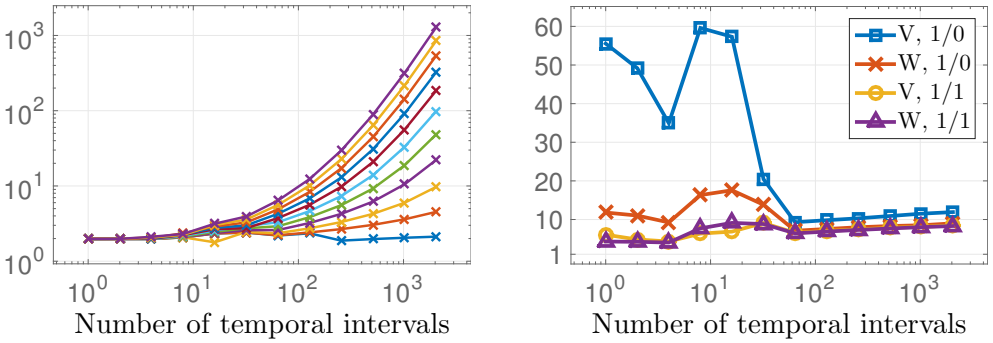

FIG. 6. *Condition number of the preconditioned system matrix for the nonsymmetric space-time variational formulation as a function of the number of temporal intervals, see Section 4.4.2. Left: Wavelet-in-time and exact computation in space. Each line from bottom to top corresponds to a value of drift velocity* $\beta = 0, 0.01, \ldots, 0.1$. *Right: For* $\beta = 0$, *wavelet-in-time multigrid-in-space preconditioner with different multigrid parameters: V- or W-cycle and the number of pre-/post-smoothing steps.*

For four configurations of the multigrid determined by whether the V- or the W-cycle is executed, and whether none or one post-smoothing step is performed, the resulting condition numbers for the space-time preconditioned system are visualized in Figure 6 (right). Based on those measurements we recommend to use the V-cycle with one pre- and one post-smoothing step, since the resulting condition number remains below *ten* and the resulting preconditioner is symmetric.

5. Conclusions. For algebraic linear systems arising from different space-time Petrov–Galerkin discretizations of linear parabolic evolution equations we have developed a wavelet-in-time multigrid-in-space preconditioner. The sparsity of the wavelet-in-time transformation is crucial to reduce the inter-process communication cost when the parallelization is done along the temporal direction. The transformation block-diagonalizes the canonical preconditioner given by the continuous space-time norms, and allows to invert the resulting spatial blocks in parallel using standard spatial multigrid methods. We have presented several numerical experiments documenting the excellent performance of the preconditioner in the regime of small Péclet numbers, together with a first application to robust preconditioning of optimality systems from optimal control constrained by parabolic PDEs. Problems such as implementation of the matrix $\mathbf{Z}$ in (40) in a general setting, and parallel computation of the preconditioner are currently under investigation.

REFERENCES

[1] Roman Andreev, *Stability of space-time Petrov-Galerkin discretizations for parabolic evolution equations*, PhD thesis, ETH Zürich, 2012. ETH Diss. No. 20842.

[2] ———, *Stability of sparse space-time finite element discretizations of linear parabolic evolution equations*, IMA J. Numer. Anal., 33 (2013), pp. 242–260.

[3] ———, *Space-time discretization of the heat equation*, Numer. Algorithms, 67 (2014), pp. 713–731.

[4] Roman Andreev and Julia Schweitzer, *Conditional space-time stability of collocation Runge–Kutta for parabolic evolution equations*, Electron. Trans. Numer. Anal., 41 (2014), pp. 62–80.

[5] Roman Andreev and Christine Tobler, *Multilevel preconditioning and low rank tensor iteration for space-time simultaneous discretizations of parabolic PDEs*, Numer. Lin. Algebra Appl., online (2014). In press.

[6] Ivo Babuška and Tadeusz Janik, *The h-p version of the finite element method for parabolic equations. I. The p version in time.*, Numerical Methods for Partial Differential Equations, 5 (1989), pp. 363–399.

[7] ———, *The h-p version of the finite element method for parabolic equations. II. The h-p version in time.*, Numerical Methods for Partial Differential Equations, 6 (1990), pp. 343–369.

[8] Haïm Brézis and Ivar Ekeland, *Un principe variationnel associé à certaines équations paraboliques. Le cas dépendant du temps*, C. R. Acad. Sci. Paris Sér. A-B, 282 (1976), pp. Ai, A1197–A1198.

[9] Nabi Chegini and Rob Stevenson, *Adaptive wavelet schemes for parabolic problems: Sparse matrices and numerical results*, SIAM J. Numer. Anal., 49 (2011), pp. 182–212.

[10] Albert Cohen, Ingrid Daubechies, and Jean-Christophe Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.

[11] Lawrence C. Evans, *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, 1998.

[12] Michael Griebel and Daniel Oeltz, *A sparse grid space-time discretization scheme for parabolic problems*, Computing, 81 (2007), pp. 1–34.

[13] Wolfgang Hackbusch, *Multigrid methods and applications*, vol. 4, Springer-Verlag, Berlin, 1985. Second printing 2003.

[14] Jacques-Louis Lions and Enrico Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York, 1972.

[15] Christian Mollet, *Stability of Petrov-Galerkin discretizations: Application to the space-time weak formulation for parabolic evolution problems*, tech. report, Universität Paderborn, 2013.

[16] Bernard Nayroles, *Deux théorèmes de minimum pour certains systèmes dissipatifs*, C. R. Acad. Sci. Paris Sér. A-B, 282 (1976), pp. Aiv, A1035–A1038.

[17] Maxim A. Olshanskii and Arnold Reusken, *On the convergence of a multigrid method for linear reaction-diffusion problems*, Computing, 65 (2000), pp. 193–202.

[18] Christoph Schwab and Rob Stevenson, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp., 78 (2009), pp. 1293–1318.

[19] Walter Zulehner, *Nonstandard norms and robust estimates for saddle point problems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 536–560.

R. Andreev

RICAM, Altenberger Str. 69, 4040 Linz, Austria

roman.andreev@oeaw.ac.at

December 2, 2014