SIAM J. SCI. COMPUT.
Vol. 38, No. 4, pp. A2173–A2208

ANALYSIS OF A NEW SPACE-TIME PARALLEL MULTIGRID ALGORITHM FOR PARABOLIC PROBLEMS*

MARTIN J. GANDER† AND MARTIN NEUMÜLLER‡

Abstract. We present and analyze a new space-time parallel multigrid method for parabolic equations. The method is based on arbitrarily high order discontinuous Galerkin discretizations in time and a finite element discretization in space. The key ingredient of the new algorithm is a block Jacobi smoother. We present a detailed convergence analysis when the algorithm is applied to the heat equation and determine asymptotically optimal smoothing parameters, a precise criterion for semi-coarsening in time or full coarsening, and give an asymptotic two grid contraction factor estimate. We then explain how to implement the new multigrid algorithm in parallel and show with numerical experiments its excellent strong and weak scalability properties.

Key words. space-time parallel methods, multigrid in space-time, DG-discretizations, strong and weak scalability, parabolic problems

AMS subject classifications. 65N55, 65F10, 65L60

DOI. 10.1137/15M1046605

1. Introduction.

About ten years ago, clock speeds of processors stopped increasing, and the only way to obtain more performance is by using more processing cores. This has led to new generations of supercomputers with millions of computing cores, and even today's small devices are multicore. In order to exploit these new architectures for high performance computing, algorithms must be developed that can use these large numbers of cores efficiently. When solving evolution partial differential equations, the time direction offers itself as a further direction for parallelization, in addition to the spatial directions. The parareal algorithm [31, 34, 1, 39, 17, 9] has sparked renewed interest in the area of time parallelization, a field that is now just over fifty years old; see the historical overview [19]. We are interested here in space-time parallel methods, which can be based on the two fundamental paradigms of domain decomposition or multigrid. Domain decomposition methods in space-time lead to waveform relaxation type methods; see [16, 18, 20] for classical Schwarz waveform relaxation, [10, 11, 12, 13, 2] for optimal and optimized variants, and [30, 35, 15] for Dirichlet–Neumann and Neumann–Neumann waveform relaxation. The spatial decompositions can be combined with the parareal to obtain algorithms that run on arbitrary decompositions of the space-time domain into space-time subdomains; see [33, 14]. Space-time multigrid methods were developed in [21, 32, 43, 29, 42, 28, 27, 45] and reached good F-cycle convergence behavior when appropriate semi-coarsening and extension operators were used. In particular we want to mention [28], where the Fourier mode analysis is used to analyze the space-time two-grid cycle which uses pointwise smoothers in space and time. For a variant for nonlinear problems, see [6, 38, 37].

We present and analyze here a new space-time parallel multigrid algorithm that has excellent strong and weak scalability properties on large scale parallel computers.

*Submitted to the journal's Methods and Algorithms for Scientific Computing section November 2, 2015; accepted for publication (in revised form) April 5, 2016; published electronically July 12, 2016.
http://www.siam.org/journals/sisc/38-4/M104660.html

†Section de Mathématiques, Geneva University, 2-4 rue du Lièvre, CP 64 CH-1211 Genève (martin.gander@unige.ch).

‡Institute of Computational Mathematics, Johannes Kepler University, Altenberger Straße 69, 4040 Linz, Austria (martin.neumueller@jku.at).

As a model problem we consider the heat equation in a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$ with boundary $\Gamma := \partial\Omega$ on the bounded time interval $[0, T]$,

$$\begin{aligned} \partial_t u(\boldsymbol{x}, t) - \Delta u(\boldsymbol{x}, t) &= f(\boldsymbol{x}, t) && \text{for } (\boldsymbol{x}, t) \in Q := \Omega \times (0, T), \\ u(\boldsymbol{x}, t) &= 0 && \text{for } (\boldsymbol{x}, t) \in \Sigma := \Gamma \times (0, T), \\ u(\boldsymbol{x}, 0) &= u_0(\boldsymbol{x}) && \text{for } (\boldsymbol{x}, t) \in \Sigma_0 := \Omega \times \{0\}. \end{aligned} \tag{1.1}$$

We divide the time interval $[0, T]$ into subintervals

$$0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T \quad \text{with } t_n = n\,\tau \text{ and } \tau = \frac{T}{N}$$

and use a standard finite element discretization in space and a discontinuous Galerkin approximation in time, which leads to the large linear system in space-time

$$[M_h \otimes K_\tau + K_h \otimes M_\tau]\,\boldsymbol{u}_{n+1} = \boldsymbol{f}_{n+1} + M_h \otimes C_\tau \boldsymbol{u}_n, \quad n = 0, 1, \ldots, N-1. \tag{1.2}$$

Here, M_h is the standard mass matrix and K_h is the standard stiffness matrix in space obtained by using the nodal basis functions $\{\varphi_i\}_{i=1}^{N_x} \subset H_0^1(\Omega)$, where N_x is the number of unknowns in space, i.e.,

$$M_h[i,j] := \int_\Omega \varphi_j(\boldsymbol{x})\varphi_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad K_h[i,j] := \int_\Omega \nabla\varphi_j(\boldsymbol{x}) \cdot \nabla\varphi_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad i,j = 1, \ldots, N_x.$$

The matrices for the time discretization, where a discontinuous Galerkin approximation with polynomials of order $p_t \in \mathbb{N}_0$ is used, are given by

$$\begin{aligned} K_\tau[k,\ell] &:= -\int_{t_{n-1}}^{t_n} \psi_\ell^n(t)\partial_t\psi_k^n(t)\mathrm{d}t + \psi_\ell^n(t_n)\psi_k^n(t_n), \quad k,\ell = 1, \ldots, N_t, \\ M_\tau[k,\ell] &:= \int_{t_{n-1}}^{t_n} \psi_\ell^n(t)\psi_k^n(t)\mathrm{d}t, \quad C_\tau[k,\ell] := \psi_\ell^{n-1}(t_{n-1})\psi_k^n(t_{n-1}). \end{aligned}$$

Here the basis functions for one time interval (t_{n-1}, t_n) are given by $\mathbb{P}^{p_t}(t_{n-1}, t_n) = \operatorname{span}\{\psi_\ell^n\}_{\ell=1}^{N_t}$, $N_t = p_t + 1$, and for $p_t = 0$, we would, for example, get a backward Euler scheme. Moreover, we denote each component of a vector $\boldsymbol{v}_n \in \mathbb{R}^{N_x N_t}$ by $\boldsymbol{v}_{n,j,\ell} \in \mathbb{R}$, where n is the index of the nth time step, j is the index with respect to the unkowns in space, and ℓ is the index corresponding to the unknowns with respect to time, i.e., $\boldsymbol{v}_{n,j} \in \mathbb{R}^{N_t}$. The right-hand side is given by

$$\boldsymbol{f}_{n+1,j,\ell} := \int_{t_n}^{t_{n+1}} \int_\Omega f(\boldsymbol{x}, t)\varphi_j(\boldsymbol{x})\psi_\ell^{n+1}(t)\mathrm{d}\boldsymbol{x}\mathrm{d}t, \quad j = 1, \ldots, N_x,\ \ell = 1, \ldots, N_t.$$

On the time interval (t_n, t_{n+1}), we can therefore define the approximation

$$u_{h,\tau}^{n+1}(\boldsymbol{x}, t) = \sum_{\ell=1}^{N_t} \sum_{j=1}^{N_x} \boldsymbol{u}_{n+1,j,\ell}\,\varphi_j(\boldsymbol{x})\,\psi_\ell^{n+1}(t),$$

where $\boldsymbol{u}_{n+1}$ is the solution of the linear system (1.2). We thus have to solve the block triangular system

$$\begin{pmatrix} A_{\tau,h} & & & \\ B_{\tau,h} & A_{\tau,h} & & \\ & \ddots & \ddots & \\ & & B_{\tau,h} & A_{\tau,h} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_1 - B_{\tau,h}\boldsymbol{u}_0 \\ \boldsymbol{f}_2 \\ \vdots \\ \boldsymbol{f}_N \end{pmatrix} \tag{1.3}$$

with $A_{\tau,h} := M_h \otimes K_\tau + K_h \otimes M_\tau$ and $B_{\tau,h} := -M_h \otimes C_\tau$.

To solve the linear system (1.3), one can simply apply a forward substitution with respect to the blocks corresponding to the time steps. Hence one has to invert the matrix $A_{\tau,h}$ for each time step, where, for example, a multigrid solver can be applied. This is the usual way that time dependent problems are solved when implicit schemes are used [40, 24, 25], but this process is entirely sequential. We want to apply a parallelizable space-time multigrid scheme to solve the global linear system (1.3) at once. First we study general properties of the discontinuous Galerkin time stepping scheme in section 2. In section 3 we present our method, and we study its properties in section 4 using local Fourier mode analysis. Numerical examples are given in section 5, and the parallel implementation is discussed in section 6, where we also show scalability studies. We give an outlook on further developments in section 7.

2. Time discretization. Here we study for $\lambda \geq 0$ the discontinuous Galerkin discretization in time for the scalar evolution equation

$$\partial_t u(t) + \lambda u(t) = f(t) \quad \text{for } t \in (0,T) \quad \text{and} \quad u(0) = u_0 \in \mathbb{R}, \tag{2.1}$$

which leads to the discrete variational problem:

Find $u_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$ such that for all $v_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$

$$\begin{aligned} &- \int_{t_{n-1}}^{t_n} u_\tau^n(t)\partial_t v_\tau^n(t)\mathrm{d}t + u_\tau^n(t_n)v_\tau^n(t_n) + \lambda \int_{t_{n-1}}^{t_n} u_\tau^n(t)v_\tau^n(t)\mathrm{d}t \\ &\qquad = \int_{t_{n-1}}^{t_n} f(t)v_\tau^n(t)\mathrm{d}t + u_\tau^{n-1}(t_{n-1})v_\tau^n(t_{n-1}). \end{aligned} \tag{2.2}$$

The problem (2.2) is equivalent to the system of linear equations

$$[K_\tau + \lambda M_\tau]\, \boldsymbol{u}_{n+1} = \boldsymbol{f}_{n+1} + C_\tau \boldsymbol{u}_n, \quad n = 0, 1, \ldots, N-1. \tag{2.3}$$

Here $\boldsymbol{u}_n \in \mathbb{R}^{N_t}$ and the right-hand side is given by

$$\boldsymbol{f}_{n,\ell} := \int_{t_{n-1}}^{t_n} f(t)\psi_\ell^n(t)\mathrm{d}t, \quad \ell = 1, \ldots, N_t \text{ and } n = 1, \ldots, N.$$

To study the properties of the discontinuous Galerkin discretization (2.2), we consider for a function $f : (t_{n-1}, t_n) \to \mathbb{R}$ the Radau quadrature rule of order $2s-1$,

$$\int_{t_{n-1}}^{t_n} f(t)\mathrm{d}t \approx \tau \sum_{k=1}^{s} b_k f(t_{n-1} + c_k \tau)$$

with the weights $b_k \in \mathbb{R}_+$ and the integration points $c_1 = 0$ and $c_2, \ldots, c_s \in [0,1]$; see [25].

THEOREM 2.1. *The discontinuous Galerkin approximation* (2.2) *of the model problem* (2.1) *is equivalent to the* (p_t+1)*-stage implicit Runge–Kutta scheme RADAU IA if the integral of the right-hand side is approximated by the Radau quadrature of order* $2p_t + 1$.

Proof. With this approximation, and using integration by parts, we obtain from (2.2) the variational problem:

Find $u_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$ such that for all $v_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$

$$\begin{aligned}\int_{t_{n-1}}^{t_n} \partial_t u_\tau^n(t) v_\tau^n(t) \mathrm{d}t + u_\tau^n(t_{n-1}) v_\tau^n(t_{n-1}) + \lambda \int_{t_{n-1}}^{t_n} u_\tau^n(t) v_\tau^n(t) \mathrm{d}t \\ = \tau \sum_{k=1}^{p_t+1} b_k f(t_{n-1} + c_k\tau) v_\tau^n(t_{n-1} + c_k\tau) + u_\tau^{n-1}(t_{n-1}) v_\tau^n(t_{n-1}).\end{aligned} \tag{2.4}$$

The idea of the proof is to apply the discontinuous collocation method introduced in [23] to the model problem (2.1). Let $c_1 = 0$ and $c_2, \ldots, c_{p_t+1} \in [0, 1]$ be the integration points of the Radau quadrature of order $2p_t+1$ with the weights $b_1, \ldots, b_{p_t+1} \in \mathbb{R}\backslash\{0\}$. Then the discontinuous collocation method is given by the following:

Find $w_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$ such that for all $k = 2, \ldots, p_t + 1$

$$\begin{aligned}w_\tau^n(t_{n-1}) - w_\tau^{n-1}(t_{n-1}) = \tau b_1 \left[f(t_{n-1}) - \partial_t w_\tau^n(t_{n-1}) - \lambda w_\tau^n(t_{n-1}) \right], \\ \partial_t w_\tau^n(t_{n-1} + c_k\tau) + \lambda w_\tau^n(t_{n-1} + c_k\tau) = f(t_{n-1} + c_k\tau).\end{aligned} \tag{2.5}$$

In [23] it was shown that the discontinuous collocation method (2.5) is equivalent to the $(p_t + 1)$-stage implicit Runge–Kutta scheme RADAU IA. Hence it remains to show the equivalence of the discontinuous Galerkin method (2.4) to the discontinuous collocation method (2.5). First, we observe that $\partial_t u_\tau^n v_\tau^n$ and $u_\tau^n v_\tau^n$ are polynomials of degree at most $2p_t$. Therefore we can replace the integrals on the left-hand side of (2.4) with the Radau quadrature of order $2p_t + 1$ and obtain

$$\begin{aligned}\tau \sum_{k=1}^{p_t+1} b_k \partial_t u_\tau^n(t_{n-1} + c_k\tau) v_\tau^n(t_{n-1} + c_k\tau) + u_\tau^n(t_{n-1}) v_\tau^n(t_{n-1}) \\ + \tau\lambda \sum_{k=1}^{p_t+1} b_k u_\tau^n(t_{n-1} + c_k\tau) v_\tau^n(t_{n-1} + c_k\tau) \\ = \tau \sum_{k=1}^{p_t+1} b_k f(t_{n-1} + c_k\tau) v_\tau^n(t_{n-1} + c_k\tau) + u_\tau^{n-1}(t_{n-1}) v_\tau^n(t_{n-1})\end{aligned} \tag{2.6}$$

with $v_\tau^n \in \mathbb{P}^{p_t}(t_{n-1}, t_n)$. As test functions v_τ^n we consider the Lagrange polynomials

$$\ell_i^n(t) = \prod_{\substack{j=1 \\ j\neq i}}^{p_t+1} \frac{t - (t_{n-1} + c_j\tau)}{\tau(c_i - c_j)} \qquad \text{for } i = 1, \ldots, p_t + 1.$$

Hence we have $\ell_i^n(t_{n-1}+c_j\tau) = 0$ for $i \neq j$ and $\ell_i^n(t_{n-1}+c_i\tau) = 1$ for $i = 1, \ldots, p_t+1$. First we use the test function $v_\tau^n = \ell_1^n$ in (2.6) and obtain

$$\tau b_1 \partial_t u_\tau^n(t_{n-1}) + u_\tau^n(t_{n-1}) + \tau\lambda b_1 u_\tau^n(t_{n-1}) = \tau b_1 f(t_{n-1}) + u_\tau^{n-1}(t_{n-1}).$$

This implies that the solution u_τ^n of (2.4) satisfies the first equation of (2.5). For the test function $v_\tau^n = \ell_k^n$, $k = 2, \ldots, p_t + 1$ we further get

$$\tau b_k \partial_t u_\tau^n(t_{n-1} + c_k\tau) + \tau\lambda b_k u_\tau^n(t_{n-1} + c_k\tau) = \tau b_k f(t_{n-1} + c_k\tau).$$

Dividing this equation by $\tau b_k \neq 0$ we see that the solution u_τ^n of the discontinuous Galerkin scheme (2.4) also satisfies the second equation of the discontinuous collocation method (2.5). Hence the solution u_τ^n of the discontinuous Galerkin scheme (2.4)

is a solution of the discontinuous collocation method (2.5). The converse is proved by reverting the arguments. ∎

The RADAU IA scheme has been introduced in the PhD thesis [5] in 1969; see also [4]. From the proof of Theorem 2.1 we see that the jump of the discrete solution at time t_{n-1} is equal to the pointwise error multiplied by the time step size τ and the weight b_1; see (2.5). Hence the height of the jump can be used as a simple error estimator for adaptive time stepping.

THEOREM 2.2. *For $s \in \mathbb{N}$, the s-stage RADAU IA scheme is of order $2s-1$ and the stability function $R(z)$ is given by the $(s-1,s)$ subdiagonal Padé approximation of the exponential function e^z. Furthermore the method is A-stable, i.e.,*

$$|R(z)| < 1 \qquad \text{for } z \in \mathbb{C} \text{ with } \Re(z) < 0.$$

Proof. The proof can be found in [25]. ∎

COROLLARY 2.3. *The stability function $R(z)$ of the discontinuous Galerkin approximation with polynomial degree $p_t \in \mathbb{N}$ is given by the (p_t, p_t+1) subdiagonal Padé approximation of the exponential function e^z. Furthermore the method is A-stable, i.e.,*

$$|R(z)| < 1 \qquad \text{for } z \in \mathbb{C} \text{ with } \Re(z) < 0.$$

Proof. For the Dahlquist test equation $\partial_t u = \lambda u, \lambda \in \mathbb{C}$ we obtain by Theorem 2.1 that the discontinuous Galerkin scheme is equivalent to the RADAU IA method. Hence the two methods have the same stability function $R(z)$. Applying Theorem 2.2 completes the proof. ∎

LEMMA 2.4. *For $\lambda \in \mathbb{C}$ the eigenvalues of the matrix $(K_\tau + \lambda M_\tau)^{-1} C_\tau \in \mathbb{C}^{N_t \times N_t}$ are given by*

$$\sigma((K_\tau + \lambda M_\tau)^{-1} C_\tau) = \{0, R(-\lambda\tau)\},$$

where $R(z)$ is the A-stability function of the given discontinuous Galerkin time stepping scheme; see Corollary 2.3.

Proof. First we notice that the eigenvalues of the matrix $(K_\tau + \lambda M_\tau)^{-1} C_\tau$ are independent of the basis $\{\psi_k\}_{k=1}^{N_t}$, which is used to compute the matrices K_τ, M_τ, and C_τ. Hence we can use basis functions $\{\psi_k\}_{k=1}^{N_t}$, where the eigenvalues of the matrix $(K_\tau + \lambda M_\tau)^{-1} C_\tau$ are easy to compute, i.e., polynomials $\psi_k \in \mathbb{P}^{p_t}(0,\tau)$ which are zero or one at $t = \tau$

$$\psi_k(\tau) = \begin{cases} 1 & k = 1, \\ 0 & k \neq 1 \end{cases} \qquad \text{for } k = 1, \ldots, N_t.$$

To study the A-Stability of the discontinuous Galerkin discretization, we consider for $\lambda \in \mathbb{C}$ the model problem

$$\partial_t u(t) = -\lambda u(t), \quad t \in (0,\tau), \quad \text{and} \quad u(0) = u_0.$$

For $n = 0$ this leads to the linear system

$$(K_\tau + \lambda M_\tau)\, \boldsymbol{u}_1 = u_0 C_\tau \boldsymbol{v}$$

with the vector $v_1 = 1$ and $v_k = 0$ for $k = 2, \ldots, N_t$ and with the solution vector $\boldsymbol{u}_1 \in \mathbb{R}^{N_t}$ for the first step. Therefore the value at the endpoint τ of the discrete solution $u_\tau^1 \in \mathbb{P}^{p_t}(0,\tau)$ is given by

$$u_\tau^1(\tau) = \boldsymbol{v}^\top \boldsymbol{u}_1 = u_0 \boldsymbol{v}^\top \left(K_\tau + \lambda M_\tau\right)^{-1} C_\tau \boldsymbol{v} \in \mathbb{R}.$$

Hence the stability function $R(z)$ with $z = -\lambda\tau$ is given by

$$R(z) = R(-\lambda\tau) = \boldsymbol{v}^\top \left(K_\tau + \lambda M_\tau\right)^{-1} C_\tau \boldsymbol{v}. \tag{2.7}$$

Since the matrix C_τ has rank one, only one eigenvalue can be nonzero and with (2.7), it is easy to see that this eigenvalue is given by $R(-\lambda\tau)$. ∎

3. Multigrid method. We present now our new space-time multigrid method to solve the linear space-time system (1.3), which we rewrite in compact form as

$$\mathcal{L}_{\tau,h} \boldsymbol{u} = \boldsymbol{f}. \tag{3.1}$$

For an introduction to multigrid methods, see [22, 41, 44, 46]. We need a hierarchical sequence of space-time meshes $\mathcal{T}_{N_L}$ for $L = 0, \ldots, M_L$, where $L = 0$ is the coarsest level and $L = M_L$ is the finest level. This sequence has to be chosen in an appropriate way; see section 4.2. In general we consider two types of coarsening strategies, i.e., coarsening only with respect to time and coarsening in space and time. For each space-time mesh $\mathcal{T}_{N_L}$ we compute the system matrix $\mathcal{L}_{\tau_L,h_L}$ for $L = 0, \ldots, M_L$. On the last (finest) level M_L, we obtain the original system (3.1), i.e., $\mathcal{L}_{\tau_{M_L},h_{M_L}} = \mathcal{L}_{\tau,h}$.

We denote by $\mathcal{S}^\nu_{\tau_L,h_L}$ the inexact damped block Jacobi smoother with $\nu \in \mathbb{N}$ steps,

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \omega_t (\widetilde{D}_{\tau_L,h_L})^{-1} \left[\boldsymbol{f} - \mathcal{L}_{\tau_L,h_L} \boldsymbol{u}^{(k)}\right]. \tag{3.2}$$

Here $\widetilde{D}^{-1}_{\tau_L,h_L}$ denotes an approximation of the inverse of the block diagonal matrix $D_{\tau_L,h_L} := \operatorname{diag}\{A_{\tau_L,h_L}\}_{n=1}^{N_L}$, where a block $A_{\tau_L,h_L} := M_{h_L} \otimes K_{\tau_L} + K_{h_L} \otimes M_{\tau_L}$ corresponds to one time step. We will consider in particular approximating $(D_{\tau_L,h_L})^{-1}$ by applying one multigrid V-cycle in space at each time step, using a standard tensor product multigrid, like in [3]. Moreover $\omega_t > 0$ is a fixed damping parameter where the optimal choice will be derived in section 4.1.

For the prolongation operator $\mathcal{P}^L$ we use the standard interpolation from coarse space-time grids to the next finer space-time grids, where we always combine two time steps to one coarse time step. The prolongation operator will thus depend on the space-time hierarchy chosen. The restriction operator is the adjoint of the prolongation operator, $\mathcal{R}^L = (\mathcal{P}^L)^\top$. For the exact definitions see section 4.2. With $\nu_1, \nu_2 \in \mathbb{N}$ we denote the number of pre- and post-smoothing steps, and $\gamma \in \mathbb{N}$ defines the cycle index, where typical choices are $\gamma = 1$ (V-cycle) and $\gamma = 2$ (W-cycle). On the coarsest level $L = 0$ we solve the linear system, which consists of only one time step, exactly by using an LU-factorization for the system matrix $\mathcal{L}_{\tau_0,h_0}$. For a given initial guess we apply this space-time multigrid cycle several times, until we have reached a given relative error reduction $\varepsilon_{\mathrm{MG}}$.

To study the convergence behavior of our space-time multigrid method, we use local Fourier mode analysis.

4. Local Fourier mode analysis. The aim of this analysis is to derive smoothing factors and two-grid convergence factors for arbitary order discontinuous Galerkin

time discretization schemes, i.e., for arbitrary polynomial degrees $p_t \in \mathbb{N}_0$. To do so, we group the unkowns for one time step and one Fourier mode together and use the ideas which are used to analyze block-systems; see [41]. Hence, our symbol for the smoother will be a matrix of dimension $N_t \times N_t$ with $N_t = p_t + 1$ and the symbol for the two-grid operators will be of size $2^{d+1}N_t \times 2^{d+1}N_t$. For simplicity we assume that $\Omega = (0,1)$ is a one dimensional domain, which is divided into uniform elements with mesh size h. The analysis for higher dimensions is more technical, but the tools and results stay the same as for the one dimensional case. The standard one dimensional mass and stiffness matrices are

$$M_h = \frac{h}{6}\begin{pmatrix} 4 & 1 & & \\ 1 & 4 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 4 \end{pmatrix}, \qquad K_h = \frac{1}{h}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}.$$

4.1. Smoothing analysis. The iteration matrix of the exact damped block Jacobi is

$$\mathcal{S}^{\nu}_{\tau_L,h_L} = \left[I - \omega_t (D_{\tau_L,h_L})^{-1}\mathcal{L}_{\tau_L,h_L}\right]^{\nu},$$

where D_{τ_L,h_L} is a block diagonal matrix with blocks A_{τ_L,h_L}. We first use the exact inverse of the diagonal matrix D_{τ_L,h_L} in our analysis, and the V-cycle approximation is studied later; see section 4.3. We denote by $N_{L_t} \in \mathbb{N}$ the number of time steps and by $N_{L_x} \in \mathbb{N}$ the degrees of freedom in space for level $L \in \mathbb{N}_0$. To transform our problem into the frequency domain, where we group all the unkowns for one time step together, we need the following definition and lemma.

DEFINITION 4.1 (Fourier modes, Fourier frequencies). *Let $N_L \in \mathbb{N}$. Then the vector valued function $\boldsymbol{\varphi}_\ell(\theta_k) := e^{\mathrm{i}\ell\theta_k}$, $\ell = 1, \dots, N_L$ is called the Fourier mode with frequency*

$$\theta_k \in \Theta_L := \left\{ \frac{2k\pi}{N_L} : k = 1 - \frac{N_L}{2}, \dots, \frac{N_L}{2} \right\} \subset (-\pi, \pi].$$

The frequencies Θ_L are further separated into low and high frequencies

$$\Theta_L^{\text{low}} := \Theta_L \cap \left(-\frac{\pi}{2}, \frac{\pi}{2}\right],$$
$$\Theta_L^{\text{high}} := \Theta_L \cap \left(\left(-\pi, -\frac{\pi}{2}\right] \cup \left(\frac{\pi}{2}, \pi\right]\right) = \Theta_L \setminus \Theta_L^{\text{low}}.$$

LEMMA 4.2. *Let $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{N_{L_t}})^\top \in \mathbb{R}^{N_t N_{L_x} N_{L_t}}$ for $N_t, N_{L_x}, N_{L_t} \in \mathbb{N}$, where we assume that N_{L_x} and N_{L_t} are even numbers. Then the vector $\boldsymbol{u}$ can be written as*

$$\boldsymbol{u} = \sum_{\theta_x \in \Theta_{L_x}} \sum_{\theta_t \in \Theta_{L_t}} \boldsymbol{\psi}(\theta_x, \theta_t)$$

with the vectors

$$\boldsymbol{\psi}_{n,j}(\theta_x, \theta_t) := U(\theta_x, \theta_t)\boldsymbol{\Phi}_{n,j}(\theta_x, \theta_t) \in \mathbb{C}^{N_t}, \quad \text{where } \boldsymbol{\Phi}_{n,j}(\theta_x, \theta_t) \in \mathbb{C}^{N_t}$$

with $\boldsymbol{\Phi}_{n,j,\ell}(\theta_x,\theta_t) := \boldsymbol{\varphi}_n(\theta_t)\boldsymbol{\varphi}_j(\theta_x)$ *for* $n = 1,\dots,N_{L_t}$, $j = 1,\dots,N_{L_x}$, $\ell = 1,\dots,N_t$ *and with the coefficient matrix*

$$U(\theta_x,\theta_t) := \operatorname{diag}(\hat{u}_1,\dots,\hat{u}_{N_t}) \in \mathbb{C}^{N_t\times N_t}$$

with the coefficients for $\theta_x \in \Theta_{L_x}$ *and* $\theta_t \in \Theta_{L_t}$

$$\hat{u}_\ell := \frac{1}{N_{L_x}}\frac{1}{N_{L_t}}\sum_{j=1}^{N_{L_x}}\sum_{n=1}^{N_{L_t}} \boldsymbol{u}_{n,j,\ell}\boldsymbol{\varphi}_j(-\theta_x)\boldsymbol{\varphi}_n(-\theta_t).$$

Note that the indexing of the vectors above is the same as defined in the introduction.

Proof. The statement of this Lemma follows by simply applying the standard discrete Fourier transform, see [46, Theorem 7.3.1], in space and time by separating the unkowns for each time step. ∎

In view of Lemma 4.2 a vector $\boldsymbol{u} = (\boldsymbol{u}_1,\boldsymbol{u}_2,\dots,\boldsymbol{u}_{N_{L_t}})^\top$ can be written as a linear combination of the vectors $\boldsymbol{\Phi}(\theta_x,\theta_t)$, where the coefficients are matrices of dimension $N_t \times N_t$. Note that the matrix $U(\theta_x,\theta_t) \in \mathbb{C}^{N_t\times N_t}$ in Lemma 4.2 is a diagonal matrix. Later on we have to analyze the action of the Fourier symbols (which are in general dense matrices of dimensions $N_t \times N_t$) onto the coefficient matrix $U(\theta_x,\theta_t)$. Hence we have to deal with arbitary dense matrices, which motivates the next definition.

DEFINITION 4.3 (Fourier space). *For the space and time level* L_x *and* L_t *let* $\theta_x \in \Theta_{L_x}$ *and* $\theta_t \in \Theta_{L_t}$. *Further let the vector* $\boldsymbol{\Phi}(\theta_x,\theta_t)$ *be defined as in Lemma* 4.2. *Then we define the linear space of Fourier modes with frequencies* (θ_x,θ_t) *as*

$$\begin{aligned}\Psi_{L_x,L_t}(\theta_x,\theta_t) &:= \operatorname{span}\{\boldsymbol{\Phi}(\theta_x,\theta_t)\}\\ &:= \big\{\boldsymbol{\psi}(\theta_x,\theta_t) \in \mathbb{C}^{N_tN_{L_x}N_{L_t}} : \boldsymbol{\psi}_{n,j}(\theta_x,\theta_t) := U\boldsymbol{\Phi}_{n,j}(\theta_x,\theta_t),\\ &\qquad\qquad \text{for } n = 1,\dots,N_{L_t}, j = 1,\dots,N_{L_x} \text{ and } U \in \mathbb{C}^{N_t\times N_t}\big\}.\end{aligned}$$

LEMMA 4.4 (shifting equality). *For the space and time level* L_x *and* L_t *let* $\theta_x \in \Theta_{L_x}$ *and* $\theta_t \in \Theta_{L_t}$. *Further let* $\boldsymbol{\psi}(\theta_x,\theta_t) \in \Psi_{L_x,L_t}(\theta_x,\theta_t)$. *Then we have the shifting equalities*

$$\boldsymbol{\psi}_{n-1,j}(\theta_x,\theta_t) = e^{-\mathrm{i}\theta_t}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t), \quad \boldsymbol{\psi}_{n,j-1}(\theta_x,\theta_t) = e^{-\mathrm{i}\theta_x}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)$$

for $n = 2,\dots,N_{L_t}$ *and* $j = 2,\dots,N_{L_x}$.

Proof. The result follows from the fact that

$$\boldsymbol{\varphi}_{n-1}(\theta) = e^{\mathrm{i}(n-1)\theta} = e^{-\mathrm{i}\theta}e^{\mathrm{i}n\theta} = e^{-\mathrm{i}\theta}\boldsymbol{\varphi}_n(\theta).$$ ∎

LEMMA 4.5 (Fourier symbol of $\mathcal{L}_{\tau_L,h_L}$). *For the frequencies* $\theta_x \in \Theta_{L_x}$ *and* $\theta_t \in \Theta_{L_t}$ *we consider the vector* $\boldsymbol{\psi}(\theta_x,\theta_t) \in \Psi_{L_x,L_t}(\theta_x,\theta_t)$. *Then for* $n = 2,\dots,N_{L_t}$ *and* $j = 2,\dots,N_{L_x}-1$ *we have*

$$\big(\mathcal{L}_{\tau_L,h_L}\boldsymbol{\psi}^{L_x,L_t}(\theta_x,\theta_t)\big)_{n,j} = \hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t),$$

where the Fourier symbol is given by

$$\hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t) := \frac{h_L}{3}(2+\cos(\theta_x))\left[K_{\tau_L} + h_L^{-2}\beta(\theta_x)M_{\tau_L} - e^{-\mathrm{i}\theta_t}C_{\tau_L}\right] \in \mathbb{C}^{N_t\times N_t}$$

with the function $\beta(\theta_x) := 6\,\frac{1-\cos(\theta_x)}{2+\cos(\theta_x)} \in [0,12]$.

Proof. Let $\boldsymbol{\psi}(\theta_x,\theta_t) \in \Psi_{L_x,L_t}(\theta_x,\theta_t)$. Then we have for $n = 2,\dots,N_{L_t}$ and using Lemma 4.4

$$\begin{aligned}\left(\mathcal{L}_{\tau_L,h_L}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_n &= B_{\tau_L,h_L}\boldsymbol{\psi}_{n-1}(\theta_x,\theta_t) + A_{\tau_L,h_L}\boldsymbol{\psi}_n(\theta_x,\theta_t)\\ &= \left(e^{-\mathrm{i}\theta_t}B_{\tau_L,h_L} + A_{\tau_L,h_L}\right)\boldsymbol{\psi}_n(\theta_x,\theta_t).\end{aligned}$$

Hence, we have to study the action of $A_{\tau,h}$ and $B_{\tau,h}$ on the local vector $\boldsymbol{\psi}_n(\theta_x,\theta_t)$. By using the definition of $B_{\tau,h}$, we obtain for $j = 2,\dots,N_{L_x}-1$ and $\ell = 1,\dots,N_t$ using Lemma 4.4

$$\begin{aligned}&\left(B_{\tau_L,h_L}\boldsymbol{\psi}_n(\theta_x,\theta_t)\right)_{j,\ell}\\ &\quad= -\sum_{i=1}^{N_{L_x}}\sum_{k=1}^{N_t} M_{h_L}[j,i]C_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,i,k}(\theta_x,\theta_t)\\ &\quad= -\sum_{k=1}^{N_t}\frac{h_L}{6}\Big(\boldsymbol{\psi}_{n,j-1,k}(\theta_x,\theta_t) + 4\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t) + \boldsymbol{\psi}_{n,j+1,k}(\theta_x,\theta_t)\Big)C_{\tau_L}[\ell,k]\\ &\quad= -\frac{h_L}{6}\sum_{k=1}^{N_t} C_{\tau_L}[\ell,k]\left(e^{-\mathrm{i}\theta_x} + 4 + e^{\mathrm{i}\theta_x}\right)\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t)\\ &\quad= -\frac{h_L}{3}\left(2+\cos(\theta_x)\right)\sum_{k=1}^{N_t} C_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t)\\ &\quad= -\frac{h_L}{3}\left(2+\cos(\theta_x)\right)\Big(C_{\tau_L}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\Big)_\ell.\end{aligned}$$

Next we study the action of the matrix $A_{\tau,h}$ on the local vector $\boldsymbol{\psi}_n(\theta_x,\theta_t)$:

$$\begin{aligned}&\left(A_{\tau_L,h_L}\boldsymbol{\psi}_n(\theta_x,\theta_t)\right)_{j,\ell}\\ &\quad= \sum_{i=1}^{N_{L_x}}\sum_{k=1}^{N_t} M_{h_L}[j,i]K_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,i,k}(\theta_x,\theta_t)\\ &\qquad+ \sum_{i=1}^{N_{L_x}}\sum_{k=1}^{N_t} K_{h_L}[j,i]M_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,i,k}(\theta_x,\theta_t)\\ &\quad= \frac{h_L}{3}\left(2+\cos(\theta_x)\right)\sum_{k=1}^{N_t} K_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t)\\ &\qquad+ \sum_{k=1}^{N_t}\frac{1}{h_L}\Big(-\boldsymbol{\psi}_{n,j-1,k}(\theta_x,\theta_t) + 2\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t) - \boldsymbol{\psi}_{n,j+1,k}(\theta_x,\theta_t)\Big)M_{\tau_L}[\ell,k]\\ &\quad= \frac{h_L}{3}\left(2+\cos(\theta_x)\right)\Big(K_{\tau_L}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\Big)_\ell\\ &\qquad+ \frac{2}{h_L}(1-\cos(\theta_x))\sum_{k=1}^{N_t} M_{\tau_L}[\ell,k]\boldsymbol{\psi}_{n,j,k}(\theta_x,\theta_t)\\ &\quad= \left(\left[\frac{h_L}{3}\left(2+\cos(\theta_x)\right)K_{\tau_L} + \frac{2}{h_L}(1-\cos(\theta_x))M_{\tau_L}\right]\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\right)_\ell,\end{aligned}$$

where we used Lemma 4.4. Hence we conclude the proof with

$$\begin{aligned}
&\left(\mathcal{L}_{\tau_L,h_L}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{n,j}\\
&\quad= \frac{h_L}{3}\left(2+\cos(\theta_x)\right)\left(K_{\tau_L}-e^{-\mathbb{i}\theta_t}C_{\tau_L}\right)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\\
&\qquad+\frac{2}{h_L}(1-\cos(\theta_x))M_{\tau_L}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\\
&\quad= \frac{h_L}{3}\left(2+\cos(\theta_x)\right)\left(K_{\tau_L}+6h_L^{-2}\frac{1-\cos(\theta_x)}{2+\cos(\theta_x)}M_{\tau_L}-e^{-\mathbb{i}\theta_t}C_{\tau_L}\right)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\\
&\quad= \frac{h_L}{3}\left(2+\cos(\theta_x)\right)\left(K_{\tau_L}+h_L^{-2}\beta(\theta_x)M_{\tau_L}-e^{-\mathbb{i}\theta_t}C_{\tau_L}\right)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t).
\end{aligned}$$

∎

If we assume periodic boundary conditions in space and time, i.e.,

$$\begin{aligned}
u(t,0)&=u(t,1) &&\text{for } t\in(0,T),\\
u(0,x)&=u(T,x) &&\text{for } x\in\Omega=(0,1),
\end{aligned}\tag{4.1}$$

we obtain from Lemma 4.5 the mapping property

$$\begin{aligned}
\mathcal{L}_{\tau_L,h_L}&:\Psi_{L_x,L_t}(\theta_x,\theta_t)\to\Psi_{L_x,L_t}(\theta_x,\theta_t),\\
&\qquad U\mapsto\hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t)U.
\end{aligned}\tag{4.2}$$

LEMMA 4.6 (mapping property of $\mathcal{S}^{\nu}_{\tau_L,h_L}$). *For the frequencies $\theta_x\in\Theta_{L_x}$ and $\theta_t\in\Theta_{L_t}$ we consider the vector $\boldsymbol{\psi}(\theta_x,\theta_t)\in\Psi_{L_x,L_t}(\theta_x,\theta_t)$. Then under the assumption of periodic boundary conditions* (4.1)*, we have for $n=1,\dots,N_{L_t}$ and $j=1,\dots,N_{L_x}$*

$$\left(\mathcal{S}^{\nu}_{\tau_L,h_L}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{n,j}=\left[\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right]^{\nu}\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t),$$

where the Fourier symbol is given by

$$\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t):=(1-\omega_t)I_{N_t}+\omega_t e^{-\mathbb{i}\theta_t}\left(K_{\tau_L}+h_L^{-2}\beta(\theta_x)M_{\tau_L}\right)^{-1}C_{\tau_L}\in\mathbb{C}^{N_t\times N_t}$$

with the function $\beta(\theta_x)$ as defined in Lemma 4.5.

Proof. Let $\boldsymbol{\psi}(\theta_x,\theta_t)\in\Psi_{L_x,L_t}(\theta_x,\theta_t)$, and then for a fixed $n=1,\dots,N_{L_t}$ and a fixed $j=1,\dots,N_{L_x}$ we have that

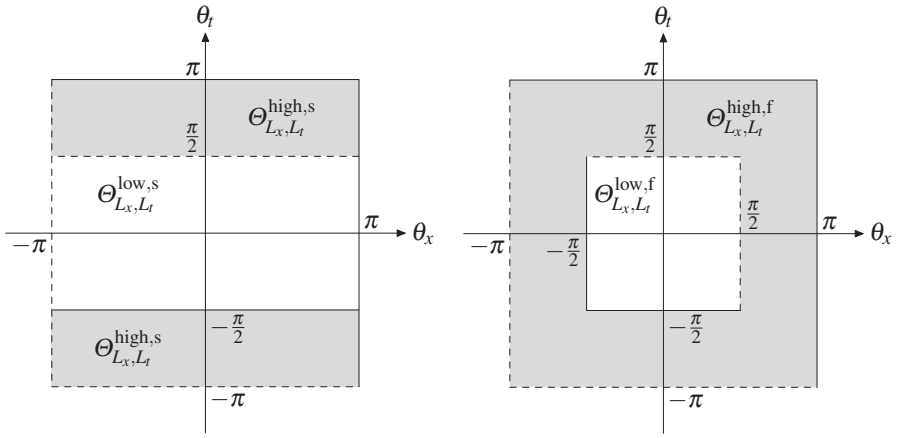
$$\begin{aligned}
\left(\mathcal{S}^{1}_{\tau_L,h_L}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{n,j}&=\left(\left(I_{N_tN_{L_x}N_{L_t}}-\omega_t(D_{\tau_L,h_L})^{-1}\mathcal{L}_{\tau_L,h_L}\right)\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{n,j}\\
&=\left(I_{N_t}-\omega_t\left(\hat{A}_{\tau_L,h_L}(\theta_x)\right)^{-1}\hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t)\right)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)\\
&=:\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\boldsymbol{\psi}_{n,j}(\theta_x,\theta_t)
\end{aligned}$$

with

$$\begin{aligned}
\hat{A}_{\tau_L,h_L}(\theta_x):&=\frac{h_L}{3}\left(2+\cos(\theta_x)\right)K_{\tau_L}+\frac{2}{h_L}(1-\cos(\theta_x))M_{\tau_L}\\
&=\frac{h_L}{3}\left(2+\cos(\theta_x)\right)\left(K_{\tau_L}+h_L^{-2}\beta(\theta_x)M_{\tau_L}\right).
\end{aligned}$$

Further calculations give

$$\left(\hat{A}_{\tau_L,h_L}(\theta_x)\right)^{-1}\hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t)=I_{N_t}-e^{-\mathbb{i}\theta_t}\left(K_{\tau_L}+h_L^{-2}\beta(\theta_x)M_{\tau_L}\right)^{-1}C_{\tau_L}.$$



(a) Semi-coarsening. (b) Full space-time coarsening.

FIG. 1. *Low and high frequencies θ_x and θ_t for semi-coarsening and full space-time coarsening.*

Hence we have

$$\begin{aligned}\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t) &= I_{N_t} - \omega_t \left(I_{N_t} - e^{-\mathrm{i}\theta_t}\left(K_{\tau_L} + h_L^{-2}\beta(\theta_x)M_{\tau_L}\right)^{-1} C_{\tau_L}\right)\\ &= (1-\omega_t)I_{N_t} + \omega_t e^{-\mathrm{i}\theta_t}\left(K_{\tau_L} + h_L^{-2}\beta(\theta_x)M_{\tau_L}\right)^{-1} C_{\tau_L}.\end{aligned}$$

By induction this completes the proof. ∎

In view of Lemma 4.6, the following mapping property holds when periodic boundary conditions are assumed:

$$\begin{aligned}\mathcal{S}^{\nu}_{\tau_L,h_L} : \Psi_{L_x,L_t}(\theta_x,\theta_t) &\to \Psi_{L_x,L_t}(\theta_x,\theta_t),\\ U &\mapsto (\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t))^{\nu}U.\end{aligned} \tag{4.3}$$

Next we will analyze the smoothing behavior for the high frequencies. To do so, we consider two coarsening strategies: semi-coarsening in time (i.e., combining two time intervals to one coarser time interval) and full space-time coarsening.

DEFINITION 4.7 (high and low frequency ranges). *For the space and time level L_x and L_t we define the set of frequencies*

$$\Theta_{L_x,L_t} := \{(\theta_x,\theta_t) : \theta_x \in \Theta_{L_x} \text{ and } \theta_t \in \Theta_{L_t}\} \subset (-\pi,\pi]^2$$

and the sets of low and high frequencies with respect to semi-coarsening in time,

$$\Theta^{\text{low,s}}_{L_x,L_t} := \Theta_{L_x,L_t} \cap (-\pi,\pi] \times \left(-\frac{\pi}{2},\frac{\pi}{2}\right], \quad \Theta^{\text{high,s}}_{L_x,L_t} := \Theta_{L_x,L_t} \setminus \Theta^{\text{low,s}}_{L_x,L_t}$$

and full space-time coarsening

$$\Theta^{\text{low,f}}_{L_x,L_t} := \Theta_{L_x,L_t} \cap \left(-\frac{\pi}{2},\frac{\pi}{2}\right]^2, \quad \Theta^{\text{high,f}}_{L_x,L_t} := \Theta_{L_x,L_t} \setminus \Theta^{\text{low,f}}_{L_x,L_t}.$$

In Figure 1, the high and low frequencies are illustrated for the two coarsening strategies. To measure the smoothing property of the damped block Jacobi method we use the following definitions.

DEFINITION 4.8 (asymptotic smoothing factors). *Let $\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)$ be the symbol of the block Jacobi smoother. Then the smoothing factor for semi-coarsening in time is*

$$\varrho_{\mathrm{s}}(\hat{\mathcal{S}}) := \max\left\{\varrho(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)) : (\theta_x,\theta_t) \in \Theta^{\mathrm{high,s}}_{L_x,L_t}\right\},$$

and the smoothing factor for full space-time coarsening is

$$\varrho_{\mathrm{f}}(\hat{\mathcal{S}}) := \max\left\{\varrho(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)) : (\theta_x,\theta_t) \in \Theta^{\mathrm{high,f}}_{L_x,L_t}\right\}.$$

To study the smoothing behavior, we need the eigenvalues of the Fourier symbol $\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)$. It will turn out in the next lemma that the A-stability of the underlying time discretization scheme plays an important role, which is reflected by the A-stability function $R(z)$.

LEMMA 4.9. *The spectral radius of the Fourier symbol $\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)$ is given by*

$$\rho\left(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right) = \max\left\{|1-\omega_t|, \hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta_t)\right\}$$

with

$$\left(\hat{\mathcal{S}}(\omega_t,\alpha,\theta_t)\right)^2 := (1-\omega_t)^2 + 2\omega_t(1-\omega_t)\alpha\cos(\theta_t) + \alpha^2\omega_t^2,$$

where $\alpha(\theta_x,\mu) := R(-\mu\beta(\theta_x))$ and $R(z)$ is the (p_t,p_t+1) subdiagonal Padé approximation of the exponential function e^z and $\mu := \tau_L h_L^{-2}$ is a discretization parameter.

Proof. The eigenvalues of the Fourier symbol

$$\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t) = (1-\omega_t)I_{N_t} + \omega_t e^{-\mathbb{i}\theta_t}\left(K_{\tau_L} + h_L^{-2}\beta(\theta_x)M_{\tau_L}\right)^{-1} C_{\tau_L}$$

are given by

$$\sigma(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)) = 1-\omega_t + e^{-\mathbb{i}\theta_k}\omega_t\sigma((K_{\tau_L} + h_L^{-2}\beta(\theta_x)M_{\tau_L})^{-1}C_{\tau_L}).$$

With Lemma 2.4 and using the definition of $\alpha(\theta_x,\mu)$ we are now able to compute the spectrum as

$$\sigma(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)) = \left\{1-\omega_t, 1-\omega_t + e^{-\mathbb{i}\theta_k}\omega_t\alpha(\theta_x,\mu)\right\}.$$

Hence we obtain the spectral radius

$$\varrho(\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)) = \max\left\{|1-\omega_t|, \left|1-\omega_t + e^{-\mathbb{i}\theta_k}\omega_t\alpha(\theta_x,\mu)\right|\right\}.$$

Direct calculations lead to

$$\left|1-\omega_t + e^{-\mathbb{i}\theta_k}\omega_t\alpha(\theta_x,\mu)\right|^2 = (1-\omega_t)^2 + 2\omega_t(1-\omega_t)\alpha(\theta_x,\mu)\cos(\theta_k) + (\alpha(\theta_x,\mu))^2\omega_t^2,$$

which completes the proof. ∎

Next, we study the smoothing behavior of the damped block Jacobi iteration for the case when semi-coarsening in time is applied.

LEMMA 4.10. *For the function*

$$\left(\hat{\mathcal{S}}(\omega_t,\alpha,\theta_t)\right)^2 := (1-\omega_t)^2 + 2\omega_t(1-\omega_t)\alpha\cos(\theta_t) + \alpha^2\omega_t^2$$

with $\alpha = \alpha(\theta_x,\mu)$ *as defined in Lemma* 4.9 *and even polynomial degrees* p_t*, the min-max principle*

$$\inf_{\omega_t\in(0,1]} \sup_{\substack{\theta_t\in[\frac{\pi}{2},\pi]\\ \theta_x\in[0,\pi]}} \hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta_t) = \frac{1}{\sqrt{2}}$$

holds for any discretization parameter $\mu \geq 0$ *with the optimal parameter*

$$\omega_t^* = \frac{1}{2}$$

and the worst case frequencies

$$\theta_t^* = \frac{\pi}{2} \qquad \textit{and} \qquad \theta_x^* = 0.$$

Proof. Since we consider even polynomial degrees p_t, the (p_t, p_t+1) subdiagonal Padé approximation $R(z)$ of the exponential function e^z is positive for all $z \leq 0$. Hence we also have that $\alpha(\theta_x,\mu) = R(-\mu\beta(\theta_x))$ is positive for all $\mu \geq 0$ and $\theta_x \in [0,\pi]$. Since $\omega_t \in (0,1]$, we obtain

$$\theta_t^* := \operatorname*{argsup}_{\theta_t\in[\frac{\pi}{2},\pi]} \hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta_t) = \frac{\pi}{2}.$$

Since $\alpha(0,\mu) = 1$ and $|\alpha(\theta_x,\mu)| \leq 1$ for all $\theta_x \in [0,\pi]$ and $\mu \geq 0$ we get

$$\theta_x^* := \operatorname*{argsup}_{\theta_x\in[0,\pi]} \hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta^*) = 0.$$

Hence we have to find the infimum of

$$\left(\hat{\mathcal{S}}(\omega_t,\alpha(\theta_x^*,\mu),\theta_t^*)\right)^2 = (1-\omega_t)^2 + \omega_t^2,$$

which is obtained for $\omega_t^* = \frac{1}{2}$. This implies that

$$\left(\hat{\mathcal{S}}(\omega_t^*,\alpha(\theta_x^*,\mu),\theta_t^*)\right)^2 = \frac{1}{2},$$

which completes the proof. ∎

THEOREM 4.11 (asymptotic smoothing factor for semi-coarsening). *For the function*

$$\left(\hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta_t)\right)^2 = (1-\omega_t)^2 + 2\omega_t(1-\omega_t)\alpha(\theta_x,\mu)\cos(\theta_t) + (\alpha(\theta_x,\mu))^2\omega_t^2,$$

where $\alpha = \alpha(\theta_x,\mu)$ *is defined as in Lemma* 4.9 *and the choice* $\omega_t^* = \frac{1}{2}$ *and any polynomial degree* $p_t \in \mathbb{N}_0$*, we have the bound*

$$\sup_{\substack{\theta_t\in[\frac{\pi}{2},\pi]\\ \theta_x\in[0,\pi]}} \hat{\mathcal{S}}(\omega_t,\alpha(\theta_x,\mu),\theta_t) \leq \frac{1}{\sqrt{2}}.$$

Proof. For even polynomial degrees p_t, we can apply Lemma 4.10 to get the bound stated. For odd polynomial degrees, the (p_t, p_t+1) subdiagonal Padé approximation $R(z)$ of the exponential function e^z is negative for large negative values of z. If the value of $\alpha(\theta_x^*, \mu) = R(-\mu\beta(\theta_x^*))$ for the optimal parameter $\theta_x^* \in [0, \pi]$ is positive, we get directly the bound of Theorem 4.11. Otherwise we obtain

$$\theta_t^* := \operatorname*{argsup}_{\theta_t \in [\frac{\pi}{2}, \pi]} \hat{\mathcal{S}}(\omega_t, \alpha(\theta_x^*, \mu), \theta_t) = \pi.$$

For a negative $\alpha(\theta_x^*, \mu)$, this implies that

$$\sup_{\substack{\theta_t \in [\frac{\pi}{2}, \pi] \\ \theta_x \in [0, \pi]}} \hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu), \theta_t) \leq \frac{1}{2}(1 + |\alpha(\theta_x^*, \mu)|) \leq \frac{3}{4}(\sqrt{3} - 1) < \frac{1}{\sqrt{2}},$$

since any subdiagonal (p_t, p_t+1) Padé approximation $R(z)$ is bounded from below by $R(z) \geq \frac{1}{2}(5 - 3\sqrt{3})$ for all $z < 0$. ∎

Theorem 4.11 shows that the asymptotic smoothing factor for semi-coarsening in time is bounded by $\varrho_{\mathrm{s}}(\hat{\mathcal{S}}) \leq \frac{1}{\sqrt{2}}$. Hence, by applying the damped block Jacobi smoother with the optimal damping parameter $\omega_t^* = \frac{1}{2}$, the error components in the high frequencies $\Theta_{L_x, L_t}^{\mathrm{high,s}}$ are asymptotically damped by a factor of at least $\frac{1}{\sqrt{2}}$.

THEOREM 4.12 (asymptotic smoothing factor for full space-time coarsening). *For the optimal choice of the damping parameter $\omega_t^* = \frac{1}{2}$, we have*

$$\sup_{\theta_t \in [0, \pi]} \hat{\mathcal{S}}(\omega_t^*, \alpha, \theta_t) = \frac{1}{2}(1 + |\alpha|)$$

with the worst case frequency

$$\theta_t^* = \begin{cases} 0 & \alpha \geq 0, \\ \pi & \alpha < 0. \end{cases}$$

Proof. Let $\alpha \in \mathbb{R}$. For the optimal damping parameter $\omega_t^* = \frac{1}{2}$ we have

$$\left(\hat{\mathcal{S}}(\omega_t^*, \alpha, \theta_t)\right)^2 = \frac{1}{4}\left(1 + 2\alpha\cos(\theta_t) + \alpha^2\right).$$

First we study the case $\alpha \geq 0$, where we get

$$\theta_t^* := \operatorname*{argsup}_{\theta_t \in [0, \pi]} \hat{\mathcal{S}}(\omega_t^*, \alpha, \theta_t) = 0.$$

For the case $\alpha < 0$ we obtain

$$\theta_t^* := \operatorname*{argsup}_{\theta_t \in [0, \pi]} \hat{\mathcal{S}}(\omega_t^*, \alpha, \theta_t) = \pi.$$

This implies that

$$\left(\hat{\mathcal{S}}(\omega_t^*, \alpha, \theta_t*)\right)^2 = \frac{1}{4}\left(1 + 2\,|\alpha| + \alpha^2\right) = \frac{1}{4}\left(1 + |\alpha|\right)^2,$$

which completes the proof. ∎

Theorem 4.12 shows that we obtain good smoothing behavior for the high frequencies with respect to the space discretization, i.e., $\theta_x \in \Theta_{L_x}^{\text{high}}$, if $\alpha = \alpha(\theta_x, \mu)$ is sufficiently small for any frequency $\theta_x \in [\frac{\pi}{2}, \pi]$. Hence combining Theorem 4.11 with Theorem 4.12, we see that good smoothing behavior can be obtained for all frequencies $(\theta_x, \theta_t) \in \Theta_{L_x,L_t}^{\text{high,f}}$ if the function $\alpha = \alpha(\theta_x, \mu)$ is sufficiently small. This results in a restriction on the discretization parameter μ. In the next subsection we will analyze the smoothing behavior with respect to this discretization parameter. This analysis will help us to switch between the two different coarsening strategies in the multigrid approach.

4.1.1. Smoothing factors for different discretization parameters. With the next lemma we will analyze the behavior of the smoothing factor $\varrho_{\text{f}}(\hat{\mathcal{S}})$ with respect to the discretization parameter μ for even polynomial degrees $p_t \in \mathbb{N}_0$.

LEMMA 4.13. *Let $p_t \in \mathbb{N}_0$ be even. Then for the optimal choice of the damping parameter $\omega_t^* = \frac{1}{2}$ we have*

$$\sup_{\substack{\theta_t \in [0,\pi] \\ \theta_x \in [\frac{\pi}{2},\pi]}} \hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu), \theta_t) = \frac{1}{2}(1 + R(-3\mu)),$$

where $R(z)$ is the (p_t, p_t+1) subdiagonal Padé approximation of the exponential function e^z.

Proof. In view of Theorem 4.12 it remains to compute the supremum

$$\sup_{\theta_x \in [\frac{\pi}{2},\pi]} \frac{1}{2}(1 + |\alpha(\theta_x, \mu)|).$$

Since for even polynomial degrees p_t the function $\alpha(\theta_x, \mu) = R(-\mu\beta(\theta_x))$ is monotonically decreasing in $\beta(\theta_x)$, the supremum is obtained for $\beta(\theta_x) = 3$, since $\beta(\theta_x) \in [3, 12]$ for $\theta_x \in [\frac{\pi}{2}, \pi]$. This implies that $\theta_x^* = \frac{\pi}{2}$, and we obtain the statement of the lemma with

$$\begin{aligned}\sup_{\substack{\theta_t \in [0,\pi] \\ \theta_x \in [\frac{\pi}{2},\pi]}} \hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu)), \theta_t) &= \hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x^*, \mu), \theta_t^*) \\ &= \frac{1}{2}(1 + |\alpha(\theta_x^*, \mu)|) = \frac{1}{2}(1 + R(-3\mu)).\end{aligned}$$

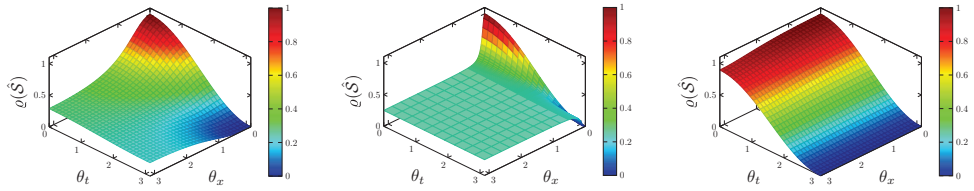
∎

The proof of Lemma 4.13 only holds for even polynomial degrees, but the result is also true for odd polynomial degrees p_t, though the proof gets more involved, since the Padé approximation $R(z)$, $z \leq 0$ is not monotonically decreasing for odd polynomial degrees.

Remark 4.14. In view of Lemma 4.13 we obtain a good smoothing behavior for the high frequencies in space $\theta_x \in \Theta_{L_x}^{\text{high}}$, i.e., $\varrho_{\text{f}}(\hat{\mathcal{S}}) \leq \frac{1}{\sqrt{2}}$ (see Theorem 4.11), if the discretization parameter μ is large enough, i.e.,

$$\mu \geq \mu_{p_t}^* \quad \text{with} \quad R(-3\mu_{p_t}^*) = \sqrt{2} - 1. \tag{4.4}$$

Hence we are able to compute the critical discretization parameter $\mu_{p_t}^*$ with respect



(a) $\mu = 1$. (b) $\mu = 100$. (c) $\mu = 0.01$.

FIG. 2. *Smoothing factor* $\hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu), \theta_t)^2$ *for* $\theta_x, \theta_t \in [0, \pi]$ *for* $p_t = 0$ *and different discretization parameters* μ.

to the polynomial degree p_t,

$$\mu_0^* = \frac{\sqrt{2}}{3} \approx 0.4714045208,$$

$$\mu_1^* = \frac{1}{3}\left(-3 - \sqrt{2} + \sqrt{11 + 12\sqrt{2}}\right) \approx 0.2915022565,$$

$$\mu_2^* \approx 0.2938105446,$$

$$\mu_3^* \approx 0.2937911168,$$

$$\mu_\infty^* \approx 0.2937911957.$$

To compute the critical discretization parameter μ_∞^*, we used the fact that the (p_t, p_t+1) subdiagonal Padé approximation $R(z)$ converges to the exponential function e^z for $z \leq 0$ as $p_t \to \infty$.

Remark 4.15. Theorem 4.12 shows that for all frequencies $(\theta_x, \theta_t) \in \Theta_{L_x, L_t}$ we have the bound

$$\hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu)), \theta_t) \leq \frac{1}{2}\left(1 + |R(-\beta(\theta_x)\mu)|\right) \leq 1.$$

Only for $\theta_x = 0$ we have that $\beta(\theta_x) = 0$, which implies $R(-\beta(\theta_x)\mu)) = 1$. Hence if the discretization parameter $\mu = \tau_L h_L^{-2}$ is large enough we have that

$$|R(-\beta(\theta_x)\mu)| \approx 0$$

for almost all frequencies $\theta_x \in \Theta_{L_x}$, which implies a good smoothing behavior for almost all frequencies; see Figures 2(a)–2(c). Only the frequencies $\theta_x \in \Theta_{L_x}$ which are close to zero imply $\hat{\mathcal{S}}(\omega_t^*, \alpha(\theta_x, \mu)), \theta_t) \approx 1$. Hence for a large discretization parameter μ the smoother itself is a good iterative solver for most frequencies, though the frequencies $\theta_x \in \Theta_{L_x}$ which are close to zero, i.e., very few low frequencies $\theta_x \in \Theta_{L_x}^{\text{low}}$, do not converge well. To obtain also a perfect solver for a large discretization parameter μ we can simply apply a correction step after one damped block Jacobi iteration by restricting the defect in space several times until we arrive at a very coarse problem. For this small problem one can solve the coarse correction exactly by solving these small problems forward in time. Afterward, we correct the solution by prolongating the coarse corrections back to the fine space-grids.

4.2. Two-grid analysis. The iteration matrices for the two-grid cycles with semi-coarsening and full space-time coarsening are

$$\mathcal{M}^{\mathrm{s}}_{\tau_L,h_L} := \mathcal{S}^{\nu_2}_{\tau_L,h_L}\left[I - \mathcal{P}^{L_x,L_t}_{\mathrm{s}}\left(\mathcal{L}_{2\tau_L,h_L}\right)^{-1}\mathcal{R}^{L_x,L_t}_{\mathrm{s}}\mathcal{L}_{\tau_L,h_L}\right]\mathcal{S}^{\nu_1}_{\tau_L,h_L},$$
$$\mathcal{M}^{\mathrm{f}}_{\tau_L,h_L} := \mathcal{S}^{\nu_2}_{\tau_L,h_L}\left[I - \mathcal{P}^{L_x,L_t}_{\mathrm{f}}\left(\mathcal{L}_{2\tau_L,2h_L}\right)^{-1}\mathcal{R}^{L_x,L_t}_{\mathrm{f}}\mathcal{L}_{\tau_L,h_L}\right]\mathcal{S}^{\nu_1}_{\tau_L,h_L}$$

with the restriction and prolongation matrices

$$\mathcal{R}^{L_x,L_t}_{\mathrm{s}} := I_{N_{L_x}} \otimes \mathcal{R}^{L_t}, \qquad \mathcal{R}^{L_x,L_t}_{\mathrm{f}} := \mathcal{R}^{L_x}_{x} \otimes \mathcal{R}^{L_t},$$
$$\mathcal{P}^{L_x,L_t}_{\mathrm{s}} := I_{N_{L_x}} \otimes \mathcal{P}^{L_t}, \qquad \mathcal{P}^{L_x,L_t}_{\mathrm{f}} := \mathcal{P}^{L_x}_{x} \otimes \mathcal{P}^{L_t}.$$

The restriction and prolongation matrices in time, i.e., $\mathcal{R}^{L_t}$ and $\mathcal{P}^{L_t}$ are given by

$$\mathcal{R}^{L_t} := \begin{pmatrix} R_1 & R_2 & & & & \\ & & R_1 & R_2 & & \\ & & & \ddots & \ddots & \\ & & & & R_1 & R_2 \end{pmatrix} \in \mathbb{R}^{N_t N_{L_t-1} \times N_t N_{L_t}}, \quad \mathcal{P}^{L_t} := (\mathcal{R}^{L_t})^\top \tag{4.5}$$

with the local prolongation matrices $R_1^\top := M_{\tau_L}^{-1}\widetilde{M}^1_{\tau_L}$ and $R_2^\top := M_{\tau_L}^{-1}\widetilde{M}^2_{\tau_L}$, where for basis functions $\{\psi_k\}_{k=1}^{N_t} \subset \mathbb{P}^{p_t}(0,\tau_L)$ and $\{\widetilde{\psi}_k\}_{k=1}^{N_t} \subset \mathbb{P}^{p_t}(0,2\tau_L)$ the local projection matrices from coarse to fine grids are defined for $k,\ell = 1,\dots,N_t$ by

$$\widetilde{M}^1_{\tau_L}[k,\ell] := \int_0^{\tau_L} \widetilde{\psi}_\ell(t)\psi_k(t)\mathrm{d}t \quad \text{and} \quad \widetilde{M}^2_{\tau_L}[k,\ell] := \int_{\tau_L}^{2\tau_L} \widetilde{\psi}_\ell(t)\psi_k(t+\tau)\mathrm{d}t.$$

The restriction and prolongation matrices in space for the one dimensional case are

$$\mathcal{R}^{L_x}_{x} := \frac{1}{2}\begin{pmatrix} 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 \end{pmatrix} \in \mathbb{R}^{N_{L_x-1} \times N_{L_x}}, \tag{4.6}$$

$$\mathcal{P}^{L_x}_{x} := (\mathcal{R}^{L_x}_{x})^\top \in \mathbb{R}^{N_{L_x} \times N_{L_x-1}}. \tag{4.7}$$

To analyze the two-grid iteration matrices $\mathcal{M}^{\mathrm{s}}_{\tau_L,h_L}$ and $\mathcal{M}^{\mathrm{f}}_{\tau_L,h_L}$ we need the following.

LEMMA 4.16. *Let* $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{N_{L_t}})^\top \in \mathbb{R}^{N_t N_{L_x} N_{L_t}}$ *for* $N_t, N_{L_x}, N_{L_t} \in \mathbb{N}$, *where we assume that* N_{L_x} *and* N_{L_t} *are even numbers. Then the vector* $\boldsymbol{u}$ *can be written as*

$$\boldsymbol{u} = \sum_{(\theta_x,\theta_t)\in\Theta^{\mathrm{low,f}}_{L_x,L_t}} \left[\boldsymbol{\psi}(\theta_x,\theta_t) + \boldsymbol{\psi}(\gamma(\theta_x),\theta_t) + \boldsymbol{\psi}(\theta_x,\gamma(\theta_t)) + \boldsymbol{\psi}(\gamma(\theta_x),\gamma(\theta_t))\right]$$

with the shifting operator $\gamma(\theta) := \theta - \mathrm{sign}(\theta)\pi$ *and the vector* $\boldsymbol{\psi}(\theta_x,\theta_t) \in \mathbb{C}^{N_t N_{L_x} N_{L_t}}$ *as in Lemma* 4.2.

Proof. The statement of this lemma follows easily from the fact that the shifting operator $\gamma : \Theta^{\mathrm{low}}_L \to \Theta^{\mathrm{high}}_L$ with $\gamma(\theta) = \theta - \mathrm{sign}(\theta)\pi$ is a one to one mapping. ∎

In view of Lemma 4.16 each vector $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N_{L_t}})^\top$ can be written as a linear combination of the vectors

$$\big\{\boldsymbol{\Phi}(\theta_x, \theta_t), \boldsymbol{\Phi}(\gamma(\theta_x), \theta_t), \boldsymbol{\Phi}(\theta_x, \gamma(\theta_t)), \boldsymbol{\Phi}(\gamma(\theta_x), \gamma(\theta_t))\big\},$$

where $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$ are only low frequencies, i.e., four fine grid modes get aliased to one coarse grid mode. This motivates us to study the convergence in terms of the four fine-grid modes coupled together by the coarse-grid.

DEFINITION 4.17 (space of harmonics). *For $N_t, N_{L_x}, N_{L_t} \in \mathbb{N}$ and the frequencies $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$ let the vector $\boldsymbol{\Phi}(\theta_x, \theta_t) \in \mathbb{C}^{N_t N_{L_x} N_{L_t}}$ be as in Lemma 4.2. Then we define the linear space of harmonics with frequencies (θ_x, θ_t) as*

$$\begin{aligned}
\mathcal{E}_{L_x,L_t}(\theta_x, \theta_t) &:= \operatorname{span}\big\{\boldsymbol{\Phi}(\theta_x, \theta_t), \boldsymbol{\Phi}(\gamma(\theta_x), \theta_t), \boldsymbol{\Phi}(\theta_x, \gamma(\theta_t)), \boldsymbol{\Phi}(\gamma(\theta_x), \gamma(\theta_t))\big\} \\
&= \big\{\boldsymbol{\psi}(\theta_x, \theta_t) \in \mathbb{C}^{N_t N_{L_x} N_{L_t}} : \\
&\qquad \boldsymbol{\psi}_{n,j}(\theta_x, \theta_t) = U_1 \boldsymbol{\Phi}_{n,j}(\theta_x, \theta_t) + U_2 \boldsymbol{\Phi}_{n,j}(\gamma(\theta_x), \theta_t) \\
&\qquad\qquad + U_3 \boldsymbol{\Phi}_{n,j}(\theta_x, \gamma(\theta_t)) + U_4 \boldsymbol{\Phi}_{n,j}(\gamma(\theta_x), \gamma(\theta_t)) \\
&\qquad \text{for } n = 1, \ldots, N_{L_t}, j = 1, \ldots, N_{L_x} \text{ and } U_1, U_2, U_3, U_4 \in \mathbb{C}^{N_t \times N_t}\big\}.
\end{aligned}$$

With the assumption of periodic boundary conditions, see (4.1), Lemma 4.2 implies for the system matrix $\mathcal{L}_{\tau_L,h_L}$ for all frequencies $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$ the mapping property:

$$\begin{aligned}
&\mathcal{L}_{\tau_L,h_L} : \mathcal{E}_{L_x,L_t}(\theta_x, \theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x, \theta_t), \\
&\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \begin{pmatrix} \hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x, \theta_t) U_1 \\ \hat{\mathcal{L}}_{\tau_L,h_L}(\gamma(\theta_x), \theta_t) U_2 \\ \hat{\mathcal{L}}_{\tau_L,h_L}(\theta_x, \gamma(\theta_t)) U_3 \\ \hat{\mathcal{L}}_{\tau_L,h_L}(\gamma(\theta_x), \gamma(\theta_t)) U_4 \end{pmatrix} =: \widetilde{\mathcal{L}}_{\tau_L,h_L}(\theta_x, \theta_t) \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix},
\end{aligned} \tag{4.8}$$

where $\widetilde{\mathcal{L}}_{\tau_L,h_L}(\theta_x, \theta_t) \in \mathbb{C}^{4N_t \times 4N_t}$ is a block diagonal matrix. With the same arguments, we obtain with Lemma 4.3 for the smoother for all frequencies $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$ the mapping property

$$\begin{aligned}
&\mathcal{S}^\nu_{\tau_L,h_L} : \mathcal{E}_{L_x,L_t}(\theta_x, \theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x, \theta_t), \\
&\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \begin{pmatrix} (\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x, \theta_t))^\nu U_1 \\ (\hat{\mathcal{S}}_{\tau_L,h_L}(\gamma(\theta_x), \theta_t))^\nu U_2 \\ (\hat{\mathcal{S}}_{\tau_L,h_L}(\theta_x, \gamma(\theta_t)))^\nu U_3 \\ (\hat{\mathcal{S}}_{\tau_L,h_L}(\gamma(\theta_x), \gamma(\theta_t)))^\nu U_4 \end{pmatrix} =: \left(\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x, \theta_t)\right)^\nu \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}
\end{aligned} \tag{4.9}$$

with the block diagonal matrix $\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x, \theta_t) \in \mathbb{C}^{4N_t \times 4N_t}$. To analyze the two-grid cycle on the space of harmonics $\mathcal{E}_{L_x,L_t}(\theta_x, \theta_t)$ for frequencies $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$, we further have to investigate the mapping properties of the restriction and prolongation operators for the two different coarsening strategies $\mathcal{R}^{L_x,L_t}_{\text{s}}, \mathcal{R}^{L_x,L_t}_{\text{f}}$, and $\mathcal{P}^{L_x,L_t}_{\text{s}}, \mathcal{P}^{L_x,L_t}_{\text{f}}$.

LEMMA 4.18 (Fourier symbols for the prolongation and restriction in space). *Let $\mathcal{R}^{L_x}_x$ and $\mathcal{P}^{L_x}_x$ be the restriction and prolongation matrices as defined in* (4.6)–(4.7). *For $\theta_x \in \Theta^{\text{low}}_{L_x}$ let $\boldsymbol{\varphi}(\theta_x) \in \mathbb{C}^{N_{L_x}}$ and $\boldsymbol{\varphi}^C(2\theta_x) \in \mathbb{C}^{N_{L_x-1}}$ be a fine and a coarse Fourier mode. Then*

$$\big(\mathcal{R}^{L_x}_x \boldsymbol{\varphi}(\theta_x)\big)_j = \hat{\mathcal{R}}_x(\theta_x) \boldsymbol{\varphi}^C_j(2\theta_x)$$

for $j = 2, \dots, N_{L_x-1} - 1$ *with the Fourier symbol* $\hat{\mathcal{R}}_x(\theta_x) := 1 + \cos(\theta_x)$. *For the prolongation operator we further have*

$$\left(\mathcal{P}_x^{L_x}\boldsymbol{\varphi}^C(2\theta_x)\right)_i = \hat{\mathcal{P}}_x(\theta_x)\boldsymbol{\varphi}_i(\theta_x) + \hat{\mathcal{P}}_x(\gamma(\theta_x))\boldsymbol{\varphi}_i(\gamma(\theta_x))$$

for $i = 2, \dots, N_{L_x} - 1$ *with the Fourier symbol* $\hat{\mathcal{P}}_x(\theta_x) := \frac{1}{2}\hat{\mathcal{R}}_x(\theta_x)$.

Proof. The proof is classical; see [41], for example. ∎

LEMMA 4.19 (Fourier symbol for the restriction in time). *Let* $\mathcal{R}^L$ *be the restriction operator defined in* (4.5), *and for the time level* $L_t \in \mathbb{N}_0$ *we consider* $\theta_t \in \Theta_{L_t}^{\text{low}}$ *and let*

$$\boldsymbol{\Phi}(\theta_t) \in \mathbb{C}^{N_t N_{L_t}} \quad \text{and} \quad \boldsymbol{\Phi}^C(\theta_t) \in \mathbb{C}^{N_t N_{L_t-1}}$$

be fine and coarse block Fourier modes with respect to time, i.e.,

$$\begin{aligned}
\boldsymbol{\Phi}_{n,\ell}(\theta_t) &:= \boldsymbol{\varphi}_n(\theta_t) \qquad \text{for } \ell = 1, \dots, N_t \text{ and } n = 1, \dots, N_{L_t},\\
\boldsymbol{\Phi}^C_{\hat{n},\ell}(\theta_t) &:= \boldsymbol{\varphi}_{\hat{n}}(\theta_t) \qquad \text{for } \ell = 1, \dots, N_t \text{ and } \hat{n} = 1, \dots, N_{L_t-1}.
\end{aligned}$$

Then we have

$$\left(\mathcal{R}^{L_t}\boldsymbol{\Phi}(\theta_t)\right)_{\hat{n}} = \hat{\mathcal{R}}(\theta_t)\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) \qquad \text{for } \hat{n} = 1, \dots, N_{L_t-1}$$

with the Fourier symbol $\hat{\mathcal{R}}(\theta_k) := e^{-\mathbb{i}\theta_k}R_1 + R_2$.

Proof. By definition we have for $\hat{n} = 1, \dots, N_{L_t-1}$

$$\begin{aligned}
\left(\mathcal{R}^{L_t}\boldsymbol{\Phi}(\theta_t)\right)_{\hat{n}} &= R_1\boldsymbol{\Phi}_{2\hat{n}-1}(\theta_t) + R_2\boldsymbol{\Phi}_{2\hat{n}}(\theta_t) = \left[e^{-\mathbb{i}\theta_\ell}R_1 + R_2\right]\boldsymbol{\Phi}_{2\hat{n}}(\theta_t)\\
&= \left[e^{-\mathbb{i}\theta_\ell}R_1 + R_2\right]\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t),
\end{aligned}$$

by using a shifting argument as in Lemma 4.4 and the fact that $\boldsymbol{\varphi}_{2\hat{n}}(\theta_t) = \boldsymbol{\varphi}_{\hat{n}}(2\theta_t)$. ∎

LEMMA 4.20 (Fourier symbol for the prolongation in time). *Let* $\mathcal{P}^L$ *be the prolongation operator defined in* (4.5), *and for the time level* $L_t \in \mathbb{N}_0$ *we consider* $\theta_t \in \Theta_{L_t}^{\text{low}}$ *and let*

$$\boldsymbol{\Phi}(\theta_t) \in \mathbb{C}^{N_t N_{L_t}} \quad \text{and} \quad \boldsymbol{\Phi}^C(\theta_t) \in \mathbb{C}^{N_t N_{L_t-1}}$$

be fine and coarse block Fourier modes with respect to time as defined in Lemma 4.5. *Then we have*

$$\left(\mathcal{P}^L\boldsymbol{\Phi}^C(2\theta_t)\right)_n = \hat{\mathcal{P}}(\theta_t)\boldsymbol{\Phi}_n(\theta_t) + \hat{\mathcal{P}}(\gamma(\theta_t))\boldsymbol{\Phi}_n(\gamma(\theta_t))$$

for $n = 1, \dots, N_{L_t}$ *with the Fourier symbol* $\hat{\mathcal{P}}(\theta_k) := \frac{1}{2}[e^{\mathbb{i}\theta_t}R_1^\top + R_2^\top]$.

Proof. By definition we have for $\hat{n} = 1, \dots, N_{L_t-1}$

$$\left(\mathcal{P}^{L_t}\boldsymbol{\Phi}^C(2\theta_t)\right)_{2\hat{n}-1} = R_1^\top\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) = R_1^\top\boldsymbol{\Phi}_{2\hat{n}}(\theta_t) = e^{\mathbb{i}\theta_t}R_1^\top\boldsymbol{\Phi}_{2\hat{n}-1}(\theta_t),$$

by using a shifting argument as in Lemma 4.4 and the fact that $\boldsymbol{\varphi}_{\hat{n}}(2\theta_t) = \boldsymbol{\varphi}_{2\hat{n}}(\theta_t)$. Similar computations also give

$$\left(\mathcal{P}^L\boldsymbol{\Phi}^C(2\theta_t)\right)_{2\hat{n}} = R_2^\top\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) = R_2^\top\boldsymbol{\Phi}_{2\hat{n}}(\theta_t).$$

Hence we have for $n = 1, \dots, N_{L_t}$

$$\left(\mathcal{P}^L\boldsymbol{\Phi}^C(2\theta_t)\right)_n = \begin{cases} e^{\mathbb{i}\theta_t}R_1^\top\boldsymbol{\Phi}_n(\theta_t) & n \text{ odd},\\ R_2^\top\boldsymbol{\Phi}_n(\theta_t) & n \text{ even} \end{cases} \in \mathbb{C}^{N_t}. \tag{4.10}$$

Further computations show for $\ell = 1, \dots, N_t$ that

$$\begin{aligned}\boldsymbol{\Phi}_{n,\ell}(\gamma(\theta_t)) = \boldsymbol{\varphi}_n(\gamma(\theta_t)) = e^{\mathbb{i} n\gamma(\theta_t)} = e^{\mathbb{i} n\theta_t - \mathbb{i}\,\mathrm{sign}(\theta_t) n\pi} = \boldsymbol{\varphi}_n(\theta_t) e^{\mathbb{i}\,\mathrm{sign}(\theta_t) n\pi} \\ = \begin{cases} -\boldsymbol{\Phi}_{n,\ell}(\theta_t) & n \text{ odd}, \\ \boldsymbol{\Phi}_{n,\ell}(\theta_t) & n \text{ even}. \end{cases}\end{aligned}$$

Moreover we have

$$\begin{aligned}\hat{\mathcal{P}}(\gamma(\theta_t)) &= \frac{1}{2}\left[e^{\mathbb{i}\gamma(\theta_t)} R_1^\top + R_2^\top\right] = \frac{1}{2}\left[e^{\mathbb{i}(\theta_t - \mathrm{sign}(\theta_t)\pi)} R_1^\top + R_2^\top\right] \\ &= \frac{1}{2}\left[-e^{\mathbb{i}\theta_t} R_1^\top + R_2^\top\right].\end{aligned}$$

Hence we obtain for $n = 1, \dots, N_{L_t}$

$$\begin{aligned}&\hat{\mathcal{P}}(\theta_t)\boldsymbol{\Phi}_n(\theta_t) + \hat{\mathcal{P}}(\gamma(\theta_t))\boldsymbol{\Phi}_n(\gamma(\theta_t)) \\ &\quad = \frac{1}{2}\left[e^{\mathbb{i}\theta_t} R_1^\top + R_2^\top\right]\boldsymbol{\Phi}_n(\theta_t) + \frac{1}{2}\left[-e^{\mathbb{i}\theta_t} R_1^\top + R_2^\top\right]\begin{cases} -\boldsymbol{\Phi}_{n,\ell}(\theta_t) & n \text{ odd}, \\ \boldsymbol{\Phi}_{n,\ell}(\theta_t) & n \text{ even} \end{cases} \\ &\quad = \begin{cases} e^{\mathbb{i}\theta_t} R_1^\top \boldsymbol{\Phi}_n(\theta_t) & n \text{ odd}, \\ R_2^\top \boldsymbol{\Phi}_n(\theta_t) & n \text{ even} \end{cases} = \left(\mathcal{P}^L \boldsymbol{\Phi}^C(2\theta_t)\right)_n,\end{aligned}$$

where we used (4.10). ∎

DEFINITION 4.21 (Fourier space for semi-coarsening). *For $N_t, N_{L_x}, N_{L_t-1} \in \mathbb{N}$ and the frequencies $(\theta_x, \theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t}$ let the vector $\boldsymbol{\Phi}(\theta_x, \theta_t) \in \mathbb{C}^{N_t N_{L_x} N_{L_t-1}}$ be defined as in Lemma* 4.2. *Then we define the linear space with frequencies $(\theta_x, 2\theta_t)$ as*

$$\begin{aligned}\Psi_{L_x,L_t-1}(\theta_x, 2\theta_t) &:= \mathrm{span}\left\{\boldsymbol{\Phi}(\theta_x, 2\theta_t), \boldsymbol{\Phi}(\gamma(\theta_x), 2\theta_t)\right\} \\ &= \big\{\boldsymbol{\psi}(\theta_x, 2\theta_t) \in \mathbb{C}^{N_t N_{L_x} N_{L_t-1}} : \\ &\qquad \boldsymbol{\psi}_{n,j}(\theta_x, 2\theta_t) = U_1 \boldsymbol{\Phi}_{n,j}(\theta_x, 2\theta_t) + U_2 \boldsymbol{\Phi}_{n,j}(\gamma(\theta_x), 2\theta_t) \\ &\qquad \text{for } n = 1, \dots, N_{L_t} - 1, j = 1, \dots, N_{L_x} \text{ and } U_1, U_2 \in \mathbb{C}^{N_t \times N_t}\big\}.\end{aligned}$$

For semi-coarsening, the next lemma shows the mapping property for the restriction operator $\mathcal{R}_{\mathrm{s}}^{L_x,L_t}$.

LEMMA 4.22 (Fourier symbol for the restriction for semi-coarsening). *The restriction operator $\mathcal{R}_{\mathrm{s}}^{L_x,L_t}$ satisfies the mapping property*

$$\mathcal{R}_{\mathrm{s}}^{L_x,L_t} : \mathcal{E}_{L_x,L_t}(\theta_x, \theta_t) \to \Psi_{L_x,L_t-1}(\theta_x, 2\theta_t)$$

with the mapping

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \widetilde{\mathcal{R}}_{\mathrm{s}}(\theta_t) \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}$$

and the matrix

$$\widetilde{\mathcal{R}}_{\mathrm{s}}(\theta_t) := \begin{pmatrix} \hat{\mathcal{R}}(\theta_t) & 0 & \hat{\mathcal{R}}(\gamma(\theta_t)) & 0 \\ 0 & \hat{\mathcal{R}}(\theta_t) & 0 & \hat{\mathcal{R}}(\gamma(\theta_t)) \end{pmatrix}$$

with the Fourier symbol $\hat{\mathcal{R}}(\theta_t) \in \mathbb{C}^{N_t \times N_t}$ as defined in Lemma 4.19.

Proof. Let $\boldsymbol{\Phi}(\theta_x,\theta_t) \in \Psi_{L_x,L_t}(\theta_x,\theta_t)$ and $\boldsymbol{\Phi}^C(\theta_x,2\theta_t) \in \Psi_{L_x,L_t-1}(\theta_x,2\theta_t)$ be defined as in Lemma 4.21. Then for $\hat{n} = 1,\dots,N_{L_t-1}$ and $j = 1,\dots,N_{L_x}$ we have, using the notation and the results of Lemma 4.19,

$$\begin{aligned}\left(\mathcal{R}_{\mathrm{s}}^{L_x,L_t}\boldsymbol{\Phi}(\theta_x,\theta_t)\right)_{\hat{n},j} &= \boldsymbol{\varphi}_j(\theta_x)\left(\mathcal{R}^{L_t}\boldsymbol{\Phi}(\theta_t)\right)_{\hat{n}} = \hat{\mathcal{R}}(\theta_t)\boldsymbol{\Phi}_{\hat{n}}^C(2\theta_t)\boldsymbol{\varphi}_j(\theta_x)\\ &= \hat{\mathcal{R}}(\theta_t)\boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\theta_t).\end{aligned}$$

Applying this result to the vector $\boldsymbol{\psi}(\theta_x,\theta_t) \in \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$ with $(\theta_x,\theta_t) \in \Theta_{L_x,L_t}^{\mathrm{f}}$ results in

$$\begin{aligned}\left(\mathcal{R}_{\mathrm{s}}^{L_x,L_t}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{\hat{n},j} &= \hat{\mathcal{R}}(\theta_t)U_1\boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\theta_t) + \hat{\mathcal{R}}(\gamma(\theta_t))U_3\boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\gamma(\theta_t))\\ &\quad + \hat{\mathcal{R}}(\theta_t)U_2\boldsymbol{\Phi}_{\hat{n},j}^C(\gamma(\theta_x),2\theta_t) + \hat{\mathcal{R}}(\gamma(\theta_t))U_4\boldsymbol{\Phi}_{\hat{n},j}^C(\gamma(\theta_x),2\gamma(\theta_t)).\end{aligned}$$

Since $\boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\gamma(\theta_t)) = \boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\theta_t)$ we further obtain

$$\begin{aligned}&= \left[\hat{\mathcal{R}}(\theta_t)U_1 + \hat{\mathcal{R}}(\gamma(\theta_t))U_3\right]\boldsymbol{\Phi}_{\hat{n},j}^C(\theta_x,2\theta_t)\\ &\quad + \left[\hat{\mathcal{R}}(\theta_t)U_2 + \hat{\mathcal{R}}(\gamma(\theta_t))U_4\right]\boldsymbol{\Phi}_{\hat{n},j}^C(\gamma(\theta_x),2\theta_t),\end{aligned}$$

which completes the proof. ∎

LEMMA 4.23 (Fourier symbol for the restriction for full-coarsening). *With the assumptions of periodic boundary conditions* (4.1) *the following mapping property for the restriction operator holds:*

$$\mathcal{R}_{\mathrm{f}}^{L_x,L_t} : \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t) \to \Psi_{L_x-1,L_t-1}(2\theta_x,2\theta_t)$$

with the mapping

$$\begin{pmatrix}U_1\\U_2\\U_3\\U_4\end{pmatrix} \mapsto \widetilde{\mathcal{R}}_{\mathrm{f}}(\theta_x,\theta_t)\begin{pmatrix}U_1\\U_2\\U_3\\U_4\end{pmatrix}$$

and the matrix

$$\widetilde{\mathcal{R}}_{\mathrm{f}}(\theta_x,\theta_t) := \begin{pmatrix}\hat{\mathcal{R}}(\theta_x,\theta_t) & \hat{\mathcal{R}}(\gamma(\theta_x),\theta_t) & \hat{\mathcal{R}}(\theta_x,\gamma(\theta_t)) & \hat{\mathcal{R}}(\gamma(\theta_x),\gamma(\theta_t))\end{pmatrix}$$

with the Fourier symbol

$$\hat{\mathcal{R}}(\theta_x,\theta_t) := \hat{\mathcal{R}}_x(\theta_x)\hat{\mathcal{R}}(\theta_t) \in \mathbb{C}^{N_t\times N_t},$$

where $\hat{\mathcal{R}}_x(\theta_x) \in \mathbb{C}$ *is defined as in Lemma* 4.18.

Proof. For the frequencies $(\theta_x,\theta_t) \in \Theta_{L_x,L_t}^{\mathrm{low}}$ let $\boldsymbol{\Phi}(\theta_x,\theta_t) \in \Psi_{L_x,L_t}(\theta_x,\theta_t)$ and $\boldsymbol{\Phi}^C(2\theta_x,2\theta_t) \in \Psi_{L_x-1,L_t-1}(2\theta_x,2\theta_t)$. By applying Lemmas 4.18 and 4.19 we obtain for $\hat{n} = 1,\dots,N_{L_t-1}$ and $\hat{j} = 2,\dots,N_{L_x-1}-1$

$$\begin{aligned}\left(\mathcal{R}_{\mathrm{f}}^{L_x,L_t}\boldsymbol{\Phi}(\theta_x,\theta_t)\right)_{\hat{n},\hat{j}} &= \left(\mathcal{R}_x^{L_x}\boldsymbol{\varphi}(\theta_x)\right)_{\hat{j}}\left(\mathcal{R}^{L_t}\boldsymbol{\Phi}(\theta_t)\right)_{\hat{n}} = \hat{\mathcal{R}}_x(\theta_x)\hat{\mathcal{R}}(\theta_t)\boldsymbol{\Phi}_{\hat{n}}^C(2\theta_t)\boldsymbol{\varphi}_{\hat{j}}^C(2\theta_x)\\ &= \hat{\mathcal{R}}(\theta_x,\theta_t)\boldsymbol{\Phi}_{\hat{n},\hat{j}}^C(2\theta_x,2\theta_t).\end{aligned}$$

Using this result for the vector $\boldsymbol{\psi}(\theta_x,\theta_t) \in \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$ with $(\theta_x,\theta_t) \in \Theta^{\mathrm{f}}_{L_x,L_t}$ results in

$$\begin{aligned}
&\left(\mathcal{R}_{\mathrm{f}}^{L_x,L_t}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{\hat{n},\hat{j}} \\
&\quad= \hat{\mathcal{R}}(\theta_x,\theta_t)U_1\boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\theta_x,2\theta_t) + \hat{\mathcal{R}}(\gamma(\theta_x),\theta_t)U_2\boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\gamma(\theta_x),2\theta_t) \\
&\qquad+ \hat{\mathcal{R}}(\theta_x,\gamma(\theta_t))U_3\boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\theta_x,2\gamma(\theta_t)) + \hat{\mathcal{R}}(\gamma(\theta_x),\gamma(\theta_t))U_4\boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\gamma(\theta_x),2\gamma(\theta_t)).
\end{aligned}$$

With the relations

$$\boldsymbol{\Phi}^C(2\theta_x,2\theta_t) = \boldsymbol{\Phi}^C(2\gamma(\theta_x),2\theta_t) = \boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\theta_x,2\gamma(\theta_t)) = \boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\gamma(\theta_x),2\gamma(\theta_t)),$$

we obtain the statement of this lemma with

$$\begin{aligned}
\left(\mathcal{R}_{\mathrm{f}}^{L_x,L_t}\boldsymbol{\psi}(\theta_x,\theta_t)\right)_{\hat{n},\hat{j}} = \big[&\hat{\mathcal{R}}(\theta_x,\theta_t)U_1 + \hat{\mathcal{R}}(\gamma(\theta_x),\theta_t)U_2 + \hat{\mathcal{R}}(\theta_x,\gamma(\theta_t))U_3 \\
&+ \hat{\mathcal{R}}(\gamma(\theta_x),\gamma(\theta_t))U_4\big]\boldsymbol{\Phi}^C_{\hat{n},\hat{j}}(2\theta_x,2\theta_t). \qquad \square
\end{aligned}$$

With the next lemmas, we will analyze the mapping properties of the prolongation operators $\mathcal{P}_{\mathrm{s}}^{L_x,L_t}$ and $\mathcal{P}_{\mathrm{f}}^{L_x,L_t}$.

LEMMA 4.24 (Fourier symbol for the prolongation for semi-coarsening). *For $(\theta_x,\theta_t) \in \Theta^{\mathrm{f}}_{L_x,L_t}$ the prolongation operator $\mathcal{P}_{\mathrm{s}}^{L_x,L_t}$ satisfies the mapping property*

$$\mathcal{P}_{\mathrm{s}}^{L_x,L_t} : \Psi_{L_x,L_t-1}(\theta_x,2\theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$$

with the mapping

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \mapsto \begin{pmatrix} \hat{\mathcal{P}}(\theta_t) & 0 \\ 0 & \hat{\mathcal{P}}(\theta_t) \\ \hat{\mathcal{P}}(\gamma(\theta_t)) & 0 \\ 0 & \hat{\mathcal{P}}(\gamma(\theta_t)) \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} =: \widetilde{\mathcal{P}}_{\mathrm{s}}(\theta_t) \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

and the Fourier symbol $\hat{\mathcal{P}}(\theta_t) \in \mathbb{C}^{N_t \times N_t}$ defined as in Lemma 4.20.

Proof. Let $\boldsymbol{\psi}^C(\theta_x,2\theta_t) \in \boldsymbol{\Psi}_{L_x,L_t-1}(\theta_x,2\theta_t)$ for $(\theta_x,\theta_t) \in \Theta^{\mathrm{f}}_{L_x,L_t}$, i.e.,

$$\begin{aligned}
\boldsymbol{\psi}^C_{\hat{n},j}(\theta_x,2\theta_t) &= \boldsymbol{\varphi}_j(\theta_x)U_1\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) + \boldsymbol{\varphi}_j(\gamma(\theta_x))U_2\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) \\
&=: \boldsymbol{\varphi}_j(\theta_x)\boldsymbol{\psi}^{C,1}_{\hat{n}}(2\theta_t) + \boldsymbol{\varphi}_j(\gamma(\theta_x))\boldsymbol{\psi}^{C,2}_{\hat{n}}(2\theta_t).
\end{aligned}$$

Let $\boldsymbol{\Phi}(\theta_x,\theta_t) \in \mathbb{C}^{N_tN_{L_x}N_{L_t}}$ be defined as in Lemma 4.2, and then we obtain with Lemma 4.20 for $n = 1,\dots,N_{L_t}$ and $j = 1,\dots,N_{L_x}$ the statement of this lemma with

$$\begin{aligned}
\left(\mathcal{P}_{\mathrm{s}}^{L_x,L_t}\boldsymbol{\psi}^C(\theta_x,2\theta_t)\right)_{n,j} &= \boldsymbol{\varphi}_j(\theta_x)\left(\mathcal{P}^{L_t}\boldsymbol{\psi}^{C,1}(2\theta_t)\right)_n + \boldsymbol{\varphi}_j(\gamma(\theta_x))\left(\mathcal{P}^{L_t}\boldsymbol{\psi}^{C,2}(2\theta_t)\right)_n \\
&= \hat{\mathcal{P}}(\theta_t)U_1\boldsymbol{\Phi}_{n,j}(\theta_x,\theta_t) + \hat{\mathcal{P}}(\gamma(\theta_t))U_1\boldsymbol{\Phi}_{n,j}(\theta_x,\gamma(\theta_t)) \\
&\quad+ \hat{\mathcal{P}}(\theta_t)U_2\boldsymbol{\Phi}_{n,j}(\gamma(\theta_x),\theta_t) + \hat{\mathcal{P}}(\gamma(\theta_t))U_2\boldsymbol{\Phi}_{n,j}(\gamma(\theta_x),\gamma(\theta_t)). \qquad \square
\end{aligned}$$

LEMMA 4.25 (Fourier symbol for the prolongation for full-coarsening). *With the assumptions of periodic boundary conditions* (4.1) *the following mapping property for the prolongation operator holds:*

$$\mathcal{P}_{\mathrm{f}}^{L_x,L_t} : \Psi_{L_x-1,L_t-1}(2\theta_x,2\theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$$

with the mapping

$$U \mapsto \begin{pmatrix} \hat{\mathcal{P}}(\theta_x,\theta_t) \\ \hat{\mathcal{P}}(\gamma(\theta_x),\theta_t) \\ \hat{\mathcal{P}}(\theta_x,\gamma(\theta_t)) \\ \hat{\mathcal{P}}(\gamma(\theta_x),\gamma(\theta_t)) \end{pmatrix} U =: \widetilde{\mathcal{P}}_{\mathrm{f}}(\theta_x,\theta_t)U$$

and the Fourier symbol

$$\hat{\mathcal{P}}(\theta_x,\theta_t) := \hat{\mathcal{P}}_x(\theta_x)\hat{\mathcal{P}}(\theta_t),$$

where $\hat{\mathcal{P}}(\theta_t)$ *is defined as in Lemma* 4.20.

Proof. Let $\boldsymbol{\psi}^C(2\theta_x,2\theta_t) \in \boldsymbol{\Psi}_{L_x-1,L_t-1}(2\theta_x,2\theta_t)$ for $(\theta_x,\theta_t) \in \Theta^{\mathrm{f}}_{L_x,L_t}$, i.e.,

$$\boldsymbol{\psi}^C_{\hat{n},\hat{j}}(2\theta_x,2\theta_t) = \boldsymbol{\varphi}^C_{\hat{j}}(2\theta_x)U\boldsymbol{\Phi}^C_{\hat{n}}(2\theta_t) =: \boldsymbol{\varphi}^C_{\hat{j}}(2\theta_x)\boldsymbol{\psi}^C_{\hat{n}}(2\theta_t).$$

Let $\boldsymbol{\Phi}(\theta_x,\theta_t) \in \mathbb{C}^{N_tN_{L_x}N_{L_t}}$ be defined as in Lemma 4.2, and then we obtain with Lemmas 4.18 and 4.20 for $n = 1,\dots,N_{L_t}$ and $j = 2,\dots,N_{L_x}-1$ the statement of this lemma with

$$\begin{aligned}
&\left(\mathcal{P}^{L_x,L_t}_{\mathrm{f}}\boldsymbol{\psi}^C(2\theta_x,2\theta_t)\right)_{n,j} \\
&\quad = \left(\mathcal{P}^{L_x}_x\boldsymbol{\varphi}^C(2\theta_x)\right)_j \left(\mathcal{P}^{L_t}\boldsymbol{\psi}^C(2\theta_t)\right)_n \\
&\quad = \hat{\mathcal{P}}(\theta_x,\theta_t)U\boldsymbol{\Phi}_{n,j}(\theta_x,\theta_t) + \hat{\mathcal{P}}(\gamma(\theta_x),\theta_t)U\boldsymbol{\Phi}_{n,j}(\gamma(\theta_x),\theta_t) \\
&\qquad + \hat{\mathcal{P}}(\theta_x,\gamma(\theta_t))U\boldsymbol{\Phi}_{n,j}(\theta_x,\gamma(\theta_t)) + \hat{\mathcal{P}}(\gamma(\theta_x),\gamma(\theta_t))U\boldsymbol{\Phi}_{n,j}(\gamma(\theta_x),\gamma(\theta_t)).
\end{aligned}$$

∎

For periodic boundary conditions (4.1), we further obtain with Lemma 4.5 the mapping property for the coarse grid correction, when semi-coarsening in time is applied,

$$\begin{aligned}
&(\mathcal{L}_{2\tau_L,h_L})^{-1} : \Psi_{L_x,L_t-1}(\theta_x,2\theta_t) \to \Psi_{L_x,L_t-1}(\theta_x,2\theta_t) \\
&\qquad\qquad \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \mapsto \left(\widetilde{\mathcal{L}}^{\mathrm{s}}_{2\tau_L,h_L}(\theta_x,2\theta_t)\right)^{-1} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \in \mathbb{C}^{2N_t\times N_t}
\end{aligned} \tag{4.11}$$

with the matrix

$$\left(\widetilde{\mathcal{L}}^{\mathrm{s}}_{2\tau_L,h_L}(\theta_x,2\theta_t)\right)^{-1} := \begin{pmatrix} \left(\hat{\mathcal{L}}_{2\tau_L,h_L}(\theta_x,2\theta_t)\right)^{-1} & 0 \\ 0 & \left(\hat{\mathcal{L}}_{2\tau_L,h_L}(\gamma(\theta_x),2\theta_t)\right)^{-1} \end{pmatrix},$$

i.e., $(\widetilde{\mathcal{L}}^{\mathrm{s}}_{2\tau_L,h_L}(\theta_x,2\theta_t))^{-1} \in \mathbb{C}^{2N_t\times 2N_t}$. For full space-time coarsening, we have the mapping property

$$\begin{aligned}
&(\mathcal{L}_{2\tau_L,2h_L})^{-1} : \Psi_{L_x-1,L_t-1}(2\theta_x,2\theta_t) \to \Psi_{L_x-1,L_t-1}(2\theta_x,2\theta_t) \\
&\qquad\qquad U \mapsto \left(\widetilde{\mathcal{L}}^{\mathrm{f}}_{2\tau_L,2h_L}(2\theta_x,2\theta_t)\right)^{-1} U \in \mathbb{C}^{N_t\times N_t}
\end{aligned} \tag{4.12}$$

with $(\widetilde{\mathcal{L}}^{\mathrm{f}}_{2\tau_L,2h_L}(2\theta_x,2\theta_t))^{-1} := (\hat{\mathcal{L}}_{2\tau_L,2h_L}(2\theta_x,2\theta_t))^{-1} \in \mathbb{C}^{N_t\times N_t}$.

We can now prove the following two theorems.

THEOREM 4.26 (Fourier symbol for the two-grid operator for semi-coarsening). *Let* $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$. *With the assumption of periodic boundary conditions* (4.1), *the following mapping property holds for the two-grid operator* $\mathcal{M}^{\text{s}}_{\tau_L,h_L}$ *with semi-coarsening in time:*

$$\mathcal{M}^{\text{s}}_{\tau_L,h_L} : \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$$

with the mapping

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \widetilde{\mathcal{M}}^{\text{s}}_{\mu}(\theta_k,\theta_t) \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}$$

and the iteration matrix

$$\widetilde{\mathcal{M}}^{\text{s}}_{\mu}(\theta_k,\theta_t) := \left(\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right)^{\nu_2} \widetilde{\mathcal{K}}_{\text{s}}(\theta_x,\theta_t) \left(\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right)^{\nu_1} \in \mathbb{C}^{4N_t \times 4N_t}$$

with

$$\widetilde{\mathcal{K}}_{\text{s}}(\theta_x,\theta_t) := I_{4N_t} - \widetilde{\mathcal{P}}_{\text{s}}(\theta_t) \left(\widetilde{\mathcal{L}}^{\text{s}}_{2\tau_L,h_L}(\theta_x,2\theta_t)\right)^{-1} \widetilde{\mathcal{R}}_{\text{s}}(\theta_t) \widetilde{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t).$$

Proof. The statement of this theorem follows by using Lemmas 4.22 and 4.24, the mapping properties (4.8), (4.9), and (4.11), and the fact that

$$\beta : \Theta^{\text{low}}_{L_t} \to \Theta_{L_t-1} \qquad \text{with} \qquad \theta_t \mapsto 2\theta_t$$

is a one to one mapping. ∎

THEOREM 4.27 (Fourier symbol for the two-grid operator for full-coarsening). *Let* $(\theta_x, \theta_t) \in \Theta^{\text{low,f}}_{L_x,L_t}$. *With the assumption of periodic boundary conditions* (4.1), *the following mapping property holds for the two-grid operator* $\mathcal{M}^{\text{f}}_{\tau_L,h_L}$ *with full space-time coarsening:*

$$\mathcal{M}^{\text{f}}_{\tau_L,h_L} : \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$$

with the mapping

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \widetilde{\mathcal{M}}^{\text{f}}_{\mu}(\theta_k,\theta_t) \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}$$

and the iteration matrix

$$\widetilde{\mathcal{M}}^{\text{f}}_{\mu}(\theta_k,\theta_t) := \left(\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right)^{\nu_2} \widetilde{\mathcal{K}}_{\text{f}}(\theta_x,\theta_t) \left(\widetilde{\mathcal{S}}_{\tau_L,h_L}(\theta_x,\theta_t)\right)^{\nu_1} \in \mathbb{C}^{4N_t \times 4N_t}$$

with

$$\widetilde{\mathcal{K}}_{\text{f}}(\theta_x,\theta_t) := I_{4N_t} - \widetilde{\mathcal{P}}_{\text{f}}(\theta_x,\theta_t) \left(\widetilde{\mathcal{L}}^{\text{f}}_{2\tau_L,2h_L}(2\theta_x,2\theta_t)\right)^{-1} \widetilde{\mathcal{R}}_{\text{f}}(\theta_x,\theta_t) \widetilde{\mathcal{L}}_{\tau_L,h_L}(\theta_x,\theta_t).$$

Proof. The statement follows in the same way as in Theorem 4.27 by using Lemmas 4.23 and 4.25 and the mapping properties (4.8), (4.9), and (4.12). ∎

In view of Lemma 4.16 we can now represent the initial error $\boldsymbol{e}^0 = \boldsymbol{u} - \boldsymbol{u}^0$ as

$$\begin{aligned}\boldsymbol{e}^0 &= \sum_{(\theta_x,\theta_t)\in\Theta^{\mathrm{low,f}}_{L_x,L_t}} \left[\boldsymbol{\psi}(\theta_x,\theta_t) + \boldsymbol{\psi}(\gamma(\theta_x),\theta_t) + \boldsymbol{\psi}(\theta_x,\gamma(\theta_t)) + \boldsymbol{\psi}(\gamma(\theta_x),\gamma(\theta_t))\right] \\ &=: \sum_{(\theta_x,\theta_t)\in\Theta^{\mathrm{low,f}}_{L_x,L_t}} \tilde{\boldsymbol{\psi}}(\theta_x,\theta_t)\end{aligned}$$

with $\tilde{\boldsymbol{\psi}}(\theta_x,\theta_t) \in \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$ for all $(\theta_x,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t}$. Using Theorems 4.26 and 4.27 we now can analyze the asymptotic convergence behavior of the two-grid cycle by simply computing the largest spectral radius of $\widetilde{\mathcal{M}}^{\mathrm{s}}_{\mu}(\theta_k,\theta_t)$ or $\widetilde{\mathcal{M}}^{\mathrm{f}}_{\mu}(\theta_k,\theta_t)$ with respect to the frequencies $(\theta_x,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t}$. This motivates the following.

DEFINITION 4.28 (asymptotic two-grid convergence factors). *For the two-grid iteration matrices $\mathcal{M}^{\mathrm{s}}_{\tau_L,h_L}$ and $\mathcal{M}^{\mathrm{f}}_{\tau_L,h_L}$ we define the asymptotic convergence factors*

$$\begin{aligned}\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu}) &:= \max\left\{\varrho(\widetilde{\mathcal{M}}^{\mathrm{s}}_{\mu}(\theta_k,\theta_t)) : (\theta_x,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t} \text{ with } \theta_x \neq 0\right\}, \\ \varrho(\hat{\mathcal{M}}^{\mathrm{f}}_{\mu}) &:= \max\left\{\varrho(\widetilde{\mathcal{M}}^{\mathrm{f}}_{\mu}(\theta_k,\theta_t)) : (\theta_x,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t} \text{ with } \theta_x \neq 0\right\}.\end{aligned}$$

Note that in the definition of the two-grid convergence factors we have neglected all frequencies $(0,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t}$, since the Fourier symbol with respect to the Laplacian is zero for $\theta_x = 0$; see also the remarks in [41, Chapter 4].

To derive the asymptotic convergence factors $\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu})$ and $\varrho(\hat{\mathcal{M}}^{\mathrm{f}}_{\mu})$ for a given discretization parameter $\mu \in \mathbb{R}_+$ and a given polynomial degree $p_t \in \mathbb{N}_0$ we have to compute the eigenvalues of

$$\widetilde{\mathcal{M}}^{\mathrm{s}}_{\mu}(\theta_k,\theta_t) \in \mathbb{C}^{4N_t\times 4N_t} \quad \text{and} \quad \widetilde{\mathcal{M}}^{\mathrm{f}}_{\mu}(\theta_k,\theta_t) \in \mathbb{C}^{4N_t\times 4N_t} \tag{4.13}$$

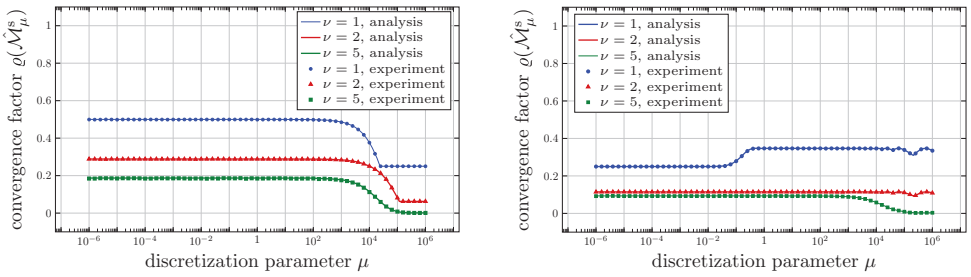
with $N_t = p_t + 1$ for each low frequency $(\theta_x,\theta_t) \in \Theta^{\mathrm{low,f}}_{L_x,L_t}$. Since it is difficult to find closed form expressions for the eigenvalues of the iteration matrices (4.13), we will compute the eigenvalues numerically. In particular we will compute the average convergence factors for the domain $\Omega = (0,1)$ with a decomposition into 1024 uniform sub intervals, i.e., $N_{L_x} = 1023$. Furthermore we will analyze the two-grid cycles for $N_{L_t} = 256$ time steps. In numerical experiments it turned out that finer space-time grids lead to almost the same two-grid convergence factors, i.e., these estimates imply a scalable two-grid method.

We plot as solid lines in the Figures 3(a)–3(b) the theoretical convergence factors $\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu})$ as functions of the discretization parameter $\mu = \tau_L h_L^{-2} \in [10^{-6}, 10^{6}]$ for different polynomial degrees $p_t \in \{0,1\}$ and different number of smoothing steps $\nu_1 = \nu_2 = \nu \in \{1,2,5\}$. We observe that the theoretical convergence factors are always bounded by $\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu}) \leq \frac{1}{2}$. For the case when semi-coarsening in time is applied we see that the two-grid cycle converges for any discretization parameter μ. We also see for polynomial degree $p_t = 1$ that the theoretical convergence factors are much smaller than the theoretical convergence factors for the lowest order case $p_t = 0$.

We also plot in Figures 3(a)–3(b) using dots, triangles, and squares the numerically measured convergence factors for solving the equation

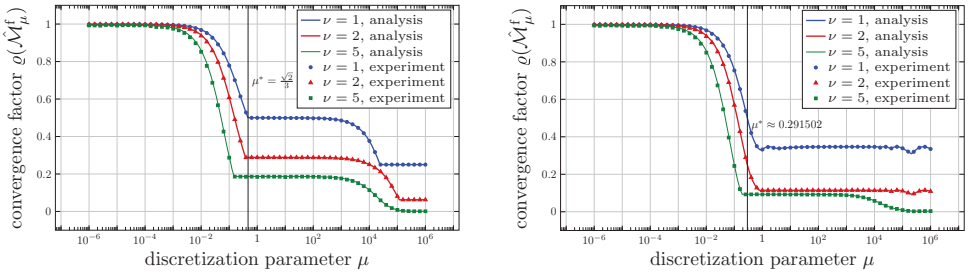
$$\mathcal{L}_{\tau_L,h_L}\boldsymbol{u} = \boldsymbol{f}$$

with the two-grid cycle when semi-coarsening in time is applied. For the numerical



(a) $p_t = 0$. (b) $p_t = 1$.

FIG. 3. *Asymptotic convergence factor* $\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu})$ *for different discretization parameters* μ *and numerical convergence factors for* $N_t = 256$ *time steps and* $N_x = 1023$.



(a) $p_t = 0$. (b) $p_t = 1$.

FIG. 4. *Average convergence factor* $\varrho(\hat{\mathcal{M}}^{\mathrm{f}}_{\mu})$ *for different discretization parameters* μ *and numerical convergence factors for* $N_t = 256$ *time steps and* $N_x = 1023$.

test we use a zero right-hand side, i.e., $\boldsymbol{f} = \boldsymbol{0}$ and a random initial vector $\boldsymbol{u}^0$ with values between zero and one. The convergence factor of the two-grid cycle is measured by

$$\max_{k=1,\dots,N_{\mathrm{iter}}} \frac{\left\|\boldsymbol{r}^{k+1}\right\|_2}{\left\|\boldsymbol{r}^{k}\right\|_2}, \quad \text{with } \boldsymbol{r}^k := \boldsymbol{f} - \boldsymbol{\mathcal{L}}_{\tau_L,h_L}\boldsymbol{u}^k,$$

where $N_{\mathrm{iter}} \in \mathbb{N}$, $N_{\mathrm{iter}} \leq 250$ is the number of two-grid iterations used until we have reached a given relative error reduction of $\varepsilon_{\mathrm{MG}} = 10^{-140}$. We observe that the numerical results agree very well with the theoretical results, even though the local Fourier mode analysis is not rigorous for the numerical simulation that does not use periodic boundary conditions.

In Figures 4(a)–4(b) we plot the theoretical convergence factors $\varrho(\hat{\mathcal{M}}^{\mathrm{s}}_{\mu})$ for the two-grid cycle $\mathcal{M}^{\mathrm{f}}_{\tau_L,h_L}$ with full space-time coarsening as function of the discretization parameter $\mu \in [10^{-6}, 10^{6}]$ for different polynomial degrees $p_t \in \{0, 1\}$. We observe that the theoretical convergence factors are bounded by $\varrho(\hat{\mathcal{M}}^{\mathrm{f}}_{\mu}) \leq \frac{1}{2}$ if the discretization parameter μ is large enough, i.e., for $\mu \geq \mu^*$. In Remark 4.14 we already computed these critical values μ^* for several polynomial degrees p_t. As before we compared the theoretical results with the numerical results when full space-time coarsening is applied. In Figures 4(a)–4(b) the measured numerical convergence factors are plotted as dots, triangles, and squares. We see that the theoretical results agree very well with the numerical results.

Overall we conclude that the two-grid cycle always converges to the exact solution of the linear system (1.3) when semi-coarsening in time is applied. Furthermore, if the discretization parameter μ is large enough, we can also apply full space-time coarsening, which leads to a smaller coarse problem compared to the semi coarsening case.

4.3. Two-grid analysis with inexact smoother. For the two-grid analysis above we used for the block Jacobi smoother

$$\mathcal{S}^{\nu}_{\tau_L,h_L} = \left[I - \omega_t (D_{\tau_L,h_L})^{-1}\mathcal{L}_{\tau_L,h_L}\right]^{\nu} \tag{4.14}$$

the exact inverse of the diagonal matrix $D_{\tau_L,h_L} = \operatorname{diag}\{A_{\tau_L,h_L}\}_{n=1}^{N_{L_t}}$. In practice, it is more efficient to use an approximation $\widetilde{D}^{-1}_{\tau_L,h_L}$ by applying one multigrid iteration in space for the blocks A_{τ_L,h_L}, see also [36], where such an approximate block Jacobi method is used directly to precondition GMRES. Hence the smoother (4.14) changes to

$$\overline{\mathcal{S}}^{\nu}_{\tau_L,h_L} := \left[I - \omega_t \left(I - \mathcal{M}_{\tau_L,h_L}\right)(D_{\tau_L,h_L})^{-1}\mathcal{L}_{\tau_L,h_L}\right]^{\nu} \tag{4.15}$$

with the matrix $\mathcal{M}_{\tau_L,h_L} := \operatorname{diag}\{\mathcal{M}^x_{\tau_L,h_L}\}_{n=1}^{N_{L_t}}$, where $\mathcal{M}^x_{\tau_L,h_L}$ is the error propergation matrix of the multigrid scheme for the matrix A_{τ_L,h_L}. In the case that the iteration matrix $\mathcal{M}^x_{\tau_L,h_L}$ is given by a two-grid cycle, we further obtain the representation

$$\mathcal{M}^x_{\tau_L,h_L} = \mathcal{S}^{x,\nu_2^x}_{\tau_L,h_L}\left[I - \overline{\mathcal{P}}^{L_x}_x A^{-1}_{\tau_L,2h_L}\overline{\mathcal{R}}^{L_x}_x A_{\tau_L,h_L}\right]\mathcal{S}^{x,\nu_1^x}_{\tau_L,h_L}$$

with a damped Jacobi smoother in space

$$\mathcal{S}^{x,\nu^x}_{\tau_L,h_L} := \left[I - \omega_x \left(D^x_{\tau_L,h_L}\right)^{-1} A_{\tau_L,h_L}\right]^{\nu^x}, \quad D^x_{\tau_L,h_L} := \operatorname{diag}\left\{\frac{2h}{3}K_{\tau_L} + \frac{2}{h}M_{\tau_L}\right\}_{r=1}^{N_{L_x}}$$

and the restriction and prolongation operators

$$\overline{\mathcal{R}}^{L_x}_x := \mathcal{R}^{L_x}_x \otimes I_{N_t} \quad \text{and} \quad \overline{\mathcal{P}}^{L_x}_x := \mathcal{P}^{L_x}_x \otimes I_{N_t}.$$

With the different smoother (4.15) we also have to analyze the two different two-grid iteration matrices

$$\overline{\mathcal{M}}^{\mathrm{s}}_{\tau_L,h_L} := \overline{\mathcal{S}}^{\nu_2}_{\tau_L,h_L}\left[I - \mathcal{P}^{L_x,L_t}_{\mathrm{s}}\left(\mathcal{L}_{2\tau_L,h_L}\right)^{-1}\mathcal{R}^{L_x,L_t}_{\mathrm{s}}\mathcal{L}_{\tau_L,h_L}\right]\overline{\mathcal{S}}^{\nu_1}_{\tau_L,h_L}, \tag{4.16}$$

$$\overline{\mathcal{M}}^{\mathrm{f}}_{\tau_L,h_L} := \overline{\mathcal{S}}^{\nu_2}_{\tau_L,h_L}\left[I - \mathcal{P}^{L_x,L_t}_{\mathrm{f}}\left(\mathcal{L}_{2\tau_L,2h_L}\right)^{-1}\mathcal{R}^{L_x,L_t}_{\mathrm{f}}\mathcal{L}_{\tau_L,h_L}\right]\overline{\mathcal{S}}^{\nu_1}_{\tau_L,h_L}. \tag{4.17}$$

Hence it remains to analyze the mapping property of the operator $\mathcal{M}_{\tau_L,h_L}$ on the space of harmonics $\mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$. By several computations we find under the assumptions of periodic boundary conditions (4.1) that

$$\mathcal{M}_{\tau_L,h_L} : \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t) \to \mathcal{E}_{L_x,L_t}(\theta_x,\theta_t)$$

with the mapping

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \mapsto \begin{pmatrix} \widetilde{M}_{\tau_L,h_L}(\theta_x) & 0 \\ 0 & \widetilde{M}_{\tau_L,h_L}(\theta_x) \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} \tag{4.18}$$

and the iteration matrix

$$\widetilde{M}_{\tau_L,h_L}(\theta_x) := \widetilde{\mathcal{S}}^{x,\nu_1^x}_{\tau_L,h_L}(\theta_x)\mathcal{K}^x_{\tau_L,h_L}(\theta_x)\widetilde{\mathcal{S}}^{x,\nu_2^x}_{\tau_L,h_L}(\theta_x) \in \mathbb{C}^{2N_t\times 2N_t},$$
$$\mathcal{K}^x_{\tau_L,h_L}(\theta_x,\theta_t) := I_{2N_t} - \widetilde{\mathcal{P}}_x(\theta_x)\widetilde{A}^{-1}_{\tau_L,2h_L}(2\theta_x)\widetilde{\mathcal{R}}_x(\theta_x)\widetilde{A}_{\tau_L,h_L}(\theta_x) \in \mathbb{C}^{2N_t\times 2N_t}$$

with the matrices

$$\widetilde{A}_{\tau_L,h_L}(\theta_x) := \begin{pmatrix} \hat{A}_{\tau_L,h_L}(\theta_x) & 0 \\ 0 & \hat{A}_{\tau_L,h_L}(\gamma(\theta_x)) \end{pmatrix} \in \mathbb{C}^{2N_t\times 2N_t},$$
$$\widetilde{A}^{-1}_{\tau_L,2h_L}(2\theta_x) := \left(\hat{A}_{\tau_L,2h_L}(2\theta_x)\right)^{-1} \in \mathbb{C}^{N_t\times N_t},$$
$$\widetilde{\mathcal{S}}^{x,\nu^x}_{\tau_L,h_L}(\theta_x) := \begin{pmatrix} \left(\hat{\mathcal{S}}_{\tau_L,h_L}(\omega_x,\theta_x)\right)^{\nu^x} & 0 \\ 0 & \left(\hat{\mathcal{S}}_{\tau_L,h_L}(\omega_x,\gamma(\theta_x))\right)^{\nu^x} \end{pmatrix} \in \mathbb{C}^{2N_t\times 2N_t},$$
$$\widetilde{\mathcal{R}}_x(\theta_x) := \begin{pmatrix} \hat{\mathcal{R}}_x(\theta_x)I_{N_t} & \hat{\mathcal{R}}_x(\gamma(\theta_x))I_{N_t} \end{pmatrix} \in \mathbb{C}^{2N_t\times N_t},$$
$$\widetilde{\mathcal{P}}_x(\theta_x) := \begin{pmatrix} \hat{\mathcal{P}}_x(\theta_x)I_{N_t} \\ \hat{\mathcal{P}}_x(\gamma(\theta_x))I_{N_t} \end{pmatrix} \in \mathbb{C}^{N_t\times 2N_t}$$

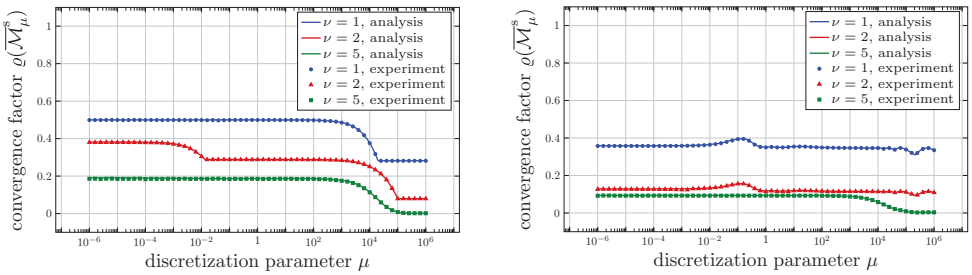
and the Fourier symbols

$$\hat{A}_{\tau_L,h_L}(\theta_x) := \frac{h_L}{3}\left(2+\cos(\theta_x)\right)K_{\tau_L} + \frac{2}{h_L}\left(1-\cos(\theta_x)\right)M_{\tau_L} \in \mathbb{C}^{N_t\times N_t},$$
$$\hat{\mathcal{S}}_{\tau_L,h_L}(\omega_x,\theta_x) := I_{N_t} - \omega_x\left(\frac{2h_L}{3}K_{\tau_L} + \frac{2}{h_L}M_{\tau_L}\right)^{-1}\hat{A}_{\tau_L,h_L}(\theta_x) \in \mathbb{C}^{N_t\times N_t}.$$

Hence we can analyze the modified two-grid iteration matrices (4.16) by taking the additional approximation with the mapping (4.18) into account. For the smoothing steps $\nu_1^x = \nu_2^x = 2$ and the damping parameter $\omega_x = \frac{2}{3}$ for the spatial multigrid component, the theoretical convergence factors with semi coarsening in time are plotted in Figures 5(a)–5(b) for the discretization parameter $\mu \in [10^{-6}, 10^6]$ with respect to the polynomial degrees $p_t \in \{0,1\}$. We observe that the theoretical convergence factors are always bounded by $\varrho(\overline{\mathcal{M}}^s_\mu) \le \frac{1}{2}$. We also notice that the theoretical convergence factors are a little bit larger for small discretization parameters μ, compared to the case when the exact inverse of the diagonal matrix D_{τ_L,h_L} is used. The numerical factors are plotted as dots, triangles, and squares in Figures 5(a)–5(b). We observe that the theoretical convergence factors coincide with the numerical results.

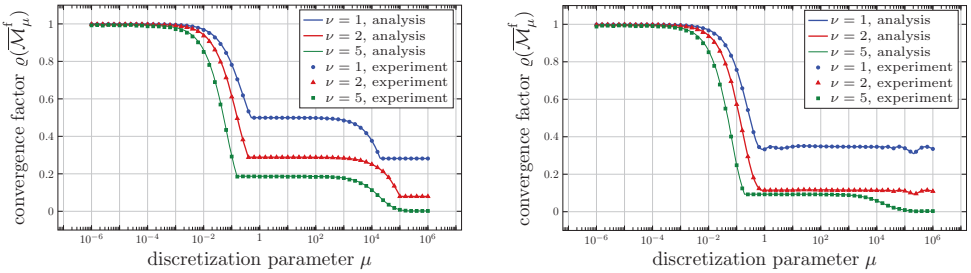
In Figures 6(a)–6(b) the convergence of the two-grid cycle for the full space-time coarsening case is studied. Here we see that the computed convergence factors are very close to the results which we obtained for the case when the exact inverse of the diagonal matrix D_{τ_L,h_L} is used.

5. Numerical examples. We present now numerical results for the multigrid version of our algorithm, for which we analyzed the two-grid cycle in section 4.2. Following our two-grid analysis, we also apply full space-time coarsening only if $\mu_L \ge \mu^*$ in the multigrid version. If $\mu_L < \mu^*$, we only apply semi-coarsening in time. In that case, we will have for the next coarser level $\mu_{L-1} = 2\tau_L h_L^{-2} = 2\mu_L$. This implies that the discretization parameter μ_{L-1} gets larger when semi-coarsening in time is used. Hence, if $\mu_{L-k} \ge \mu^*$ for $k < L$ we can apply full space-time coarsening to reduce the computational costs. If full space-time coarsening is applied, we have



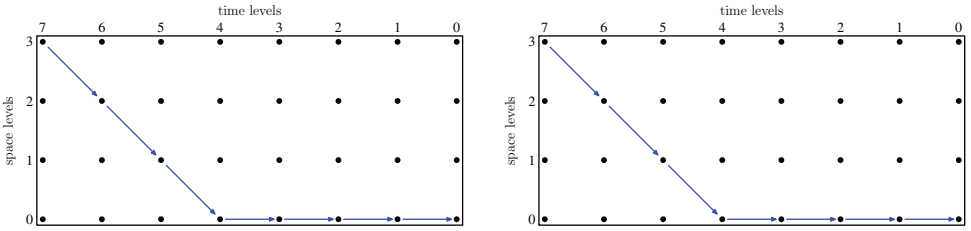
(a) $p_t = 0$. (b) $p_t = 1$.

FIG. 5. *Average convergence factor* $\varrho(\overline{\mathcal{M}}^{\mathrm{s}}_{\mu})$ *for different discretization parameters* μ *and numerical convergence factors for* $N_t = 256$ *time steps and* $N_x = 1023$.



(a) $p_t = 0$. (b) $p_t = 1$.

FIG. 6. *Average convergence factor* $\varrho(\overline{\mathcal{M}}^{\mathrm{f}}_{\mu})$ *for different discretization parameters* μ *and numerical convergence factors for* $N_t = 256$ *time steps and* $N_x = 1023$.



(a) Discretization parameter on the finest space-time level: $\mu_7 = 10\mu^*$.

(b) Discretization parameter on the finest space-time level: $\mu_7 = \mu^*$.

FIG. 7. *Space-time coarsening for different discretization parameters* μ_L.

$\mu_{L-1} = 2\tau_L (2h_L)^{-2} = \frac{1}{2}\mu_L$, which results in a smaller discretization parameter μ_{L-1}. We therefore will combine semi-coarsening in time or full space-time coarsening in the right way to get to the next coarser space-time level. For different discretization parameters $\mu = c\mu^*$, $c \in \{1, 10\}$, this coarsening strategy is shown in Figure 7 for eight time and four space levels. The restriction and prolongation operators for the space-time multigrid scheme are then defined by the given coarsening strategy.

We now show examples to illustrate the performance of our new space-time multigrid method.

TABLE 1
Multigrid iterations for Example 5.1.

		Time levels														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Space levels	0	1	7	7	7	7	7	7	7	8	8	9	9	9	9	9
	1	1	7	7	7	7	7	7	7	8	8	9	9	9	9	9
	2	1	7	7	7	7	7	8	7	8	8	9	9	9	9	9
	3	1	7	7	7	7	8	8	8	8	8	9	9	9	9	9
	4	1	7	7	7	8	8	8	8	8	8	8	9	9	9	9
	5	1	7	7	7	7	7	8	8	8	8	8	9	9	9	9

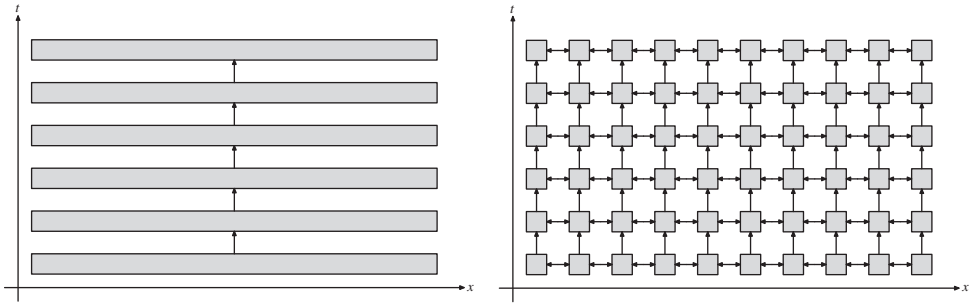
Example 5.1 (multigrid iterations). In this example we consider the spatial domain $\Omega = (0,1)^3$ and the simulation interval $(0,T)$ with $T = 1$. The initial decomposition for the spatial domain Ω is given by 12 tetrahedra. We use several uniform refinement levels to study the convergence behavior of the space-time multigrid solver with respect to the space-discretization. For the coarsest time level we use one time step, i.e., $\tau_0 = 1$ and we increase the number of time steps for finer time levels by a factor of two. For the space discretization we use P1 conforming finite elements, and for the time discretization we use piecewise linear discontinuous ansatz functions, i.e., $p_t = 1$. To test the performance of the space-time multigrid method we use a zero right-hand side, i.e., $\boldsymbol{f} = \boldsymbol{0}$, and as an initial guess $\boldsymbol{u}^0$ we use a random vector with values between zero and one. For the space-time multigrid solver we use $\nu_1 = \nu_2 = 2$, $\omega_t = \frac{1}{2}$, $\gamma = 1$. We apply for each block A_{τ_L,h_L} one geometric multigrid V-cycle to approximate the inverse of the diagonal matrix D_{τ_L,h_L}. For this multigrid cycle we use $\nu_1^x = \nu_2^x = 2$, $\omega_x = \frac{2}{3}$, $\gamma_x = 1$. For the smoother we use a damped block Jacobi smoother, where a block consists of the unknowns with respect to time and is of size $p_t + 1$; see also subsection 4.3. We apply the space-time multigrid solver until we have reached a given relative error reduction of $\varepsilon_{\mathrm{MG}} = 10^{-8}$. In Table 1, the iteration numbers for several space and time levels are given. We observe that the iteration numbers stay bounded independently of the mesh size h_{L_x}, the time step size τ_{L_t}, and the number of time steps $N_{L_t} = 2^{L_t}$.

Example 5.2 (high order time discretizations). In this example we study the convergence of the space-time multigrid method for different polynomial degrees p_t, which are used for the underlying time discretization. To do so, we consider the spatial domain $\Omega = (0,1)^2$ and the simulation interval $(0,T)$ with $T = 1024$. For the space-time discretization we use tensor product space-time elements with piecewise linear continuous ansatz functions in space, and for the discretization in time we use a fixed time step size $\tau = 1$. For the initial triangulation of the spatial domain Ω we consider four triangles, which are refined uniformly several times. For the space-time multigrid approach we use the same parameters as in Example 5.1. We solve the linear system (3.1) with zero right-hand side, i.e., $\boldsymbol{f} = \boldsymbol{0}$, and for the initial vector $\boldsymbol{u}^0$ we use a random vector with values between zero and one. We apply the space-time multigrid solver until we have reached a relative error reduction of $\varepsilon_{\mathrm{MG}} = 10^{-8}$. In Table 2 the iteration numbers for different polynomial degrees p_t and different space levels are given. We observe that the iteration numbers are bounded, independently of the ansatz functions for the time discretization.

6. Parallelization. One big advantage of our new space-time multigrid method is that it can be parallelized also in the time direction, i.e., the damped block Jacobi smoother can be executed in parallel in time. For each time step we have to apply one

TABLE 2
Multigrid iterations with respect to the polynomial degree p_t.

Space levels	Polynomial degree p_t													
	0	1	2	3	4	5	10	15	20	25	30	35	40	45
0	7	7	6	6	6	6	5	5	4	4	4	4	5	5
1	7	7	7	7	7	7	7	7	7	7	7	7	7	7
2	7	7	7	7	7	7	7	7	7	7	7	7	7	7
3	7	7	7	7	7	7	7	7	7	7	7	7	7	7
4	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	7	7	7	7	7	7	7	7	7	7	7	7	7	7
6	7	7	7	7	7	7	7	7	7	7	7	7	7	7
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7



(a) Parallelization only in the time direction. (b) Full space-time parallelization.

FIG. 8. *Communication pattern on a fixed level.*

multigrid cycle in space to approximate the inverse of the diagonal matrix D_{τ_L,h_L}. The application of this space multigrid cycle can also be done in parallel, where one may use parallel packages like in [8, 7, 26]. Hence the problem (3.1) can be fully parallelized in space and time; see Figure 8.

The next example shows the excellent weak and strong scaling properties of our new space-time multigrid method.

Example 6.1 (parallel computations). In this example we consider the spatial domain $\Omega = (0,1)^3$, which is decomposed into 49,152 tetrahedra. For the discretization in space we use P1 conforming finite elements, and for the time discretization we use polynomials of order $p_t = 3$, which allows from an approximation point of view the time step size $\tau = 10^{-1}$. For the multigrid solver in space we use the best possible settings such that we obtain the smallest computational times when we apply the usual forward substitution. In particular we use a damped Gauss–Seidel smoother with the damping parameter $\omega_x = 1.285$, and we apply one pre- and one post-smoothing step, i.e., $\nu_1^x = \nu_2^x = 1$. We also tune the multigrid parameters with respect to time, such that we also obtain the best possible computational times for the presented space-time multigrid solver. Here we use $\nu_1 = \nu_2 = 1$ smoothing steps, and since $\mu = \tau h^{-2}$ is large enough we use for the damping parameter $\omega_t = 1$; see also Remark 4.15. Of course we could also use the asymptotic optimal damping parameter $\omega_t^* = \frac{1}{2}$ which would lead to slightly more multigrid iterations for this case. To study the scaling behavior only with respect to time we keep the discretization parameter μ fixed for all space-time levels, i.e., we apply no refinement in space and fix the time step size. Of course it is possible to use also parallel solvers like [8, 7] with respect to the problem

TABLE 3
Scaling results with solving times in seconds for Example 6.1.

(a) Weak scaling results.

Cores	Time steps	dof	iter	Time	fwd. sub.
1	2	59 768	7	28.8	19.0
2	4	119 536	7	29.8	37.9
4	8	239 072	7	29.8	75.9
8	16	478 144	7	29.9	152.2
16	32	956 288	7	29.9	305.4
32	64	1 912 576	7	29.9	613.6
64	128	3 825 152	7	29.9	1 220.7
128	256	7 650 304	7	29.9	2 448.4
256	512	15 300 608	7	30.0	4 882.4
512	1 024	30 601 216	7	29.9	9 744.2
1 024	2 048	61 202 432	7	30.0	19 636.9
2 048	4 096	122 404 864	7	29.9	38 993.1
4 096	8 192	244 809 728	7	30.0	81 219.6
8 192	16 384	489 619 456	7	30.0	162 551.0
16 384	32 768	979 238 912	7	30.0	313 122.0
32 768	65 536	1 958 477 824	7	30.0	625 686.0
65 536	131 072	3 916 955 648	7	30.0	1 250 210.0
131 072	262 144	7 833 911 296	7	30.0	2 500 350.0
262 144	524 288	15 667 822 592	7	30.0	4 988 060.0

(b) Strong scaling results.

Cores	Time steps	dof	iter	Time
1	512	15 300 608	7	7 635.2
2	512	15 300 608	7	3 821.7
4	512	15 300 608	7	1 909.9
8	512	15 300 608	7	954.2
16	512	15 300 608	7	477.2
32	512	15 300 608	7	238.9
64	512	15 300 608	7	119.5
128	512	15 300 608	7	59.7
256	512	15 300 608	7	30.0
512	524 288	15 667 822 592	7	15 205.9
1 024	524 288	15 667 822 592	7	7 651.5
2 048	524 288	15 667 822 592	7	3 825.3
4 096	524 288	15 667 822 592	7	1 913.4
8 192	524 288	15 667 822 592	7	956.6
16 384	524 288	15 667 822 592	7	478.1
32 768	524 288	15 667 822 592	7	239.3
65 536	524 288	15 667 822 592	7	119.6
131 072	524 288	15 667 822 592	7	59.8
262 144	524 288	15 667 822 592	7	30.0

in space. This leads to the fully parallel space-time solver as mentioned in Figure 8, which will be analyzed in some future work.

To show the parallel performance, we first study the weak scaling behavior of the new multigrid method. We use a fixed number of time steps per core, i.e., 2 time steps for each core, and we increase the number of cores when we increase the number of time steps. Hence the computational cost for one space-time multigrid cycle stays almost the same for each core. But the cost for the communication grows, since the space-time hierarchy gets bigger, when we increase the number of time steps. In Table 3(a), we give timings for solving the linear system (3.1) for a different number of time steps. We see that the multigrid iterations stay bounded if we increase the problem size and that the computational costs stay completely constant if we increase the number of cores. We also compare the presented space-time multigrid solver with the usual forward substitution. For this we apply for each time step the space multigrid solver with the best possible settings from above. We run the space multigrid solver until we obtain the same relative error tolerance as for the space-time multigrid method. In Table 3(a) the timings for the forward substitution are compared with the parallel space-time multigrid solver. Note that for the forward substitution we have no parallelization, neither in space nor in time. We see that the space-time multigrid solver is already faster when we use only two cores. Furthermore, when we increase the number of cores we observe that the space-time multigrid approach completely outperforms the forward substitution. Then the largest problem takes 58 days to be solved by forward substitution on one core, compared to 30 seconds with our new space-time multigrid method on 524,288 cores.

To test the strong scaling behavior, we fix the problem size for 1 to 256 cores by using 512 time steps, which results in a linear system with 15,300,608 unknowns. For 512 up to 262,144 cores we use 524,288 time steps, which results in 15,667,822,592 unknowns. Then we increase the number of cores, which results in smaller problems per computing core. In Table 3(b) the timings are given for different numbers of cores. We see that the computational costs are divided by a factor very close to two if we double the number of cores. All the parallel computations of this example were performed on Vulcan BlueGene/Q Supercomputer in Livermore, CA.

Example 6.2 (parallel tests for one spatial unkown). To show the scaling behavior with respect to time in more detail we test the weak and strong scaling behavior of our new space-time multigrid algorithm for the case when we have only one degree of

TABLE 4
Scaling results with solving times in seconds for Example 6.2.

(a) Weak scaling results.

Cores	Time steps	$p_t = 0$	$p_t = 1$	$p_t = 5$	$p_t = 10$	$p_t = 20$
1	32 768	3.96	2.45	3.60	7.97	17.77
2	65 536	5.30	3.39	5.10	10.74	25.83
4	131 072	5.34	3.38	5.45	10.84	26.13
8	262 144	5.95	3.77	5.75	11.70	28.54
16	524 288	5.94	3.77	5.78	11.69	28.61
32	1 048 576	5.96	3.80	5.87	11.73	27.21
64	2 097 152	7.92	4.73	6.68	12.62	27.96
128	4 194 304	8.03	4.74	6.66	12.76	28.52
256	8 388 608	8.09	4.91	6.80	12.90	28.29
512	16 777 216	8.06	4.79	6.75	12.82	29.13
1 024	33 554 432	7.96	4.75	6.69	13.01	28.89
2 048	67 108 864	8.06	4.79	6.73	13.02	28.86
4 096	134 217 728	8.14	4.80	6.77	12.77	29.18
8 192	268 435 456	8.14	4.89	6.84	13.03	29.23
16 384	536 870 912	8.10	4.80	6.82	13.25	29.52
32 768	1 073 741 824	8.21	4.94	6.90	13.19	29.03

(b) Strong scaling results.

Cores	Time steps	$p_t = 0$	$p_t = 1$	$p_t = 5$	$p_t = 10$	$p_t = 20$
1	1 048 576	129.11	78.79	117.99	254.95	535.43
2	1 048 576	85.66	54.69	82.36	172.46	396.71
4	1 048 576	42.90	27.22	41.17	86.81	199.71
8	1 048 576	23.83	15.08	22.90	46.87	107.77
16	1 048 576	11.91	7.57	11.50	23.40	54.06
32	1 048 576	5.96	3.80	5.87	11.73	27.21
64	1 048 576	3.98	2.45	3.31	6.50	13.70
128	1 048 576	1.97	1.18	1.67	3.23	6.97
256	1 048 576	0.984	0.598	0.808	1.57	3.49
512	1 048 576	0.508	0.299	0.407	0.787	1.77
1 024	1 048 576	0.264	0.155	0.210	0.444	0.904
2 048	1 048 576	0.146	0.0864	0.114	0.210	0.465
4 096	1 048 576	0.0861	0.0506	0.0653	0.116	0.243
8 192	1 048 576	0.0548	0.0329	0.0405	0.0743	0.129
16 384	1 048 576	0.0400	0.0230	0.0272	0.0424	0.0767
32 768	1 048 576	0.0511	0.0241	0.0288	0.0376	0.0608

freedom in space. In particular we study the case

$$M_h = (1) \quad \text{and} \quad K_h = (1).$$

We use different polynomial degrees $p_t \in \{0, 1, 5, 10, 20\}$ and a fixed time step size $\tau = 10^{-6}$. For a random initial guess and a zero right-hand side we run the algorithm until we have reached a relative error reduction of $\varepsilon_{\mathrm{MG}} = 10^{-8}$. We first study the weak scaling behavior by using a fixed number of time steps per core (32,768), and we increase the number of cores when increasing the number of time steps. In Table 4(a), we give computation times for different numbers of cores and polynomial degrees. We observe excellent weak scaling, i.e., the computation times remain bounded when we increase the number of cores. We next study the strong scaling behavior by fixing the problem size, i.e., we use 1,048,576 time steps in this example. Then we increase the number of cores from 1 up to 32,768. In Table 4(b) the computation times are given for different number of cores and polynomial degrees. We observe that the computation costs are basically divided by a factor of two if we double the number of cores, but for 32,768 cores and $p_t \in \{0, 1, 5\}$ we obtain no speedup any more, since the local problems are to small, i.e., for $p_t = 0$ one core has to solve for only 32 unknowns.

These computations were performed on the Monte Rosa supercomputer at the Swiss National Supercomputing Centre CSCS in Lugano.

7. Conclusions. We presented a new space-time multigrid method for the heat equation and used local Fourier mode analysis to give precise asymptotic convergence and parameter estimates for the two-grid cycle. We showed that this asymptotic analysis predicts very well the performance of the new algorithm, and our parallel implementation gave excellent weak and strong scaling results for a large number of processors.

This new space-time multigrid algorithm can be used not only for the heat equation; it is also applicable to general parabolic problems. It has successfully been applied to the time dependent Stokes equations, where one obtains similar speed up results as for the heat equation. Furthermore, this technique has been applied successfully to parabolic control problems, but the analysis and the results will appear elsewhere.

Acknowledgments. We thank Ernst Hairer for his help with Theorem 2.1. We also want to thank Panayot S. Vassilevski for giving us the opportunity to compute on the LC machines in Livermore.

REFERENCES

[1] G. Bal, *On the convergence and the stability of the parareal algorithm to solve partial differential equations*, Lect. Notes Comput. Sci. Eng. 40, Springer, Berlin, 2005, pp. 425–432.

[2] D. Bennequin, M. J. Gander, and L. Halpern, *A homographic best approximation problem with application to optimized Schwarz waveform relaxation*, Math. Comp., 78 (2009), pp. 185–223.

[3] S. Börm and R. Hiptmair, *Analysis of tensor product multigrid*, Numer. Algorithms, 26 (2001), pp. 219–234.

[4] F. Chipman, *A-stable Runge–Kutta processes*, Nordisk Tidskr. Informationsbehandling (BIT), 11 (1971), pp. 384–388.

[5] B. Ehle, *On Padé Approximations to the Exponential Function and A-Stable Methods for the Numerical Solution of Initial Value Problems*, Ph.D. thesis, Technical report CSRR 2010, Department AACS University of Waterloo, Ontario, Canada, 1969.

[6] M. Emmett and M. L. Minion, *Toward an efficient parallel in time method for partial differential equations*, Commun. Appl. Math. Comput. Sci, 7 (2012), pp. 105–132.

[7] R. Falgout, J. Jones, and U. Yang, *The design and implementation of hypre, a library of parallel high performance preconditioners*, Lect. Notes Comput. Sci. Eng. 51, 2006, pp. 267–294.

[8] R. Falgout and U. Yang, *hypre: A Library of High Performance Preconditioners*, in Proceedings of the International Conference on Computational Science—Part III, 2002, pp. 632–641.

[9] M. J. Gander and E. Hairer, *Nonlinear convergence analysis for the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering XVII, O. B. Widlund and D. E. Keyes, eds., Lect. Notes Comput. Sci. Eng. 60, Springer, Berlin, 2008, pp. 45–56.

[10] M. J. Gander, L. Halpern, and F. Nataf, *Optimal convergence for overlapping and non-overlapping Schwarz waveform relaxation*, in Proceedings of the Eleventh International Conference of Domain Decomposition Methods. C.-H. Lai, P. Bjørstad, M. Cross, and O. Widlund, eds., ddm.org, 1999.

[11] M. J. Gander, L. Halpern, and F. Nataf, *Optimal Schwarz waveform relaxation for the one dimensional wave equation*, SIAM J. Numer. Anal., 41 (2003), pp. 1643–1681.

[12] M. J. Gander and L. Halpern, *Absorbing boundary conditions for the wave equation and parallel computing*, Math. Comp., 74 (2004), pp. 153–176.

[13] M. J. Gander and L. Halpern, *Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems*, SIAM J. Numer. Anal., 45 (2007), pp. 666–697.

[14] M. J. Gander, Y.-L. Jiang, and R.-J. Li, *Parareal Schwarz waveform relaxation methods*, in Domain Decomposition Methods in Science and Engineering XX, O. B. Widlund and D. E. Keyes, eds., Lect. Notes Comput. Sci. Eng. 60, Springer, Berlin, 2013, pp. 45–56.

[15] M. J. Gander, F. Kwok, and B. Mandal, *Dirichlet–Neumann and Neumann–Neumann Waveform Relaxation Algorithms for Parabolic Problems*, ETNA, submitted.
[16] M. J. Gander and A. M. Stuart, *Space-time continuous analysis of waveform relaxation for the heat equation*, SIAM J. Sci. Comput., 19 (1998), pp. 2014–2031.
[17] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput., 29 (2007), pp. 556–578.
[18] M. J. Gander, *A waveform relaxation algorithm with overlapping splitting for reaction diffusion equations*, Numer. Linear Algebra Appl., 6 (1998), pp. 125–145.
[19] M. J. Gander, 50 *years of time parallel time integration*, in Multiple Shooting and Time Domain Decomposition Methods, Springer, Berlin, 2014.
[20] E. Giladi and H. B. Keller, *Space time domain decomposition for parabolic problems*, Numer. Math., 93 (2002), pp. 279–313.
[21] W. Hackbusch, *Parabolic multigrid methods*, Comput. Methods Appl. Sci. Eng., 7 (1984), pp. 189–197.
[22] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer, Berlin, 1985.
[23] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Mat. 31, Springer, Berlin, 2010.
[24] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations.* I. *Nonstiff Problems*, Springer Ser. Comput. Math. 8, Springer, Berlin, 1993.
[25] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations.* II. *Stiff and Differential-Algebraic Problems*, Springer Ser. Comput. Mat. 14, Springer, Berlin, 2010.
[26] I. Heppner, M. Lampe, A. Nägel, S. Reiter, M. Rupp, A. Vogel, and G. Wittum, *Software Framework ug4: Parallel Multigrid on the Hermit Supercomputer*, High Performance Comput. Sci. Eng., 12 (2013), pp. 435–449.
[27] G. Horton, S. Vandewalle, and P. Worley, *An algorithm with polylog parallel complexity for solving parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 531–541.
[28] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 848–864.
[29] G. Horton, *The time-parallel multigrid method*, Comm. Appl. Numer. Methods, 8 (1992), pp. 585–595.
[30] F. Kwok, *Neumann–Neumann Waveform Relaxation for the Time-Dependent Heat Equation*, in Domain Decomposition Methods in Science and Engineering, DD21, Springer, Berlin, 2014.
[31] J.-L. Lions, Y. Maday, and G. Turinici, *A parareal in time discretization of PDEs*, C. R. Acad. Sci. Paris Ser. I, 332 (2001), pp. 661–668.
[32] C. Lubich and A. Ostermann, *Multigrid dynamic iteration for parabolic equations*, BIT, 27 (1987), pp. 216–234.
[33] Y. Maday and G. Turinici, *The Parareal in Time Iterative Solver: A Further Direction to Parallel Implementation*, Lect. Notes Comput. Sci. Eng. 40, Springer, Berlin, 2005, pp. 441–448.
[34] Y. Maday, *A parareal in time procedure for the control of partial differential equations*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 387–392.
[35] B. Mandal, *A time-dependent Dirichlet–Neumann method for the heat equation*, in Domain Decomposition Methods in Science and Engineering, DD21, Springer, Berlin, 2014.
[36] E. G. McDonald and A. J. Wathen, *A Simple Proposal for Parallel Computing over Time of an Evolutionary Process with Implicit Time Stepping*, The Mathematical Institute, University of Oxford, Eprints Archive, 2014.
[37] R. Speck, D. Ruprecht, M. Emmett, M. Minion, M. Bolten, and R. Krause, *A multi-level spectral deferred correction method*, BIT, 55 (2015), pp. 843–867.
[38] R. Speck, D. Ruprecht, R. Krause, M. Emmett, M. Minion, M. Winkel, and P. Gibbon, *A massively space-time parallel n-body solver*, in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, IEEE Computer Society Press, Los Alamitos, CA, 2012, p. 92.
[39] G. Staff and E. Rønquist, *Stability of the Parareal Algorithm*, Lect. Notes Comput. Sci. Eng. 40, Springer, Berlin, 2005, pp. 449–456.
[40] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 2006.
[41] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, San Diego, CA, 2001.
[42] S. Vandewalle and E. de Velde, *Space-time concurrent multigrid waveform relaxation*, Ann. Numer. Math., 1 (1994), pp. 347–360.

[43] S. Vandewalle and R. Piessens, *Efficient parallel algorithms for solving initial-boundary value and time-periodic parabolic partial differential equations*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1330–1346.
[44] P. Vassilevski, *Multilevel Block Factorization Preconditioners*, Springer, Berlin, 2008.
[45] T. Weinzierl and T. Köppl, *A geometric space-time multigrid algorithm for the heat equation*, Numer. Math. Theory Methods Appl., 5 (2012), pp. 110–130.
[46] P. Wesseling, *An Introduction to Multigrid Methods*, John Wiley & Sons, New York, 1992. Corrected Reprint, R.T. Edwards, Philadelphia, PA, 2004.