Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

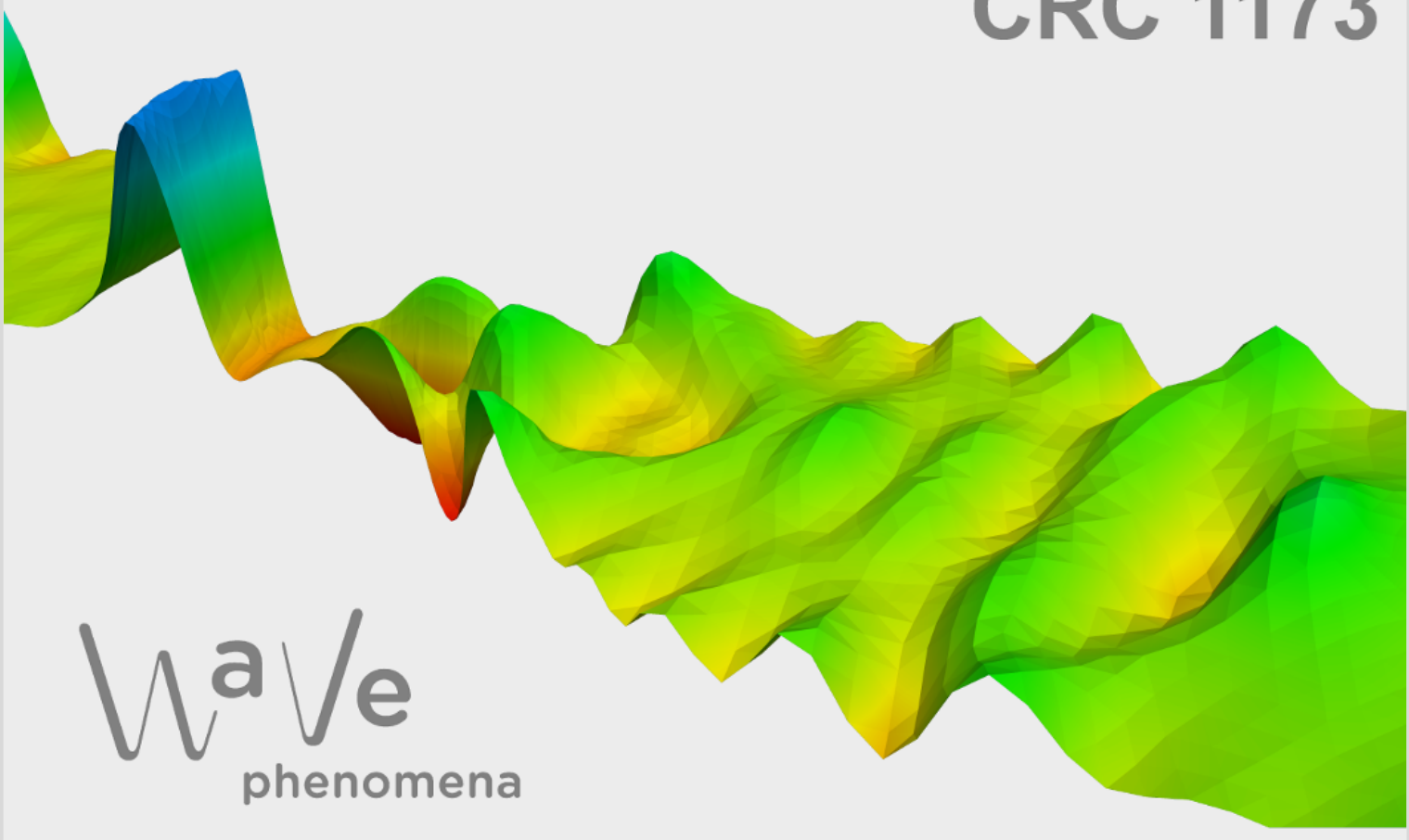
Space-time discontinuous Galerkin discretizations for linear first-order hyperbolic evolution systems

Willy Dörfler, Stefan Findeisen, Christian Wieners

CRC Preprint 2015/4, October 2015

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



KIT – University of the State of Baden-Wuerttemberg and
National Research Centre of the Helmholtz Association

www.kit.edu

Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Funded by

ISSN 2365-662X

SPACE-TIME DISCONTINUOUS GALERKIN DISCRETIZATIONS FOR LINEAR FIRST-ORDER HYPERBOLIC EVOLUTION SYSTEMS

Willy Dörfler, Stefan Findeisen and Christian Wieners[1]

Abstract. We introduce a space-time discretization for linear first-order hyperbolic evolution systems using a discontinuous Galerkin approximation in space and a Petrov–Galerkin scheme in time. We show well-posedness and convergence of the discrete system. Then we introduce an adaptive strategy based on goal-oriented dual-weighted error estimation. The full space-time linear system is solved with a parallel multilevel preconditioner. Numerical experiments for the linear transport equation and the Maxwell equation in 2D underline the efficiency of the overall adaptive solution process.

1991 Mathematics Subject Classification. 65N30.

October 21, 2015.

1. Introduction

Space-time methods for time-dependent PDEs discretize the full problem in the space-time cylinder, and then the corresponding large algebraic system also is solved for the full problem. This is in contrast to the method of lines or Rothe's method, which first use a discretization *either* in space *or* in time and then apply standard techniques for the other variable. Our methods are based on treating space *and* time simultaneously in a variational manner. Depending on the choice of the ansatz and the test spaces, the methods become either explicit or implicit. Explicit methods are computationally efficient but suffer from severe limitations for the time step size, where the length of the time edge of the space-time elements is restricted by the smallest local resolution scale in space. To circumvent these restrictions, we focus on implicit methods.

A fully implicit space-time approach allows for flexible adaptive discretizations which combine adaptivity in space with local time stepping. A further motivation for developing space-time methods is the design of modern computer facilities with an enormous number of processor cores, where the parallel realization of conventional methods becomes inefficient. Since these machines allow a fully implicit space-time approach, new parallel solution techniques are required to solve the huge linear systems, particularly for time-dependent applications in three spatial dimensions.

In recent years, discontinuous Galerkin (DG) methods in space have become very popular, e.g., see [HW08] for time-dependent first-order systems, where this discretization is coupled with explicit time integration. An application of this method to acoustic and elastic waves is considered in [DKT07] combined with an adaptive space-time *hp*-strategy. Here, we extend these spatial DG discretization by a Petrov–Galerkin method in time with continuous ansatz space and discontinuous test space (see, e.g., [BR99]). A space-time method for

Keywords and phrases: space-time methods, discontinuous Galerkin finite elements, linear hyperbolic systems, transport equation, wave equation, Maxwell's equations

[1] Institut für Angewandte und Numerische Mathematik, KIT, Karlsruhe, Germany. Email: christian.wieners@kit.edu

elastic waves with a second-order formulation in space and implicit discontinuous Galerkin time discretization is considered in [KB13].

An alternative discontinuous Petrov–Galerkin (DPG) approach is proposed by L. Demkowicz (see [DG14] for an overview and [EDCM14] for space-time applications) for general linear first-order systems, where weak approximations are constructed by introducing skeleton variables. The application of this technique to the time-harmonic case is analyzed in [ZMD$^+$11]. For acoustic and elastic waves, the hybridization in space (applied to the second-order formulation) is presented in [NPC11], and a hybrid space-time discontinuous Galerkin method is proposed in [WTF14]. Both methods are implicit in every time slab, and only Dirichlet traces are used for the hybrid coupling. Space-time (Trefftz) discontinuous Galerkin methods for wave problems are analyzed in [EKSW14, KMPS15].

Error estimation for linear wave equations (considered as second-order equations) is studied in [BR99, OPD05, GS09] and for more general hyperbolic systems in [HS06]. Simple residual error indications are not sufficient for wave problems since, in the hyperbolic case, the error is transported and thus not correlated to large local residuals. Reliable error control requires the adjoint problem, as it is introduced for goal-oriented techniques in [BR99], to be solved. This technique requires a variational approach, since this allows for an error representation with respect to a given linear error functional.

In principle, all parallel solution methods in space apply also to implicit time integration schemes. Parallel strategies in time are studied extensively on the basis of the 'parareal' idea [LMT01, ARW95, GV07]. A general overview over the most popular algorithms and software packages is given in [Gan15]. Methods such as MGRID [FFK$^+$14] and PFASST [EM12] were developed under the aspect that they can be easily incorporated into existing time sequential code. In addition, solution concepts specially adapted to the full space-time problem were proposed. E.g., the wavefront method extends a spatial domain decomposition into time slices, see [GHN03] for an application to the one-dimensional wave equation. In [GN14] a space-time multigrid method for parabolic problems is analyzed. A multigrid method for higher order discontinuous Galerkin discretizations of advection problems is proposed in [vdVR12].

In this paper we present an fully implicit and parallel adaptive space-time discontinuous Galerkin discretization for linear first-order hyperbolic problems. It is structured as follows. In Sect. 2 we introduce a setting for linear hyperbolic operators by reference to applications in the field of linear transport and acoustic and electro-magnetic waves, and we establish the well-posedness of the space-time variational problem based on a technique developed in [WW14]. Then, following the setting established in [HPS$^+$14], we consider a semi-discrete discontinuous Galerkin discretization in spatial direction with upwind flux. On this basis we define an implicit Petrov–Galerkin space-time discretization in Sect. 4, and we prove well-posedness of the discrete method and convergence on tensor product space-time meshes. Next we propose a goal-oriented space-time error indicator based on the explicit computation of the dual solution. In Sect. 6 a multilevel preconditioner with semi-coarsening first in time and then in space is defined. Within the parallel finite element software system M++ [Wie10] the adaptive method and the multilevel solution method is realized, and the efficiency of the full scheme is demonstrated for two models, the linear transport equation and Maxwell's equations in 2D.

2. A space-time setting for linear hyperbolic operators

Let $\Omega \subseteq \mathbb{R}^D$ be a bounded Lipschitz domain, and let $H \subseteq \mathrm{L}_2(\Omega)^J$ be a Hilbert space with weighted inner product $(\mathbf{v},\mathbf{w})_H = (M\mathbf{v},\mathbf{w})_{0,\Omega}$, where $M \in \mathrm{L}_\infty(\Omega)^{J\times J}$ is uniformly positive and symmetric. We consider a linear operator $A : D(A) \to H$ with domain $D(A) \subset H$. For given initial function $\mathbf{u}_0 \in D(A)$, final time $T > 0$ and right-hand side $\mathbf{f} \in \mathrm{L}_2(0,T;H)$, we study the evolution equation

$$M\partial_t \mathbf{u}(t) + A\mathbf{u}(t) = \mathbf{f}(t)\,, \qquad t \in (0,T)\,, \qquad \mathbf{u}(0) = \mathbf{u}_0\,. \tag{1}$$

We specialize A to the case of linear balance laws determined by a *flux function* $\mathbf{F}(\mathbf{v}) = [B_1\mathbf{v},\dots,B_D\mathbf{v}]$ with symmetric matrices $B_j \in \mathrm{L}_\infty(\Omega)^{J,J}$ such that

$$A\mathbf{v} = \operatorname{div}\mathbf{F}(\mathbf{v}) = \sum_{d=1}^{D} \partial_d(B_d\mathbf{v}) \in \mathrm{L}_2(\Omega)^J\,, \qquad \mathbf{v} \in D(A)\,.$$

Since the matrices B_d are symmetric, any linear combination $n_1B_1 + \cdots + n_DB_D$ for $\mathbf{n} = (n_1,\dots,n_D)^\top \in \mathbb{R}^D$ is diagonalizable with real eigenvalues, so that (1) is a linear hyperbolic system [Eva10, Chap. 7.3]. We will consider the following examples.

Linear transport. For a scalar model problem ($J = 1$), we consider the transport equation to determine $u : \Omega \times (0,T) \to \mathbb{R}$ such that

$$\rho\partial_t u + \operatorname{div}(u\mathbf{q}) = f\,, \qquad \text{on } \Omega\times(0,T)\,, \qquad u(0,.) = u_0\,,$$

for a given vector field $\mathbf{q} \in \mathrm{W}^1_\infty(\Omega)^D$ with $\operatorname{div}\mathbf{q} = 0$ and a density distribution $\rho \in \mathrm{L}_\infty(\Omega)$ satisfying $\rho \geq \rho_0$ a.e. for some $\rho_0 > 0$. This defines the inflow and outflow boundary $\Gamma_{\mathrm{in}} = \overline{\{x \in \partial\Omega\colon \mathbf{q}(x)\cdot\mathbf{n}(x) < 0\}}$ and $\Gamma_{\mathrm{out}} = \overline{\{x \in \partial\Omega\colon \mathbf{q}(x)\cdot\mathbf{n}(x) > 0\}}$ (where $\mathbf{n}$ is the outer unit normal), the flux function $\mathbf{F}(u) = u\mathbf{q}$, hence $Au = \operatorname{div}(u\mathbf{q})$ with domain $D(A) = \{u \in \mathrm{H}^1(\Omega)\colon u = 0 \text{ on } \Gamma_{\mathrm{in}}\}$, $H = \mathrm{L}_2(\Omega)$, and $Mu = \rho u$. For the adjoint operator A^* the roles of the inflow and outflow boundary are interchanged and hence $A^*u = -\operatorname{div}(u\mathbf{q})$ with domain $D(A^*) = \{u^* \in \mathrm{H}^1(\Omega)\colon u^* = 0 \text{ on } \Gamma_{\mathrm{out}}\}$.

Acoustic waves. Acoustic waves in isotropic and homogeneous media (with density $\rho \equiv 1$) are described by

$$\partial_t p + \operatorname{div}\mathbf{v} = f\,, \qquad \partial_t\mathbf{v} + \boldsymbol{\nabla} p = \mathbf{0}$$

for the pressure $p : \Omega\times(0,T) \to \mathbb{R}$ and the velocity $\mathbf{v} : \Omega\times(0,T) \to \mathbb{R}^D$. We set $\mathbf{u} = (\mathbf{v},p)$, $H = \mathrm{L}_2(\Omega)^{D+1}$ and $M(\mathbf{v},p) = (\mathbf{v},p)$. The operator A is defined by $A(\mathbf{v},p) = (\boldsymbol{\nabla}p, \operatorname{div}\mathbf{v})$ which corresponds to the flux function $\mathbf{F}$ given by $B_j = \mathbf{e}_j \otimes \mathbf{e}_{D+1} + \mathbf{e}_{D+1}\otimes\mathbf{e}_j$. In the case of homogeneous Dirichlet boundary conditions, the domain is given by $D(A) = D(A^*) = \mathrm{H}(\operatorname{div},\Omega) \times \mathrm{H}^1_0(\Omega)$ with $A^*(\mathbf{v},p) = (-\boldsymbol{\nabla}p, -\operatorname{div}\mathbf{v})$.

Electro-magnetic waves. For given permeability μ and permittivity ε, electro-magnetic waves are determined by the first-order system for the electric field $\mathbf{E}\colon \Omega\times(0,T) \to \mathbb{R}^3$ and magnetic field $\mathbf{H}\colon \Omega\times(0,T) \to \mathbb{R}^3$

$$\varepsilon\partial_t\mathbf{E} - \operatorname{curl}\mathbf{H} = \mathbf{f}\,, \quad \mu\partial_t\mathbf{H} + \operatorname{curl}\mathbf{E} = \mathbf{0}\,, \quad \operatorname{div}(\varepsilon\mathbf{E}) = \rho\,, \quad \operatorname{div}(\mu\mathbf{H}) = 0$$

for the $J = 6$ components $(\mathbf{E},\mathbf{H})$. Here, we set $H = \mathrm{L}_2(\Omega)^3 \times \mathrm{L}_2(\Omega)^3$, $M(\mathbf{E},\mathbf{H}) = (\varepsilon\mathbf{E},\mu\mathbf{H})$, and the operators $A(\mathbf{E},\mathbf{H}) = (-\operatorname{curl}\mathbf{H}, \operatorname{curl}\mathbf{E}) = -A^*(\mathbf{E},\mathbf{H})$ in $D(A) = D(A^*) = \mathrm{H}_0(\operatorname{curl},\Omega)\times\mathrm{H}(\operatorname{curl},\Omega)$ for a perfect conducting boundary. Here, the matrices B_j are given by $B_j(\mathbf{E},\mathbf{H}) = (-\mathbf{e}_j\times\mathbf{H}, \mathbf{e}_j\times\mathbf{E})$. The divergence constraints require the compatibility condition $\operatorname{div}\mathbf{f} = \rho$ for the right-hand side. Note that in case of polarized electro-magnetic waves this 3D setting can be reduced to a 2D setting.

The variational setting. In the abstract setting, we consider the operator $L = M\partial_t + A$ on the space-time cylinder $Q = \Omega \times (0,T)$ with the domain $V = D(L)$, where V is the closure of $\{\mathbf{v} \in \mathrm{C}^1(0,T;D(A))\colon \mathbf{v}(0) = \mathbf{0}\}$ with respect to the weighted graph norm $\|\mathbf{v}\|_V^2 = (M\mathbf{v},\mathbf{v})_{0,Q} + (M^{-1}L\mathbf{v},L\mathbf{v})_{0,Q}$. The corresponding dual space V^* is the closure of $\{\mathbf{v}^* \in \mathrm{C}^1(0,T;D(A^*))\colon \mathbf{v}^*(T) = \mathbf{0}\}$. Then we define $W = \overline{L(V)} \subseteq \mathrm{L}_2(0,T;H)$ with the weighted norm $\|\mathbf{w}\|_W^2 = (M\mathbf{w},\mathbf{w})_{0,Q}$. Note that in terms of this definition, the norm in V also reads $\|\mathbf{v}\|_V^2 = \|\mathbf{v}\|_W^2 + \|M^{-1}L\mathbf{v}\|_W^2$.

In the subsequent analysis, we assume homogeneous initial and boundary conditions that are included in the domain $D(L)$. Our considerations extend to initial values $\mathbf{u}_0 \neq \mathbf{0}$ by replacing $\mathbf{f}(t)$ with $\mathbf{f}(t) - A\mathbf{u}_0$ in (1). Also inhomogeneous boundary conditions can be analyzed by modifying the right-hand side when the existence of a sufficiently smooth extension of the boundary data can be assumed.

We define the bilinear form $b\colon V \times W \longrightarrow \mathbb{R}$ with $b(\mathbf{v},\mathbf{w}) = (L\mathbf{v},\mathbf{w})_{0,Q}$, and we establish the standard Babuška setting (see, e.g., [Bra07, Thm. III.3.6]).

Lemma 1. *Assume that $(A\mathbf{v},\mathbf{v})_{0,\Omega} \geq 0$ for $\mathbf{v} \in D(A)$. Then, the bilinear form $b(\cdot,\cdot)$ is continuous and inf-sup stable in $V \times W$ with $\beta = (4T^2+1)^{-1/2}$, i.e.,*

$$\sup_{\mathbf{w}\in W\setminus\{\mathbf{0}\}} \frac{b(\mathbf{v},\mathbf{w})}{\|\mathbf{w}\|_W} \geq \beta \,\|\mathbf{v}\|_V\,, \qquad \mathbf{v} \in V\,.$$

Proof. The continuity follows from the upper bound $|b(\mathbf{v},\mathbf{w})| \leq \|\mathbf{v}\|_V\|\mathbf{w}\|_W$. To prove the inf-sup condition we first note that for all $\mathbf{v} \in \mathrm{C}^1(0,T;D(A))$ with $\mathbf{v}(0) = \mathbf{0}$ we have

$$\begin{aligned}
\|\mathbf{v}\|_W^2 &= \int_0^T \big(M\mathbf{v}(t),\mathbf{v}(t)\big)_{0,\Omega}\,\mathrm{d}t = \int_0^T \Big(\big(M\mathbf{v}(t),\mathbf{v}(t)\big)_{0,\Omega} - \big(M\mathbf{v}(0),\mathbf{v}(0)\big)_{0,\Omega}\Big)\,\mathrm{d}t \\
&= \int_0^T\int_0^t \partial_t\big(M\mathbf{v}(s),\mathbf{v}(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t = 2\int_0^T\int_0^t \big(M\partial_t\mathbf{v}(s),\mathbf{v}(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t \\
&\leq 2\int_0^T\int_0^t \big(M\partial_t\mathbf{v}(s) + A\mathbf{v}(s),\mathbf{v}(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t \\
&\leq 2\int_0^T\int_0^t \big(M^{-1}L\mathbf{v}(s),L\mathbf{v}(s)\big)_{0,\Omega}^{1/2}\big(M\mathbf{v}(s),\mathbf{v}(s)\big)_{0,\Omega}^{1/2}\,\mathrm{d}s\,\mathrm{d}t \leq 2T\,\|M^{-1}L\mathbf{v}\|_W\|\mathbf{v}\|_W\,.
\end{aligned}$$

This yields $\|\mathbf{v}\|_W \leq 2T\,\|M^{-1}L\mathbf{v}\|_W$ for $\mathbf{v} \in V$. Let $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ and take $\mathbf{w} = M^{-1}L\mathbf{v} \in W \setminus \{\mathbf{0}\}$, then

$$\sup_{\mathbf{w}\in W\setminus\{\mathbf{0}\}} \frac{b(\mathbf{v},\mathbf{w})}{\|\mathbf{w}\|_W} \geq \frac{b(\mathbf{v},M^{-1}L\mathbf{v})}{\|M^{-1}L\mathbf{v}\|_W} = \frac{(L\mathbf{v},M^{-1}L\mathbf{v})_{0,\Omega}}{\|M^{-1}L\mathbf{v}\|_W} = \|M^{-1}L\mathbf{v}\|_W \geq \frac{1}{\sqrt{4T^2+1}}\,\|\mathbf{v}\|_V\,,$$

where the final inequality follows from $\|\mathbf{v}\|_V^2 = \|\mathbf{v}\|_W^2 + \|M^{-1}L\mathbf{v}\|_W^2 \leq (4T^2+1)\|M^{-1}L\mathbf{v}\|_W^2$. □

The inf-sup stability ensures that the operator $L \in \mathcal{L}(V,W)$ is injective and that the range is closed. Thus, the operator is surjective by construction and the inverse L^{-1} is bounded in $\mathcal{L}(W,V)$. This yields directly the following result [Bra07, Thm. III.3.6].

Theorem 1. *For given $\mathbf{f} \in \mathrm{L}_2(Q)^J$ there exists a unique solution $\mathbf{u} \in V$ of*

$$(L\mathbf{u},\mathbf{w})_{0,Q} = (\mathbf{f},\mathbf{w})_{0,Q}\,, \qquad \mathbf{w} \in W \tag{2}$$

satisfying the a priori bound $\|\mathbf{u}\|_V \leq \sqrt{4T^2+1}\,\|M^{-1/2}\mathbf{f}\|_{0,Q}$.

3. A semi-discrete discontinuous Galerkin discretization in space

In this section we consider the semi-discrete evolution equation

$$M_h \partial_t \mathbf{u}_h(t) + A_h \mathbf{u}_h(t) = \mathbf{f}(t)\,, \qquad t \in (0,T)\,, \tag{3}$$

in a finite dimensional subspace $H_h \subset H$ associated to the mesh size h of the underlying mesh defined below. The discrete operator $A_h \in \mathcal{L}(H_h, H_h)$ will be constructed from a discontinuous Galerkin discretization. The discrete mass operator $M_h \in \mathcal{L}(H_h, H_h)$ is the Galerkin approximation of M defined by

$$(M_h \mathbf{v}_h, \mathbf{w}_h)_{0,\Omega} = (M \mathbf{v}_h, \mathbf{w}_h)_{0,\Omega} \qquad \mathbf{v}_h, \mathbf{w}_h \in H_h\,. \tag{4}$$

Note that the discrete mass operator M_h is represented by a block diagonal positive definite matrix.

We assume that Ω is a bounded polyhedral Lipschitz domain decomposed into a finite number of open elements $K \subset \Omega$ such that $\overline{\Omega} = \bigcup_{K\in\mathcal{K}} \overline{K}$. Let $\mathcal{F}_K$ be the set of faces of K, and for inner faces $f \in \mathcal{F}_K$ let K_f be the neighboring cell such that $f = \partial K \cap \partial K_f$, and let $\mathbf{n}_K$ be the outer unit normal vector on ∂K. The outer unit normal vector field on $\partial\Omega$ is denoted by $\mathbf{n}$.

Integration by parts of $A\mathbf{v} = \operatorname{div} \mathbf{F}(\mathbf{v})$ gives for smooth ansatz functions $\mathbf{v}$ and smooth test functions $\boldsymbol{\phi}_K$

$$(A\mathbf{v}, \boldsymbol{\phi}_K)_{0,K} = -(\mathbf{F}(\mathbf{v}), \nabla \boldsymbol{\phi}_K)_{0,K} + \sum_{f\in\mathcal{F}_K} (\mathbf{n}_K \cdot \mathbf{F}(\mathbf{v}), \boldsymbol{\phi}_K)_{0,f}\,.$$

This formulation is now the basis for the discretization. We select polynomial degrees p_K, and we define the local spaces $H_{h,K} = \mathbb{P}_{p_K}(K)^J$ and the global discontinuous Galerkin space $H_h = \{\mathbf{v}_h \in \mathrm{L}_2(\Omega)^J \colon \mathbf{v}_h|_K \in H_{h,K} \text{ for all } K \in \mathcal{K}\}$. For $\mathbf{v}_h \in H_h$ we define $\mathbf{v}_{h,K} = \mathbf{v}_h|_K \in H_{h,K}$ for the restriction to K.

We then define the discrete linear operator $A_h \in \mathcal{L}(H_h, H_h)$ for $\mathbf{v}_h \in H_h$ and $\boldsymbol{\phi}_{h,K} \in H_{h,K}$ by

$$(A_h \mathbf{v}_h, \boldsymbol{\phi}_{h,K})_{0,K} = -(\mathbf{F}(\mathbf{v}_{h,K}), \nabla \boldsymbol{\phi}_{h,K})_{0,K} + \sum_{f\in\mathcal{F}_K} \big(\mathbf{n}_K \cdot \mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h), \boldsymbol{\phi}_{h,K}\big)_{0,f}\,,$$

where $\mathbf{n}_K \cdot \mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h)$ is the upwind flux obtained from local solutions of Riemann problems, see [HPS+14, Sect. 2]. Again using integration by parts, we obtain

$$(A_h \mathbf{v}_h, \boldsymbol{\phi}_{h,K})_{0,K} = \big(\operatorname{div} \mathbf{F}(\mathbf{v}_{h,K}), \boldsymbol{\phi}_{h,K}\big)_{0,K} + \sum_{f\in\mathcal{F}_K} \big(\mathbf{n}_K \cdot (\mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h) - \mathbf{F}(\mathbf{v}_{h,K})), \boldsymbol{\phi}_{h,K}\big)_{0,f}\,. \tag{5}$$

On inner faces $f = \partial K \cap \partial K_f$ it is a consistency requirement that the difference $\mathbf{n}_K \cdot (\mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h) - \mathbf{F}(\mathbf{v}_{h,K}))$ only depends on $[\mathbf{v}_h]_{K,f} = \mathbf{v}_{h,K_f} - \mathbf{v}_{h,K}$, and that $\mathbf{n}_K \cdot (\mathbf{F}_K^{\mathrm{num}}(\mathbf{v}) - \mathbf{F}(\mathbf{v})) = 0$ on all faces $f \in \mathcal{F}_K$ for $\mathbf{v} \in D(A)$. In particular, this yields

$$(A\mathbf{v}, \boldsymbol{\phi}_h)_{0,\Omega} = (A_h \mathbf{v}, \boldsymbol{\phi}_h)_{0,\Omega}\,, \qquad \mathbf{v} \in D(A)\,,\ \boldsymbol{\phi}_h \in H_h\,, \tag{6}$$

and

$$\sum_{K\in\mathcal{K}} \big(\mathbf{n}_K \cdot \mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_{h,K}), \mathbf{v}\big)_{0,\partial K} = 0\,, \qquad \mathbf{v}_h \in H_h\,,\ \mathbf{v} \in D(A) \cap \mathrm{H}^1(\Omega)^J\,. \tag{7}$$

The upwind flux guarantees that the discrete operator is non-negative, i.e., $(A_h \mathbf{v}_h, \mathbf{v}_h)_{0,\Omega} \geq 0$ for $\mathbf{v}_h \in H_h$. For the examples in Sect. 2, the numerical upwind flux in homogeneous media is given as follows (see [HPS+14] for the explicit solution of Riemann problems in heterogeneous media).

Linear transport. We have $\mathbf{n}\cdot\mathbf{F}(u) = u\,\mathbf{n}\cdot\mathbf{q}$ and

$$\mathbf{n}_K\cdot\mathbf{F}_K^{\mathrm{num}}(u_h) - \mathbf{n}_K\cdot\mathbf{F}(u_{h,K}) = \frac{1}{2}\big(\mathbf{n}_K\cdot\mathbf{q} - |\mathbf{n}_K\cdot\mathbf{q}|\big)\,[u_h]_{K,f}\,,$$

with $[u_h]_{K,f} = -2u_{h,K}$ on $f\subset\Gamma_{\mathrm{in}}$, and $[u_h]_{K,f} = 0$ on $f\subset\partial\Omega\setminus\Gamma_{\mathrm{in}}$.

Acoustic waves. We have $\mathbf{n}\cdot\mathbf{F}(\mathbf{v},p) = \big(p\mathbf{n}, \mathbf{n}\cdot\mathbf{v}\big)^\top$ and

$$\mathbf{n}_K\cdot\mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h,p_h) - \mathbf{n}_K\cdot\mathbf{F}(\mathbf{v}_{h,K},p_{h,K}) = \frac{1}{2}\big([p_h]_{K,f} - [\mathbf{v}_h]_{K,f}\cdot\mathbf{n}_K\big)\begin{pmatrix}\mathbf{n}_K\\ -1\end{pmatrix}.$$

On Dirichlet boundary faces $f = \partial K\cap\partial\Omega$, we set $[p_h]_{K,f} = 2p_{h,K}$ and $[\mathbf{v}_h]_{K,f}\cdot\mathbf{n}_K = 0$.

Electro-magnetic waves. We have $\mathbf{n}\cdot\mathbf{F}(\mathbf{E},\mathbf{H}) = \big(-\mathbf{n}\times\mathbf{H}, \mathbf{n}\times\mathbf{E}\big)^\top$ and

$$\mathbf{n}_K\cdot\mathbf{F}_K^{\mathrm{num}}(\mathbf{E}_h,\mathbf{H}_h) - \mathbf{n}_K\cdot\mathbf{F}(\mathbf{E}_{h,K},\mathbf{H}_{h,K}) = \frac{1}{2}\begin{pmatrix}-\mathbf{n}_K\times[\mathbf{H}_h]_{K,f}\\ \mathbf{n}_K\times[\mathbf{E}_h]_{K,f}\end{pmatrix} + \begin{pmatrix}\frac{\sqrt{\varepsilon}}{\sqrt{\mu}}\mathbf{n}_K\times(\mathbf{n}_K\times[\mathbf{E}_h]_{K,f})\\ \frac{\sqrt{\mu}}{\sqrt{\varepsilon}}\mathbf{n}_K\times(\mathbf{n}_K\times[\mathbf{H}_h]_{K,f})\end{pmatrix}.$$

The perfect conducting boundary conditions on $f = \partial K\cap\partial\Omega$ are modelled by the (only virtual) definition of $\mathbf{n}_K\times\mathbf{E}_{h,K_f} = -\mathbf{n}_K\times\mathbf{E}_{h,K}$ and $\mathbf{n}_K\times\mathbf{H}_{h,K_f} = \mathbf{n}_f\times\mathbf{H}_{h,K}$, i.e., $\mathbf{n}_K\times[\mathbf{E}]_{K,f} = -2\mathbf{n}_K\times\mathbf{E}_{h,K}$ and $\mathbf{n}_K\times[\mathbf{H}]_{K,f} = \mathbf{0}$.

4. A Petrov–Galerkin space-time discretization

Let $\overline{Q} = \bigcup_{R\in\mathcal{R}}\overline{R}$ be a decomposition of the space-time cylinder into space-time cells $R = K\times I$ with $K\subset\Omega$ and $I = (t_-,t_+)\subset(0,T)$. For every R we choose local ansatz and test spaces $V_{h,R}, W_{h,R}\subset \mathrm{L}_2(R)^J$ with $W_{h,R}\subset\partial_t V_{h,R}$, and we define the global ansatz and test space

$$V_h = \Big\{\mathbf{v}_h\in\mathrm{H}^1(0,T;H)\colon \mathbf{v}_h(\mathbf{x},0) = \mathbf{0} \text{ for a.a. } \mathbf{x}\in\Omega \text{ and } \mathbf{v}_{h,R} = \mathbf{v}_h|_R\in V_{h,R}\Big\},$$
$$W_h = \Big\{\mathbf{w}_h\in\mathrm{L}_2(0,T;H)\colon \mathbf{w}_{h,R} = \mathbf{w}_h|_R\in W_{h,R}\Big\}.$$

By construction, functions in W_h are discontinuous in space and time, and functions in V_h are continuous in time, i.e., $\mathbf{v}_h(\mathbf{x},\cdot)$ is continuous on $[0,T]$ for a.a. $\mathbf{x}\in\Omega$.

In addition we aim for $\dim(V_h) = \dim(W_h)$, which restricts the choice of $V_{h,R}$. In the most simple case this can be achieved for a tensor product space-time discretization with a fixed mesh $\mathcal{K}$ in space and a time series $0 = t_0 < t_1 < \cdots < t_N = T$, i.e., $\mathcal{R} = \bigcup_{K\in\mathcal{K}}\bigcup_{n=1}^N K\times(t_{n-1},t_n)$. Then, we can select a discrete space H_h with $H_{h,K} = \mathbb{P}_p(K)^J$ independently of t, and in every time slice we define $W_{h,R} = \mathbb{P}_p(K)$ constant in time on $R = K\times(t_{n-1},t_n)$. This yields in this case piecewise linear approximations in time

$$V_h = \Big\{\mathbf{v}_h\in\mathrm{H}^1(0,T;H)\,:\,\mathbf{v}_h(\mathbf{x},0) = \mathbf{0}\,,\ \mathbf{v}_h(\mathbf{x},t_n)\in H_h \text{ for a.a. } \mathbf{x}\in\Omega \text{ and } n = 1,\dots,N\,,\text{ and}$$
$$\mathbf{v}_h(\mathbf{x},t) = \frac{t_n - t}{t_n - t_{n-1}}\mathbf{v}_h(\mathbf{x},t_{n-1}) + \frac{t - t_{n-1}}{t_n - t_{n-1}}\mathbf{v}_h(\mathbf{x},t_n) \text{ for } t\in(t_{n-1},t_n)\Big\}.$$

In the general case, we select locally in space and time polynomial degrees p_R and q_R in R, and we set for the local test space $W_{h,R} = \big(\mathbb{P}_{p_R}(K)\otimes\mathbb{P}_{q_R-1}\big)^J$. Then we define for $R\in\mathcal{R}$

$$V_{h,R} = \Big\{\mathbf{v}_{h,R}\in\mathrm{L}_2(R)^J\,:\,\mathbf{v}_{h,R}(\mathbf{x},t) = \frac{t_+ - t}{t_+ - t_-}\mathbf{v}_h(\mathbf{x},t_-) + \frac{t - t_-}{t_+ - t_-}\mathbf{w}_{h,R}(\mathbf{x},t)\,,$$
$$\mathbf{v}_h\in V_h|_{[0,t_-]}\,,\ \mathbf{w}_{h,R}\in W_{h,R}\,,\ (\mathbf{x},t)\in R = K\times(t_-,t_+)\Big\}.$$

This yields $\mathbf{v}_{h,R}(\mathbf{x},\cdot)\in\mathbb{P}_{q_R}^J$ for $\mathbf{v}_{h,R}\in V_{h,R}$ and $(\mathbf{x},\cdot)\in R$.

The discontinuous Galerkin operator in space is extended to the space-time setting defining $A_h\mathbf{v}_h \in W_h$ by

$$\big(A_h\mathbf{v}_h, \mathbf{w}_h\big)_{0,Q} = \sum_{R=K\times I} \Big(\big(\operatorname{div} \mathbf{F}(\mathbf{v}_{h,R}), \mathbf{w}_{h,R}\big)_{0,R} + \sum_{f\in\mathcal{F}_K} \big(\mathbf{n}_K \cdot (\mathbf{F}_K^{\mathrm{num}}(\mathbf{v}_h) - \mathbf{F}(\mathbf{v}_{h,R})), \mathbf{w}_{h,R}\big)_{0,f\times I} \Big) \tag{8}$$

for $\mathbf{v}_h \in V_h$ and $\mathbf{w}_h \in W_h$. Moreover, we define the discrete space-time operator $L_h = M_h\partial_t + A_h \in \mathcal{L}(V_h, W_h)$ and the discrete bilinear form $b_h(\cdot,\cdot) = \big(L_h\cdot,\cdot\big)_{0,Q}$.

In order to show that a solution to our Petrov–Galerkin scheme exists, we check the inf-sup stability of the discrete bilinear form $b_h(\cdot,\cdot)$. For this purpose we introduce the weighted L_2-projection $\Pi_h\colon W \to W_h$ defined by

$$\big(\Pi_h\mathbf{v}, \mathbf{w}_h\big)_W = \big(\mathbf{v}, \mathbf{w}_h\big)_W\,, \qquad \mathbf{w}_h \in W_h\,.$$

The restriction $\Pi_{h,R} = \Pi_h|_R$ can be evaluated locally, since W_h is discontinuous in $\mathcal{R}$. Moreover, (4) yields $M_h(W_h) = W_h$, $\Pi_h M_h = M_h\Pi_h$ and $\Pi_h M_h^{-1} = M_h^{-1}\Pi_h$. The stability and convergence analysis is considered with respect to the discrete norm

$$\|\mathbf{v}_h\|_{V_h}^2 = \|\mathbf{v}_h\|_W^2 + \|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W^2\,.$$

Lemma 2. *Assume that $\Pi_h\partial_t\mathbf{v}_h = \partial_t\mathbf{v}_h$ for $\mathbf{v}_h \in V_h$ and $(A_h\mathbf{v}_h(t), \Pi_h\mathbf{v}_h(t))_{0,Q} \geq 0$ for almost all $t \in (0,T)$. Then, the bilinear form $b_h(\cdot,\cdot)$ is bounded and inf-sup stable in $V_h \times W_h$ with $\beta = 1/\sqrt{1+4T^2}$, i.e.,*

$$\sup_{\mathbf{w}_h\in W_h} \frac{b_h(\mathbf{v}_h, \mathbf{w}_h)}{\|\mathbf{w}_h\|_W} \geq \beta\,\|\mathbf{v}_h\|_{V_h}\,, \qquad \mathbf{v}_h \in V_h\,.$$

Proof. For $\mathbf{v}_h \in V_h$ and $\mathbf{w}_h \in W_h$ we have

$$b(\mathbf{v}_h, \mathbf{w}_h) = \big(L_h\mathbf{v}_h, \mathbf{w}_h\big)_{0,Q} = \big(M_h\Pi_h M_h^{-1} L_h\mathbf{v}_h, \mathbf{w}_h\big)_{0,Q} \leq \|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W \|\mathbf{w}_h\|_W \leq \|\mathbf{v}_h\|_V \|\mathbf{w}_h\|_W\,,$$

which shows that $b_h(\cdot,\cdot)$ is bounded in $V_h \times W_h$. Using the assumptions on Π_h we obtain for $\mathbf{v}_h \in V_h$

$$\big(M_h\partial_t\mathbf{v}_h, \mathbf{v}_h\big)_{0,\Omega} = \big(M_h\partial_t\mathbf{v}_h, \Pi_h\mathbf{v}_h\big)_{0,\Omega} \leq \big(M_h\partial_t\mathbf{v}_h + A_h\mathbf{v}_h, \Pi_h\mathbf{v}_h\big)_{0,\Omega} = \big(\Pi_h L_h\mathbf{v}_h, \mathbf{v}_h\big)_{0,\Omega}$$

and thus, transferring the proof of Lem. 1 to the discrete setting,

$$\begin{aligned}
\|\mathbf{v}_h\|_W^2 &= \int_0^T \big(M_h\mathbf{v}_h(t), \mathbf{v}_h(t)\big)_{0,\Omega}\,\mathrm{d}t = \int_0^T \Big(\big(M_h\mathbf{v}_h(t), \mathbf{v}_h(t)\big)_{0,\Omega} - \big(M_h\mathbf{v}_h(0), \mathbf{v}_h(0)\big)_{0,\Omega} \Big)\,\mathrm{d}t \\
&= \int_0^T \int_0^t \partial_t \big(M_h\mathbf{v}_h(s), \mathbf{v}_h(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t = 2\int_0^T \int_0^t \big(M_h\partial_t\mathbf{v}_h(s), \mathbf{v}_h(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t \\
&\leq 2\int_0^T \int_0^t \big(\Pi_h L_h\mathbf{v}_h(s), \mathbf{v}_h(s)\big)_{0,\Omega}\,\mathrm{d}s\,\mathrm{d}t \leq 2\int_0^T \int_0^T \big|\big(M_h\Pi_h M_h^{-1} L_h\mathbf{v}_h(s), \mathbf{v}_h(s)\big)_{0,\Omega}\big|\,\mathrm{d}s\,\mathrm{d}t \\
&= 2\int_0^T \big|\big(\Pi_h M_h^{-1} L_h\mathbf{v}_h, \mathbf{v}_h\big)_W\big|\,\mathrm{d}t \leq 2T\|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W \|\mathbf{v}_h\|_W\,.
\end{aligned}$$

This yields $\|\mathbf{v}_h\|_W \leq 2T\|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W$ and thus $\|\mathbf{v}_h\|_{V_h} \leq \sqrt{1+4T^2}\|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W$, which implies the inf-sup stability using $b_h(\mathbf{v}_h, \mathbf{w}_h) = (L_h\mathbf{v}_h, \mathbf{w}_h)_{0,Q} = (\Pi_h M_h^{-1} L_h\mathbf{v}_h, \mathbf{w}_h)_W$ and inserting $\mathbf{w}_h = \Pi_h M_h^{-1} L_h\mathbf{v}_h$

$$\sup_{\mathbf{w}_h\in W_h\setminus\{\mathbf{0}\}} \frac{b_h(\mathbf{v}_h, \mathbf{w}_h)}{\|\mathbf{w}_h\|_W} = \sup_{\mathbf{w}_h\in W_h} \frac{(\Pi_h M_h^{-1} L_h\mathbf{v}_h, \mathbf{w}_h)_W}{\|\mathbf{w}_h\|_W} \geq \|\Pi_h M_h^{-1} L_h\mathbf{v}_h\|_W \geq \frac{1}{\sqrt{1+4T^2}}\|\mathbf{v}_h\|_{V_h}\,.$$

□

As in Thm. 1, this shows the existence of a unique discrete Petrov–Galerkin solution (provided that the assumptions in Lem. 2 are satisfied).

Theorem 2. *For given* $\mathbf{f} \in \mathrm{L}_2(Q)^J$ *there exists a unique solution* $\mathbf{u}_h \in V_h$ *of*

$$(L_h \mathbf{u}_h, \mathbf{w}_h)_{0,Q} = (\mathbf{f}, \mathbf{w}_h)_{0,Q}\,, \qquad \mathbf{w}_h \in W_h\,, \tag{9}$$

satisfying the a priori bound $\|\mathbf{u}_h\|_{V_h} \le \sqrt{4T^2+1}\|M_h^{-1}\Pi_h \mathbf{f}\|_W$.

The convergence will be analyzed with respect to the discrete norm $\|\cdot\|_{V_h}$. For $\mathbf{v} \in V$ the consistency of the numerical flux in (8) yields $\big(A_h\mathbf{v}, \mathbf{w}_h\big)_{0,Q} = \big(\operatorname{div}\mathbf{F}(\mathbf{v}), \mathbf{w}_h\big)_{0,Q}$ so that $A_h\mathbf{v} = \Pi_h \operatorname{div}\mathbf{F}(\mathbf{v})$. This shows that A_h and thus also $\|\cdot\|_{V_h}$ can be evaluated in $V + V_h$ and that $b_h(\cdot,\cdot)$ is continuous with respect to this extension.

Theorem 3. *Let* $\mathbf{u} \in V$ *be the solution of* (2) *and* $\mathbf{u}_h \in V_h$ *its approximation solving* (9)*. Then, we have*

$$\|\mathbf{u} - \mathbf{u}_h\|_{V_h} \le (1+\beta^{-1}) \inf_{\mathbf{v}_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_{V_h}\,.$$

If in addition the solution is sufficiently smooth, we obtain the a priori error estimate

$$\|\mathbf{u} - \mathbf{u}_h\|_{V_h} \le C\big(\triangle t^q + \triangle x^p\big)\Big(\|\partial_t^{q+1}\mathbf{u}\|_{0,Q} + \|\mathrm{D}^{p+1}\mathbf{u}\|_{0,Q}\Big)$$

for $\triangle t$, $\triangle x$ *and* $p, q \ge 1$ *with* $\triangle t \le t_+ - t_-$, $\triangle x \le \operatorname{diam} K$, $p \le p_R$ *and* $q \le q_R$ *for all* $R = K \times (t_-, t_+)$.

Proof. The consistency (6) of the discontinuous Galerkin method yields $(A_h\mathbf{u}(t), \mathbf{w}_h(t))_{0,\Omega} = (A\mathbf{u}(t), \mathbf{w}_h(t))_{0,\Omega}$ and thus also consistency of the Petrov–Galerkin setting, i.e., $b_h(\mathbf{u}, \mathbf{w}_h) = b(\mathbf{u}, \mathbf{w}_h) = (\mathbf{f}, \mathbf{w}_h)_{0,Q} = b_h(\mathbf{u}_h, \mathbf{w}_h)$. This gives for all $\mathbf{v}_h \in V_h$ and $\mathbf{w}_h \in W_h$

$$b_h(\mathbf{v}_h - \mathbf{u}_h, \mathbf{w}_h) = b_h(\mathbf{v}_h - \mathbf{u}, \mathbf{w}_h) \le \|\mathbf{v}_h - \mathbf{u}\|_{V_h}\|\mathbf{w}_h\|_W$$

and thus

$$\begin{aligned}\|\mathbf{u} - \mathbf{u}_h\|_{V_h} &\le \|\mathbf{u} - \mathbf{v}_h\|_{V_h} + \|\mathbf{v}_h - \mathbf{u}_h\|_{V_h} \\ &\le \|\mathbf{u} - \mathbf{v}_h\|_{V_h} + \beta^{-1} \sup_{\mathbf{w}_h \in W_h \setminus \{\mathbf{0}\}} \frac{b_h(\mathbf{v}_h - \mathbf{u}_h, \mathbf{w}_h)}{\|\mathbf{w}_h\|_W} \le (1+\beta^{-1})\,\|\mathbf{u} - \mathbf{v}_h\|_{V_h}\,.\end{aligned}$$

Now we assume that the solution is regular satisfying $\mathbf{u} \in \mathrm{H}^{q+1}(0,T;\mathrm{L}_2(\Omega)^J) \cap \mathrm{L}_2(0,T;\mathrm{H}^{p+1}(\Omega)^J)$. We have by consistency $A_h\mathbf{v}_h = A\mathbf{v}_h$ for all $\mathbf{v}_h \in V_h \cap \mathrm{H}^1(\Omega)^J$, so that the error estimate yields

$$\|\mathbf{u} - \mathbf{u}_h\|_{V_h} \le (1+\beta^{-1}) \inf_{\mathbf{v}_h \in V_h \cap \mathrm{H}^1(\Omega)^J} \|\mathbf{u} - \mathbf{v}_h\|_{V_h} \le C\Big(\|\partial_t(\mathbf{u} - I_h\mathbf{u})\|_{0,Q} + \|\mathrm{D}(\mathbf{u} - I_h\mathbf{u})\|_{0,Q}\Big)\,,$$

where $I_h \colon V \to V_h \cap \mathrm{H}^1(\Omega)^J$ is a suitable Clément-type interpolation operator. By standard assumptions on the right hand side and the mesh regularity we obtain a bound depending on $\triangle t$ in time and $\triangle x$ in space. □

We check the assumptions of Lem. 2 only for the special case of a tensor product discretization with $p_R \equiv p$ and $q_R \equiv 1$ for all R. Note that for this case the Petrov–Galerkin method in time is equivalent to the implicit midpoint rule, see also [BR99]. For $\mathbf{v}_h \in V_h$ we define $\mathbf{v}_{h,n} = \mathbf{v}_h(\cdot, t_n)$. This yields for $t \in (t_{n-1}, t_n)$

$$\mathbf{v}_h(\mathbf{x},t) = \frac{t_n - t}{t_n - t_{n-1}} \mathbf{v}_{h,n-1}(\mathbf{x}) + \frac{t - t_{n-1}}{t_n - t_{n-1}} \mathbf{v}_{h,n}(\mathbf{x}) \,, \qquad \partial_t \mathbf{v}_n(\mathbf{x},t) = \frac{1}{t_n - t_{n-1}} \Big(\mathbf{v}_{h,n}(\mathbf{x}) - \mathbf{v}_{h,n-1}(\mathbf{x}) \Big)$$

and thus $\partial_t \mathbf{v}_h = \Pi_h \partial_t \mathbf{v}_h \in W_h$ and $\Pi_h \mathbf{v}_h(\mathbf{x},t) = \frac{1}{2}(\mathbf{v}_{h,n-1} + \mathbf{v}_{h,n})(\mathbf{x})$. Together with $\big(A_h \mathbf{v}_h, \mathbf{v}_h\big)_{0,\Omega} \geq 0$ this implies

$$\begin{aligned} \int_{t_{n-1}}^{t_n} \big(A_h \mathbf{v}_h(s), \Pi_h \mathbf{v}_h(s)\big)_{0,\Omega} \,\mathrm{d}s &= \frac{1}{2(t_n - t_{n-1})} \int_{t_{n-1}}^{t_n} \big((t_n - s) A_h \mathbf{v}_{h,n-1} + (s - t_{n-1}) A_h \mathbf{v}_{h,n} \,,\, \mathbf{v}_{h,n-1} + \mathbf{v}_{h,n}\big)_{0,\Omega} \,\mathrm{d}s \\ &= \frac{t_n - t_{n-1}}{4} \big(A_h(\mathbf{v}_{h,n-1} + \mathbf{v}_{h,n}), \mathbf{v}_{h,n-1} + \mathbf{v}_{h,n}\big)_{0,\Omega} \geq 0 \,. \end{aligned}$$

Its an open task to extend this to more general cases.

5. Duality based goal-oriented error estimation

In order to develop an adaptive strategy for the selection of the local polynomial degrees p_R, q_R we derive an error indicator with respect to a given linear goal functional $E \in W'$. Following the framework in [BR03], we define the adjoint problem and solve the dual problem. Then, the error is estimated in terms of the local residual and the dual weight.

The adjoint operator $L^* = -\partial_t + A^*$ in space and time is defined on the the adjoint Hilbert space

$$V^* = \big\{ \mathbf{w} \in W : \text{there exists } \mathbf{g} \in W \text{ such that } (L\mathbf{v}, \mathbf{w})_{0,Q} = (\mathbf{v}, \mathbf{g})_{0,Q} \text{ for all } \mathbf{v} \in V \big\}$$

and is characterized by

$$(\mathbf{v}, L^* \mathbf{w})_{0,Q} = (L\mathbf{v}, \mathbf{w})_{0,Q} \,, \qquad \mathbf{v} \in V \,,\ \mathbf{w} \in V^* \,.$$

Note that we have $\mathbf{w}(T) = \mathbf{0}$ for $\mathbf{w} \in V^*$, so that the adjoint space-time problem can be solved backward in time. In case of the linear hyperbolic problems discussed in Sect. 2 it holds

$$(\operatorname{div} \mathbf{F}(\mathbf{v}), \mathbf{w})_{0,\Omega} = -(\mathbf{v}, \operatorname{div} \mathbf{F}(\mathbf{w}))_{0,\Omega} \,, \qquad \mathbf{v} \in V \,,\ \mathbf{w} \in V^* \,, \tag{10}$$

so that we have $A^* = -A$ on $V \cap V^*$.

For the evaluation of the error functional E we introduce the dual solution $\mathbf{u}^* \in V^*$ with

$$(\mathbf{w}, L^* \mathbf{u}^*)_{0,Q} = \langle E, \mathbf{w} \rangle \,, \qquad \mathbf{w} \in W \,.$$

Let $\mathbf{u} \in V$ be the solution of (2), and $\mathbf{u}_h \in V_h$ its approximation solving (9). Now we derive an estimate for the error functional in the case that the dual solution is sufficiently smooth such that $\mathbf{u}^*(\cdot,t)|_f \in \mathrm{L}_2(f)^J$ for all faces $f \in \mathcal{F}_h$ and a.a. $t \in (0,T)$. Inserting the consistency of the numerical flux (7) and using (10) yields for all $\mathbf{w}_h \in W_h$

$$
\begin{aligned}
\langle E, \mathbf{u} - \mathbf{u}_h \rangle &= \big(\mathbf{u} - \mathbf{u}_h, -M\partial_t \mathbf{u}^* - \operatorname{div} \mathbf{F}(\mathbf{u}^*)\big)_{0,Q} \\
&= \big(\mathbf{u}, -M\partial_t \mathbf{u}^* - \operatorname{div} \mathbf{F}(\mathbf{u}^*)\big)_{0,Q} - \big(\mathbf{u}_h, -M\partial_t \mathbf{u}^* - \operatorname{div} \mathbf{F}(\mathbf{u}^*)\big)_{0,Q} \\
&= \big(M\partial_t \mathbf{u} + \operatorname{div} \mathbf{F}(\mathbf{u}), \mathbf{u}^*\big)_{0,Q} - \big(\mathbf{u}, \mathbf{n} \cdot \mathbf{F}(\mathbf{u}^*)\big)_{0,\partial Q} \\
&\quad - \sum_{R \in \mathcal{R}} \Big(\big(M\partial_t \mathbf{u}_h + \operatorname{div} \mathbf{F}(\mathbf{u}_h), \mathbf{u}^*\big)_{0,R} - \big(\mathbf{u}_h, \mathbf{n}_R \cdot \mathbf{F}(\mathbf{u}^*)\big)_{0,\partial R} \Big) \\
&= \big(\mathbf{f}, \mathbf{u}^*\big)_{0,Q} - \sum_{R = K \times I \in \mathcal{R}} \Big(\big(M\partial_t \mathbf{u}_h + \operatorname{div} \mathbf{F}(\mathbf{u}_h), \mathbf{u}^*\big)_{0,R} - \big(\mathbf{u}_h, \mathbf{n}_K \cdot \mathbf{F}(\mathbf{u}^*)\big)_{0,\partial K \times I} \Big) \\
&= \sum_{R = K \times I \in \mathcal{R}} \Big(\big(\mathbf{f} - M\partial_t \mathbf{u}_h - \operatorname{div} \mathbf{F}(\mathbf{u}_h), \mathbf{u}^*\big)_{0,R} - \big(\mathbf{n}_K \cdot \mathbf{F}(\mathbf{u}_h), \mathbf{u}^*\big)_{0,\partial K \times I} \Big) \\
&= \sum_{R = K \times I \in \mathcal{R}} \Big(\big(\mathbf{f} - M\partial_t \mathbf{u}_h - \operatorname{div} \mathbf{F}(\mathbf{u}_h), \mathbf{u}^*\big)_{0,R} - \big(\mathbf{n}_K \cdot (\mathbf{F}(\mathbf{u}_h) - \mathbf{F}^{\mathrm{num}}(\mathbf{u}_h)), \mathbf{u}^*\big)_{0,\partial K \times I} \Big) \\
&= \sum_{R = K \times I \in \mathcal{R}} \Big(\big(\mathbf{f} - M\partial_t \mathbf{u}_h - \operatorname{div} \mathbf{F}(\mathbf{u}_h), \mathbf{u}^* - \mathbf{w}_h\big)_{0,R} - \big(\mathbf{n}_K \cdot (\mathbf{F}(\mathbf{u}_h) - \mathbf{F}^{\mathrm{num}}(\mathbf{u}_h)), \mathbf{u}^* - \mathbf{w}_h\big)_{0,\partial K \times I} \Big)
\end{aligned}
$$

and thus, inserting some projection $\mathbf{w}_h = \Pi_h \mathbf{u}^*$,

$$
\begin{aligned}
\big|\langle E, \mathbf{u} - \mathbf{u}_h \rangle\big| \le \sum_{R = K \times I \in \mathcal{R}} \Big(& \big\| \rho\partial_t \mathbf{u}_h + \operatorname{div} \mathbf{F}(\mathbf{u}_h) - \mathbf{f} \big\|_{0,R} \big\| \mathbf{u}^* - \Pi_h \mathbf{u}^* \|_{0,R} \\
& + \big\| \mathbf{n}_K \cdot (\mathbf{F}(\mathbf{u}_R) - \mathbf{F}^*(\mathbf{u}_h)) \big\|_{0,\partial K \times I} \big\| \mathbf{u}^* - \Pi_h \mathbf{u}^* \big\|_{0,\partial K \times I} \Big) .
\end{aligned} \tag{11}
$$

In the applications, the projection error of the dual solution $\mathbf{u}^* - \Pi_h \mathbf{u}^*$ is estimated, e.g., by $\mathbf{u}_h^* - I_h \mathbf{u}_h^*$, where $\mathbf{u}_h^* \in W_h$ is a numerical approximation of the dual solution given by

$$
b_h(\mathbf{v}_h, \mathbf{u}_h^*) = \langle E, \mathbf{v}_h \rangle , \qquad \mathbf{v}_h \in V_h
$$

(using the transposed finite element matrix) and I_h is a higher-order recovery operator (or a lower order interpolation operator). Then, the error bound (11) is estimated by $\sum_{R \in \mathcal{R}} \eta_R$ with

$$
\eta_R = \|\mathbf{f} - L_h \mathbf{u}_h\|_{0,R} \|\mathbf{u}^* - \Pi_h \mathbf{u}^*\|_{0,R} + \|\mathbf{n}_K \cdot (\mathbf{F}(\mathbf{u}_h) - \mathbf{F}^*(\mathbf{u}_h))\|_{0,\partial K \times I} \|\mathbf{u}^* - \Pi_h \mathbf{u}^*\|_{0,\partial K \times I} .
$$

Remark 4. *The error indicator construction extends to nonlinear goal functionals $E \in \mathrm{C}^1(W)$. Then, the dual solution $\mathbf{u}^* \in V^*$ depend on the primal solution, i.e.,*

$$
(\mathbf{w}, L^* \mathbf{u}^*)_{0,Q} = \langle E'(\mathbf{u}), \mathbf{w} \rangle , \quad \mathbf{w} \in W .
$$

The estimate (11) *applies also to $\big|E(\mathbf{u}) - E(\mathbf{u}_h)\big|$, since we have*

$$
E(\mathbf{u}) - E(\mathbf{u}_h) = \int_0^1 \big\langle E'(\mathbf{u}_h + s(\mathbf{u} - \mathbf{u}_h)), \mathbf{u} - \mathbf{u}_h \big\rangle \, \mathrm{d}s \approx \langle E'(\mathbf{u}_h), \mathbf{u} - \mathbf{u}_h \rangle
$$

up to higher order in $\|\mathbf{u} - \mathbf{u}_h\|_{0,Q}$ [HR03].

6. Space-time multilevel preconditioner

In this section we address the numerical aspects in particular solution methods for the discrete hyperbolic space-time problem. First we describe the realization of our discretization using nodal basis functions in space and time, and then a multilevel preconditioner is introduced, and it is tested for different settings to derive a suitable solution strategy.

Nodal Discretisation. Here we consider the case of a tensor product space-time mesh $\mathcal{R} = \bigcup_{n=1}^{N} \mathcal{R}^n$ with time slices $\mathcal{R}^n = \bigcup_{K\in\mathcal{K}} K \times (t_{n-1}, t_n)$ and variable polynomial degrees p_R, q_R in every space-time cell R. Let $\{\boldsymbol{\psi}^n_{R,j}\}_{j=1,\dots,\dim W_{h,R}}$ be a basis of $W_{h,R}$ and define $W^n_h = \operatorname{span}\Big\{ \bigcup_{R\in\mathcal{R}^n} \bigcup_{j=1}^{\dim W_{h,R}} \boldsymbol{\psi}^n_{R,j} \Big\}$. Then, $\mathbf{v}_h \in V_h$ is represented by

$$\mathbf{v}_h(\mathbf{x},t) = \frac{t_n - t}{t_n - t_{n-1}} \mathbf{w}_h^{n-1}(\mathbf{x}, t_{n-1}) + \frac{t - t_{n-1}}{t_n - t_{n-1}} \mathbf{w}_h^n(\mathbf{x},t) \quad \text{for} \quad (\mathbf{x},t) \in K \times (t_{n-1}, t_n)$$

with $\mathbf{w}_h^0 = \mathbf{0}$ and $\mathbf{w}_h^n \in W_h^n$, $n = 1, \dots, N$. The corresponding coefficient vector is denoted by $\underline{v} = (\underline{v}^1, \dots, \underline{v}^N)^\top$, where $\underline{v}^n \in \mathbb{R}^{\dim W_h^n}$ is the coefficient vector of $\mathbf{w}_h^n = \sum_{R\in\mathcal{R}^n} \sum_{j=1}^{\dim W_{h,R}} \underline{v}_{R,j} \boldsymbol{\psi}^n_{R,j}$. With respect to this basis, the discrete space-time system (2) has the matrix representation $\underline{L}\,\underline{u} = \underline{f}$ with the block matrix

$$\underline{L} = \begin{pmatrix} \underline{D}^1 & & & \\ \underline{C}^1 & \underline{D}^2 & & \\ & \ddots & \ddots & \\ & & \underline{C}^{N-1} & \underline{D}^N \end{pmatrix}$$

with matrix entries

$$\underline{D}^n_{R',k,R,j} = \int_{t_{n-1}}^{t_n} \int_\Omega L_h\Big(\frac{t - t_{n-1}}{t_n - t_{n-1}} \boldsymbol{\psi}^n_{R,j}(\mathbf{x},t) \Big) \boldsymbol{\psi}^n_{R',k}(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\mathrm{d}t\,,$$
$$\underline{C}^n_{R',k,R,j} = \int_{t_{n-1}}^{t_n} \int_\Omega L_h\Big(\frac{t_n - t}{t_n - t_{n-1}} \boldsymbol{\psi}^n_{R,j}(\mathbf{x},t_{n-1}) \Big) \boldsymbol{\psi}^n_{R',k}(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\mathrm{d}t$$

and the right-hand side $\underline{f} = (\underline{f}^1, \dots, \underline{u}^N)$ with $\underline{f}^n_{j,R} = (\mathbf{f}, \boldsymbol{\psi}^n_{R,j})_{0,R}$. Sequentially, this system can be solved by a block-Gauss–Seidel method (corresponding to implicit time integration)

$$\underline{D}^1 \underline{u}^1 = \underline{f}^1\,, \qquad \underline{D}^2 \underline{u}^2 = \underline{f}^2 - \underline{C}^1 \underline{u}^1\,, \dots, \qquad \underline{D}^N \underline{u}^N = \underline{f}^N - \underline{C}^{N-1} \underline{u}^{N-1}\,,$$

provided that $\underline{D}^n$ can be inverted efficiently. In parallel, this requires a distribution only in space (see Fig. 1). Here, we discuss parallel multilevel preconditioners with a distribution of the full space-time mesh, cf. Fig. 2.

Multilevel methods. For space-time multilevel preconditioners we consider hierarchies in space and time. Therefore, let $\mathcal{R}_{0,0}$ be the coarse space-time mesh, and let $\mathcal{R}_{l,k}$ be the discretization obtained by $l = 1, \dots, l_{\max}$ uniform refinements in space and $k = 1, \dots, k_{\max}$ refinements in time. Let $V_{l,k}$ be the approximation spaces on $\mathcal{R}_{l,k}$ with fixed polynomial degrees $p_R \equiv p$ and $q_R \equiv q$. Let $\underline{L}_{l,k}$ be the corresponding matrix representations of the discrete operator L_h in $V_{l,k}$.

The multilevel preconditioner combines smoothing operations on different levels and requires transfer matrices between the levels. Since the spaces are nested, we can define prolongation matrices $\underline{P}^{l,k}_{l-1,k}$ and $\underline{P}^{l,k}_{l,k-1}$ representing the natural injections $V_{l-1,k} \subset V_{l,k}$ in space and $V_{l,k-1} \subset V_{l,k}$ in time. Correspondingly, the restriction matrices $\underline{R}^{l,k}_{l-1,k}$ and $\underline{R}^{l,k}_{l,k-1}$ represent the L_2 projections of the test spaces $W_{l,k} \supset W_{l-1,k}$ and $W_{l,k} \supset W_{l,k-1}$.

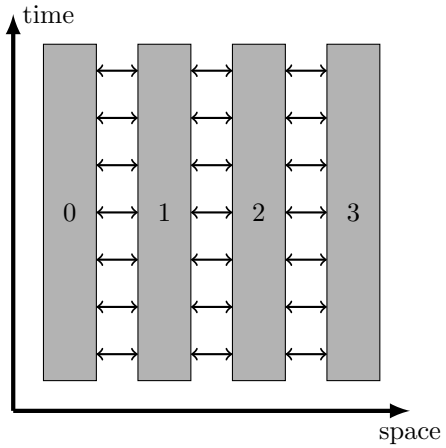


FIGURE 1. Spatial distribution of mesh cells to 4 processes and required communication (arrows).

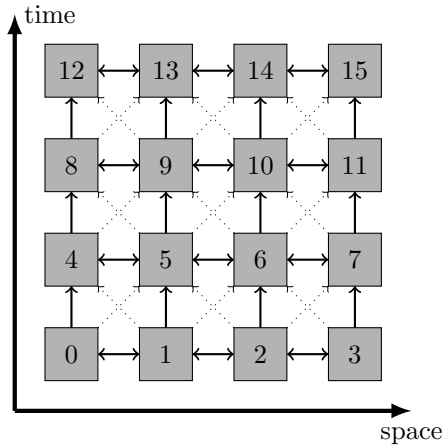


FIGURE 2. Space-time distribution of mesh cells to 16 processes and required communication (arrows).

For the smoothing operations on level (l,k) we consider the block-Jacobi preconditioner or the block-Gauss–Seidel preconditioner (where all components corresponding to a space-time cell R build blocks)

$$\underline{B}^{\mathrm{J}}_{l,k} = \theta_{l,k}\,\mathrm{block\_diag}(\underline{L}_{l,k})^{-1}\,, \qquad \underline{B}^{\mathrm{GS}}_{l,k} = \theta_{l,k}\left(\mathrm{block\_lower}(\underline{L}_{l,k}) + \mathrm{block\_diag}(\underline{L}_{l,k})\right)^{-1}$$

with damping parameter $\theta_{l,k}$. The corresponding iteration matrices are given by $\underline{S}^{\mathrm{J}}_{l,k} = \mathrm{id}_{l,k} - \underline{B}^{\mathrm{J}}_{l,k}\underline{L}_{l,k}$ and $\underline{S}^{\mathrm{GS}}_{l,k} = \mathrm{id}_{l,k} - \underline{B}^{\mathrm{GS}}_{l,k}\underline{L}_{l,k}$, and the number of pre- and postsmoothing steps are denoted by $\nu^{\mathrm{pre}}_{l,k}$ and $\nu^{\mathrm{post}}_{l,k}$.

Now, the multilevel preconditioner $\underline{B}^{\mathrm{ML}}_{l,k}$ is defined recursively. On the coarse level, we use a parallel direct linear solver $\underline{B}^{\mathrm{ML}}_{0,0} = \left(\underline{L}_{0,0}\right)^{-1}$. Then, we have two options: restricting in time defines $\underline{B}^{\mathrm{ML}}_{l,k}$ by

$$\mathrm{id}_{l,k} - \underline{B}^{\mathrm{ML}}_{l,k}\underline{L}_{l,k} = \left(\mathrm{id}_{l,k} - \underline{B}^{\mathrm{J}}_{l,k}\underline{L}_{l,k}\right)^{\nu^{\mathrm{pre}}_{l,k}} \left(\mathrm{id}_{l,k} - \underline{P}^{l,k}_{l,k-1}\underline{B}^{\mathrm{ML}}_{l,k-1}\underline{R}^{l,k}_{l,k-1}\underline{L}_{l,k}\right)\left(\mathrm{id}_{l,k} - \underline{B}^{\mathrm{J}}_{l,k}\underline{L}_{l,k}\right)^{\nu^{\mathrm{post}}_{l,k}}$$

with Jacobi smoothing (cf. Fig. 3), and restricting in space yields

$$\mathrm{id}_{l,k} - \underline{B}^{\mathrm{ML}}_{l,k}\underline{L}_{l,k} = \left(\mathrm{id}_{l,k} - \underline{B}^{\mathrm{GS}}_{l,k}\underline{L}_{l,k}\right)^{\nu^{\mathrm{pre}}_{l,k}} \left(\mathrm{id}_{l,k} - \underline{P}^{l,k}_{l,k-1}\underline{B}^{\mathrm{ML}}_{l-1,k}\underline{R}^{l,k}_{l,k-1}\underline{L}_{l,k}\right)\left(\mathrm{id}_{l,k} - \underline{B}^{\mathrm{GS}}_{l,k}\underline{L}_{l,k}\right)^{\nu^{\mathrm{post}}_{l,k}}$$

with Gauss–Seidel smoothing, cf. Fig. 4 for an illustration of the two options and Alg. 1 for the recursive realization of the multilevel preconditioner.

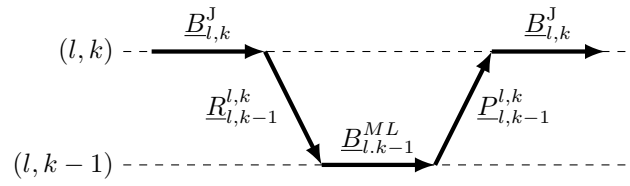


FIGURE 3. Two level in time coarsening strategy.

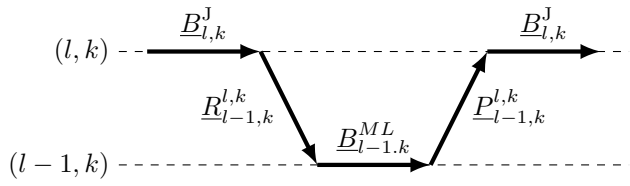


FIGURE 4. Two level in space coarsening strategy.

Algorithm 1 Multilevel preconditioner $\underline{c}_{l,k} = \underline{B}^{\mathrm{ML}}_{l,k}\underline{r}_{l,k}$ with smoother $\underline{B}^{\mathrm{SM}}_{l,k} = \underline{B}^{\mathrm{J}}_{l,k}$ or $\underline{B}^{\mathrm{GS}}_{l,k}$

$\underline{c}_{l,k} = \underline{0}$
for $\nu = 1, \dots, \nu^{\mathrm{pre}}_{lk}$ **do**
  $\underline{w}_{l,k} = \underline{B}^{\mathrm{SM}}_{l,k}\underline{r}_{l,k}$
  $\underline{c}_{l,k} := \underline{c}_{l,k} + \underline{w}_{l,k}$ and $\underline{r}_{l,k} := \underline{r}_{l,k} - \underline{L}_{l,k}\underline{w}_{l,k}$
$\underline{r}_{l-1,k} = \underline{R}^{l,k}_{l-1,k}\underline{r}_{l,k}$ or $\underline{r}_{l,k-1} = \underline{R}^{l,k}_{l,k-1}\underline{r}_{l,k}$
$\underline{c}_{l-1,k} = \underline{B}^{\mathrm{ML}}_{l-1,k}\underline{r}_{l-1,k}$ or $\underline{c}_{l,k-1} = \underline{B}^{\mathrm{ML}}_{l,k-1}\underline{r}_{l,k-1}$
$\underline{w}_{l,k} = \underline{P}^{l,k}_{l-1,k}\underline{c}_{l-1,k}$ or $\underline{w}_{l,k} = \underline{P}^{l,k}_{l,k-1}\underline{c}_{l,k-1}$
$\underline{c}_{l,k} := \underline{c}_{l,k} + \underline{w}_{l,k}$ and $\underline{r}_{l,k} := \underline{r}_{l,k} - \underline{L}_{l,k}\underline{w}_{l,k}$
for $\nu = 1, \dots, \nu^{\mathrm{post}}_{lk}$ **do**
  $\underline{w}_{l,k} = \underline{B}^{\mathrm{SM}}_{l,k}\underline{r}_{l,k}$
  $\underline{c}_{l,k} := \underline{c}_{l,k} + \underline{w}_{l,k}$ and $\underline{r}_{l,k} := \underline{r}_{l,k} - \underline{L}_{l,k}\underline{w}_{l,k}$

The different multilevel strategies are tested for the linear transport equation with fixed polynomial degrees $(p, q) = (2, 2)$. We consider a divergence free vector field $\mathbf{q}(\mathbf{x}) = 2\pi\big(-\mathbf{x}_2, \mathbf{x}_1\big)^\top$ on $\Omega = (-10, 10)^2$ with homogeneous right-hand side $f = 0$, constant density $\rho \equiv 1$, final time $T = 1$, and starting with a 2D Gaussian pulse $u_0(\mathbf{x}) = \exp\big(-1.4\,((\mathbf{x}_1 - 5)^2 + \mathbf{x}_2^2)\big)$.

Several tests indicate that a block-Jacobi smoother with $\nu_{l,k} = 2$ smoothing steps and damping parameter $\theta_{\mathrm{J}} = 0.5$ in time, and a block-Gauss–Seidel smoother with $\nu_h = 5$ pre- and postsmoothing steps and no damping ($\theta_{\mathrm{GS}} = 1$) in space is a suitable choice. The contraction number of the two-level method on different space-time meshes $\mathcal{R}_{l,k}$ is estimated by the averaged convergence rate of the preconditioned linear iteration

$$\underline{u}^{\nu+1} = \underline{u}^{\nu} + B^{\mathrm{ML}}(\underline{f} - \underline{L}\,\underline{u}^{\nu})\,, \qquad \underline{u}^0 = \underline{0}\,,$$

see Tab. 1.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
	degrees of freedom on the space-time mesh $\mathcal{R}_{l,k}$ with polynomial degrees $(p,q) = (2,2)$					
$l=1$	768×16	768×32	768×64	768×128	768×256	768×512
$l=2$	$3\,072 \times 16$	$3\,072 \times 32$	$3\,072 \times 64$	$3\,072 \times 128$	$3\,072 \times 256$	$3\,072 \times 512$
$l=3$	$12\,288 \times 16$	$12\,288 \times 32$	$12\,288 \times 64$	$12\,288 \times 128$	$12\,288 \times 256$	$12\,288 \times 512$
$l=4$	$49\,152 \times 16$	$49\,152 \times 32$	$49\,152 \times 64$	$49\,152 \times 128$	$49\,152 \times 256$	$49\,152 \times 512$
	two-level iteration in time with Jacobi smoothing ($\nu_{l,k} = 2$, $\theta_{l,k} = 0.5$)					
$l=1$	26 (4.97e-1)	24 (4.65e-1)	9 (1.31e-1)	7 (5.87e-2)	7 (5.81e-2)	7 (5.77e-2)
$l=2$		38 (6.09e-1)	32 (5.65e-1)	7 (6.88e-2)	7 (5.22e-2)	7 (5.30e-2)
$l=3$			57 (7.19e-1)	48 (6.83e-1)	6 (4.52e-2)	6 (4.30e-2)
$l=4$				106 (8.40e-1)	94 (8.20e-1)	6 (3.51e-2)
	two-level iteration in time with Gauss-Seidel smoothing ($\nu_{l,k} = 5$)					
$l=1$	4 (1.31e-4)	4 (1.36e-4)	4 (1.85e-4)	4 (1.79e-4)	4 (1.68e-4)	4 (1.68e-4)
$l=2$	5 (1.35e-2)	5 (5.50e-3)	5 (4.72e-3)	5 (5.50e-3)	5 (5.28e-3)	5 (4.94e-3)
$l=3$	8 (7.57e-2)	7 (3.63e-2)	7 (2.26e-2)	6 (4.24e-2)	6 (4.07e-2)	6 (3.89e-2)
$l=4$	15 (2.70e-1)	11 (1.61e-1)	10 (1.34e-1)	9 (1.27e-1)	9 (1.24e-1)	9 (1.20e-1)

TABLE 1. Degrees of freedom of the transport example in the space-time domain on different space-time levels (starting with $128 = 16 \times 8$ space-time cells in $\mathcal{R}_{0,0}$), and iteration steps and averaged rates for a residual reduction by the factor 10^{-8} of the linear iteration with two-level multilevel preconditioners in time or space.

One observes that coarsening in time leads to stable multilevel behavior (the number of iteration steps are bounded by a constant) as long as $k - l \geq 2$, i.e., the ratio between $\triangle t$ and $\triangle x$ is bounded. For coarsening in space, we observe that the iteration steps are independent of the time level k, but not bounded in l. At least the increase is small enough to achieve a benefit by using a multilevel method. For higher spatial dimensions $D > 1$ coarsening in space is cheaper than coarsening in time, since refining in time doubles the effort whereas refining in space increases the effort by a factor $2D > 2$.

This and the previous observations motivate a strategy for the space-time multilevel solver, where we at first only coarse in space until the lowest spatial level is reached. Afterwards we coarse in time up to a lowest temporal level where $\triangle t/\triangle x$ is still small enough. The full multilevel V-cycle is illustrated in Fig. 5.

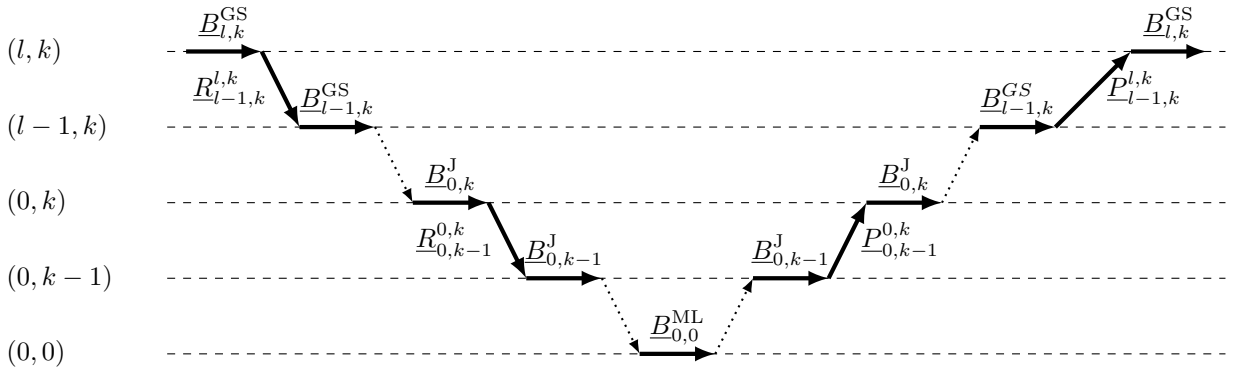


FIGURE 5. Full space-time coarsening strategy: coarsening in space up to level $(0,k)$, then coarsening in time up to level $(0,0)$; finally, solve exact on the coarsest level.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$l=1$	4 (1.25e-4)	4 (1.26e-4)	4 (1.80e-4)	4 (1.92e-4)	4 (1.86e-4)	4 (1.75e-4)
$l=2$	5 (1.35e-2)	5 (5.50e-3)	5 (4.71e-3)	5 (5.50e-3)	5 (5.28e-3)	5 (4.94e-3)
$l=3$	8 (7.57e-2)	7 (3.63e-2)	7 (2.25e-2)	6 (4.24e-2)	6 (4.07e-2)	6 (3.89e-2)
$l=4$	15 (2.73e-1)	11 (1.61e-1)	10 (1.34e-1)	9 (1.27e-1)	9 (1.24e-1)	9 (1.20e-1)

TABLE 2. Iteration steps and averaged rates for a full space-time multilevel method for the transport problem. Smoother: Jacobi ($\nu_{k,l} = 2$, $\theta_{l,k} = 0.5$) in time, Gauss–Seidel ($\nu_{l,k} = 5$) in space.

The results for this strategy applied to the test problem are given in Tab. 2. Due to the problems, observed for the two-level in space strategy, we achieve a moderate growth of iteration steps, when refining in space. We observe the same behavior for a 2D Maxwell test problem in $Q = (0,1)^2 \times (0,1)$, where the initial and boundary conditions are given by $\mathbf{u}_0(\mathbf{x},t) = \big(0, -\sin(2\pi(\mathbf{x}_1 - t)), \sin(2\pi(\mathbf{x}_1 - t))\big)^\top$, see Tab. 3.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
	degrees of freedom on the space-time mesh $\mathcal{R}_{l,k}$ with polynomial degrees $(p,q)=(2,2)$					
$l=1$	$1\,152\times 16$	$1\,152\times 32$	$1\,152\times 64$	$1\,152\times 128$	$1\,152\times 256$	$1\,152\times 512$
$l=2$	$4\,608\times 16$	$4\,608\times 32$	$4\,608\times 64$	$4\,608\times 128$	$4\,608\times 256$	$4\,608\times 512$
$l=3$	$18\,432\times 16$	$18\,432\times 32$	$18\,432\times 64$	$18\,432\times 128$	$18\,432\times 256$	$18\,432\times 512$
$l=4$	$73\,728\times 16$	$73\,728\times 32$	$73\,728\times 64$	$73\,728\times 128$	$73\,728\times 256$	$73\,728\times 512$
	multilevel iteration in space and time					
$l=1$	4 (3.42e-3)	4 (3.93e-3)	4 (4.03e-3)	4 (3.91e-3)	4 (3.70e-3)	4 (3.44e-3)
$l=2$	6 (3.18e-2)	6 (3.24e-2)	6 (3.13e-2)	6 (3.00e-2)	6 (2.85e-2)	6 (2.71e-2)
$l=3$	10 (1.31e-1)	10 (1.35e-1)	10 (1.31e-1)	10 (1.28e-1)	9 (1.59e-1)	9 (1.53e-1)
$l=4$	17 (3.62e-1)	17 (3.50e-1)	17 (3.44e-1)	17 (3.39e-1)	16 (3.68e-1)	16 (3.61e-1)

TABLE 3. Degrees of freedom for the Maxwell example in the space-time domain on different space-time meshes (starting with $64 = 8\times 8$ space-time cells in $\mathcal{R}_{0,0}$), and iteration steps and averaged rates for a full space-time multilevel method for the Maxwell example with Jacobi smoothing in time ($\nu_{l,k}=2$, $\theta_{l,k}=0.5$) and Gauss–Seidel smoothing in space ($\nu_{l,k}=5$).

In the adaptive case a coarse cell may correspond to a set of fine space-time cells of different polynomial degrees. To set up a polynomial distribution on the subspaces $V_{l-1,k}$ or $V_{l,k-1}$ (and correspondingly on $W_{l-1,k}$ or $W_{l,k-1}$) which does not impair the convergence rate, we apply the following strategy. For every coarse cell we use the highest polynomial degree in space and time on the subset of fine cells. Hence, we interpolate all solutions on the fine cells to this highest polynomial degree and use the restriction or prolongation matrices of the uniformly refined case. For the adaptive computations in the next section we observe the same (cf. Tab. 5) or slightly better (cf. Tab. 6 and Tab. 7) convergence behavior of the multilevel preconditioner.

7. Numerical Tests for space-time adaptivity

Finally we present results for the full adaptive method. We test the convergence properties for two examples, the linear transport equation for a configuration with known solution, which serves as a test problem to verify our methods, and a more sophisticated configuration for electro-magnetic waves in two spatial dimensions which is closer to practical applications. Here we use a generalized minimal residual solver (GMRES) equipped with the multilevel preconditioner from Sect. 6 and a residual reduction of 10^{-8} as stopping criterion. The adaptive strategy is described in Alg. 2 depending on a parameter $\vartheta < 1$ for the adaptive selection criterion.

Algorithm 2 Adaptive algorithm.

1: assemble the linear system $\underline{L}\,\underline{u} = \underline{f}$ with low order polynomial degrees
2: compute $\mathbf{u}_h$ by using GMRES with multilevel preconditioner
3: **while** $\max_R(p_R) < p_{\max}$ and $\max_R(q_R) < q_{\max}$ **do**
4: compute $\mathbf{u}_h^*$ and a recovery $I_h\mathbf{u}_h^*$
5: compute η_R on every cell R
6: mark space-time cell R if $\eta_R > \vartheta \max_{R'} \eta_{R'}$
7: increase polynomial degrees on marked cells
8: redistribute cells on processes for better load balancing
9: compute $\mathbf{u}_h$ with a new distribution of p_R and q_R

Linear transport. In the following numerical example we investigate the performance and reliability of our p-adaptive algorithm in comparison with uniform refinement for the example on the previous section. Since the characteristics for the transport vector $\mathbf{q}(\mathbf{x}) = 2\pi\,\big(-\mathbf{x}_2, \mathbf{x}_1\big)^\top$ are circles, we find $u(\mathbf{x}, 1) = u_0(\mathbf{x})$. We start with an initial coarse mesh with $1024 = 64 \times 16$ space-time cells which is refined 3 times in space and time up to 524 288 cells. The coarse problem is solved by using a parallel direct solver [MW11]. Furthermore we use low order polynomial degrees $(p, q) = (1, 1)$ as initial distribution on Q. In this test we aim to minimize the error $\triangle E = |E(u) - E(u_h)|$ towards the quadratic energy functional

$$E(v) = \frac{1}{2}(\rho v, v)_{0,Q}$$

using the dual error indicator derived in Sect. 5. Hence the adaptive strategy minimizes the energy error in Q.

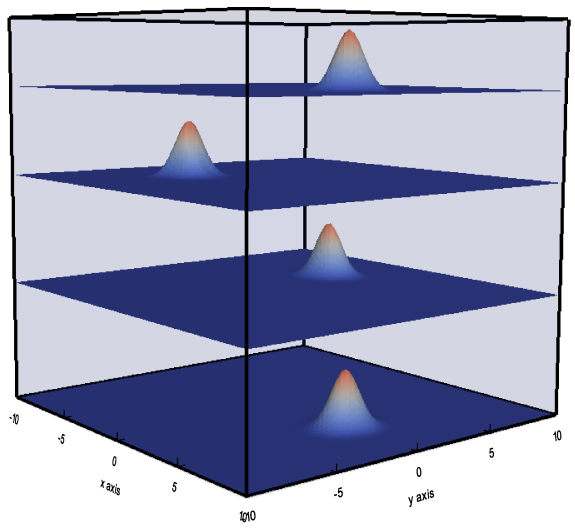


FIGURE 6. Solution of the transport equation in the space-time domain Q, sliced at times $t = 0, 0.3, 0.6, 1$.

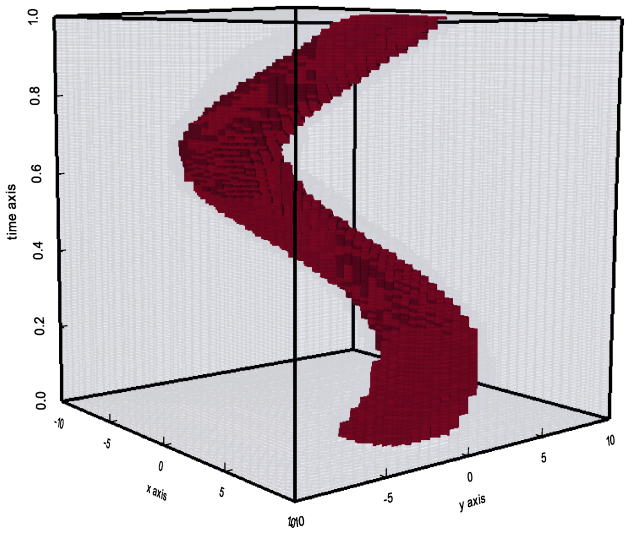


FIGURE 7. Location of the highest polynomial degrees in the space-time domain Q.

The exact solution of the dual problem

$$-\rho\partial_t u^* - \operatorname{div}(u^*\mathbf{q}) = \rho u\,, \qquad \text{on } \Omega \times (0, T)\,, \qquad u^*(T) = 0\,,$$

with homogeneous Dirichlet boundary conditions is given by $u^*(\mathbf{x}, t) = (T - t)\,u(\mathbf{x}, t)$, since for all $w \in V$

$$\begin{aligned}(w, L^*u^*)_{0,Q} &= \big(w, -\rho\partial_t u^* - \mathbf{q}\cdot\nabla u^*\big)_{0,Q} = \big(w, \rho u - (T-t)(\rho\partial_t u + \mathbf{q}\cdot\nabla u)\big)_{0,Q} \\ &= \big(\rho u - (T-t)(\rho\partial_t u + \operatorname{div}(\mathbf{q}u)), w\big)_{0,Q} = \big(\rho u - (T-t)f, w\big)_{0,Q} = \big(\rho u, w\big)_{0,Q} = \langle E'(u), w\rangle\,.\end{aligned}$$

Thus, u^* also corresponds to a Gaussian pulse (travelling backwards in time).

The adaptive results are given in Tab. 5 and Fig. 8. First we observe that the estimation for the dual error $e_h^* = u_h^* - I_h u_h^*$ approximates the exact dual error $e^* = u^* - u_h^*$ well. Using the solutions u and u_h the exact errors $\triangle E = |E(u) - E(u_h)|$ and $\|u_h(T) - u(T)\|_{0,\Omega}$ can be computed. Furthermore $\triangle E$ can be estimated using (11) with approximations e_h^* and u_h. These results, denoted as $\triangle E_h$, almost coincide with $\triangle E$. Finally the

sum over all cell-wise estimated errors $\eta_{\mathcal{R}} = \sum_{R\in\mathcal{R}} \eta_R$ computed by (11) shows the same asymptotic behavior, which is required for reliable error estimation. Fig. 6 shows the adaptive solution in the space time domain. In comparison with Fig. 7 we see that highest polynomial degrees are only used in areas where the pulse is actually located, whereas lowest polynomial degrees are used everywhere else.

The benefit of adaptive strategies becomes clear in Fig. 9, where we compare the adaptive solution with a uniformly refined solution (see Tab. 4). On the last refinement level we achieve the same errors $\triangle E$ and $\|u_h(T) - u(T)\|_{0,\Omega}$ by using only approx. 3.3 million degrees of freedom. This corresponds to a reduction of about 90% compared to the uniformly refined case. The benefit depends on the underlying problem. But if the solution is strongly located (e.g., a Gaussian pulse or a single wavefront) or one is only interested in small parts of the solution (as in our next example), it is possible to save a large amount of computational resources.

level	uniform poly. deg. (p,q)	#DoFs	GMRES steps (rate)	$\triangle E$	$\|u_h(T) - u(T)\|_{0,\Omega}$
$l=1$	(1,1)	1 585 152	10 (7.19e-2)	5.10e-2	4.06e-1
$l=2$	(2,2)	6 340 608	10 (1.30e-1)	2.14e-3	1.97e-2
$l=3$	(3,3)	15 851 520	10 (1.54e-1)	3.78e-5	8.52e-4
$l=4$	(4,4)	31 703 040	11 (1.67e-1)	4.41e-7	5.16e-4

TABLE 4. Results for the transport equation with uniform mesh with $524\,288 = 4\,096 \times 128$ space-time cells and different polynomial degrees.

level	#DoFs (effort)	GMRES steps (rate)	$\triangle E$	$\|u_h(T) - u(T)\|_{0,\Omega}$	$\|e^* - e_h^*\|_{0,Q}$	$\triangle E_h$	E
$l=1$	1 585 152	10 (7.19e-2)	5.10e-2	4.06e-1	1.78e-1	4.08e-2	7.60e-1
$l=2$	1 894 176 (30%)	10 (9.53e-2)	2.14e-3	2.02e-2	9.98e-3	2.63e-3	3.59e-2
$l=3$	2 381 598 (15%)	10 (1.43e-1)	3.79e-5	1.87e-3	8.40e-4	4.44e-5	7.33e-4
$l=4$	3 303 810 (10%)	11 (1.23e-1)	4.31e-7	5.22e-4	5.29e-4	4.94e-7	1.47e-5

TABLE 5. Adaptive refinement on a mesh with $524\,288 = 4\,096 \times 128$ space-time cells (ϑ = 1e-4).

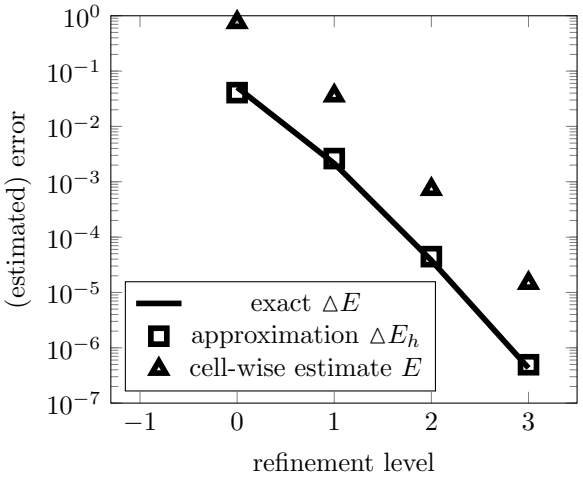


FIGURE 8. Error estimation.

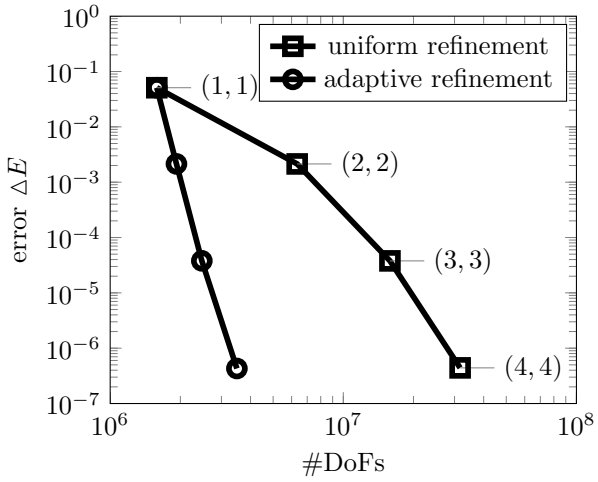


FIGURE 9. Adaptive performance.

Electro-magnetic waves. We consider a 2D transverse electric wave $\mathbf{u} = \big(\mathbf{H}_1, \mathbf{H}_2, \mathbf{E}_3\big)^\top$ with wavelength $\lambda = 1$. It is scattered by a double slit with slit gap $a = 3$ and slit width $b = 1$. The scattered wave enters the computational domain $Q = (0, 6) \times (-6, 6) \times (0, 8)$ on the left (see Fig. 10). Furthermore we apply constant material parameters $\mu = \varepsilon = 1$ and reflecting boundary conditions. Behind the double slit one observes a diffraction pattern with several local intensity extrema. In applications one is often only interested in certain small parts of the scattered wave. Therefore we choose the region of interest as $S = (5.5, 6) \times (0, 2) \times (0, 8)$ to resolve the first minimum at the right end of our computational domain as good as possible. Hence the energy error functional is given as

$$E(\mathbf{v}) = \frac{1}{2}(M\mathbf{v}, \mathbf{v})_{0,S}\,.$$

The corresponds to a screen or receiver somewhere in S to receive and measure the scattered wave as illustrated in Fig. 10. Since in this setting the exact value of $E(\mathbf{u})$ is not known, we approximate $\triangle E$ by extrapolation

$$E(\mathbf{u}) \approx E_{\text{ex}} = \frac{r}{r-1}E_l - \frac{1}{r-1}E_{l-1}\,, \qquad r = \frac{|E_{l-1} - E_{l-2}|}{|E_l - E_{l-1}|}\,, \qquad E_l = E(\mathbf{u}_l)\,.$$

We perform two tests on two different levels with 256 and 1024 processes, respectively. In the first case the initial coarse mesh consists of $9\,472 = 148 \times 64$ space-time cells and is refined 2 times in space and time up to 606 208 cells. We use 256 processes to compute a uniform and an adaptive refined solution. Similarly to the first example we observe from Tab. 6 that the estimated value of the error functional coincidence in both cases and we are able to save about 73% of the degrees of freedom.

In the second case we use 1024 processes for a problem that is refined once more in space and time up to 4 849 664 cells. To be able do some reasonable load balancing according to the degrees of freedom on each cell, we have to refine the coarse mesh too (i.e., 592×128 cells). We see from Tab. 7 that we save about 79% of the degrees of freedom and hence are still able to compute an accurate adaptive solution (with respect to E), although uniform refinement for higher degrees is not possible. In both cases we used a Gauss–Seidel preconditioned GMRES solver for the coarse problem. Fig. 11 shows the time evolution of the scattered wave computed on 1024 processes is shown. The diffraction pattern and the result of the adaptive error estimation are clearly visible. The adaptive solution uses highest polynomial degrees in areas where it is necessary to have a high resolution in S and lowest polynomial everywhere else.

		uniform refinement				adaptive refinement			
level	(p, q)	#DoFs	steps (rate)	E_l	$\lvert E_l - E_{\text{ex}}\rvert$	#DoFs (effort)	steps (rate)	E_l	$\lvert E_l - E_{\text{ex}}\rvert$
$l = 1$	$(1, 1)$	5 477 184	10 (1.14e-1)	8.9307e-2	3.1303e-1	5 477 184	10 (1.14e-1)	8.9307e-2	3.1303e-1
$l = 2$	$(2, 2)$	21 908 736	17 (2.84e-1)	3.7612e-1	2.6220e-2	9 645 930 (44%)	13 (2.33e-1)	3.7611e-1	2.6230e-2
$l = 3$	$(3, 3)$	54 771 840	24 (4.48e-1)	4.0089e-1	1.4502e-3	17 309 043 (32%)	18 (3.37e-1)	4.0089e-1	1.4502e-3
$l = 4$	$(4, 4)$	109 543 680	34 (5.68e-1)	4.0226e-1	8.0209e-5	29 064 348 (27%)	23 (4.39e-1)	4.0226e-1	8.0209e-5

TABLE 6. Uniform vs. adaptive refinement on $606\,208 = 2\,368 \times 256$ space-time cells distributed to 256 processes (ϑ = 1e-3, E_{ex} = 4.0234e-1).

		uniform refinement				adaptive refinement			
level	(p, q)	#DoFs	steps (rate)	E_l	$\lvert E_l - E_{\text{ex}}\rvert$	#DoFs (effort)	steps (rate)	E_l	$\lvert E_l - E_{\text{ex}}\rvert$
$l = 1$	$(1, 1)$	43 732 224	17 (3.04e-1)	3.0520e-1	9.7373e-2	43 732 224	17 (3.04e-1)	3.0520e-1	9.7373e-2
$l = 2$	$(2, 2)$	174 928 896	31 (5.36e-1)	4.0081e-1	1.7630e-3	68 437 899 (39%)	21 (4.01e-1)	4.0082e-1	1.7530e-3
$l = 3$	$(3, 3)$	437 322 240	out of memory			115 207 920 (26%)	28 (5.07e-1)	4.0250e-1	7.3043e-5
$l = 4$	$(4, 4)$	874 644 480	out of memory			184 208 094 (21%)	37 (5.82e-1)	4.0257e-1	3.0435e-6

TABLE 7. Uniform vs. adaptive refinement on $4\,849\,664 = 9\,472 \times 512$ space-time cells distributed to 1024 processes (ϑ = 1e-3, E_{ex} = 4.0257e-1).

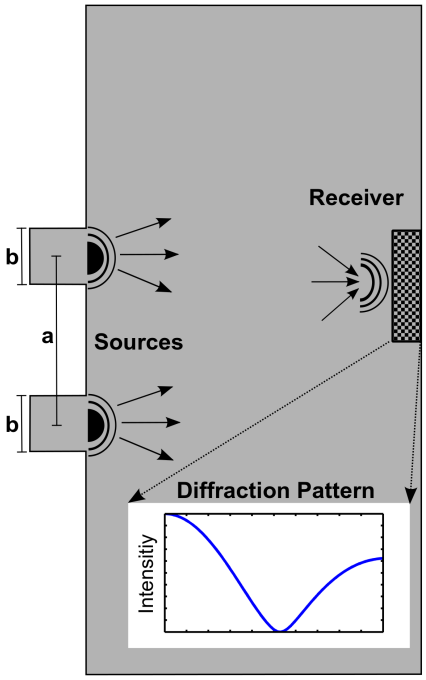


FIGURE 10. Schematic illustration of the double-slit experiment. Slit gap $a = 3$ and slit width $b = 1$.

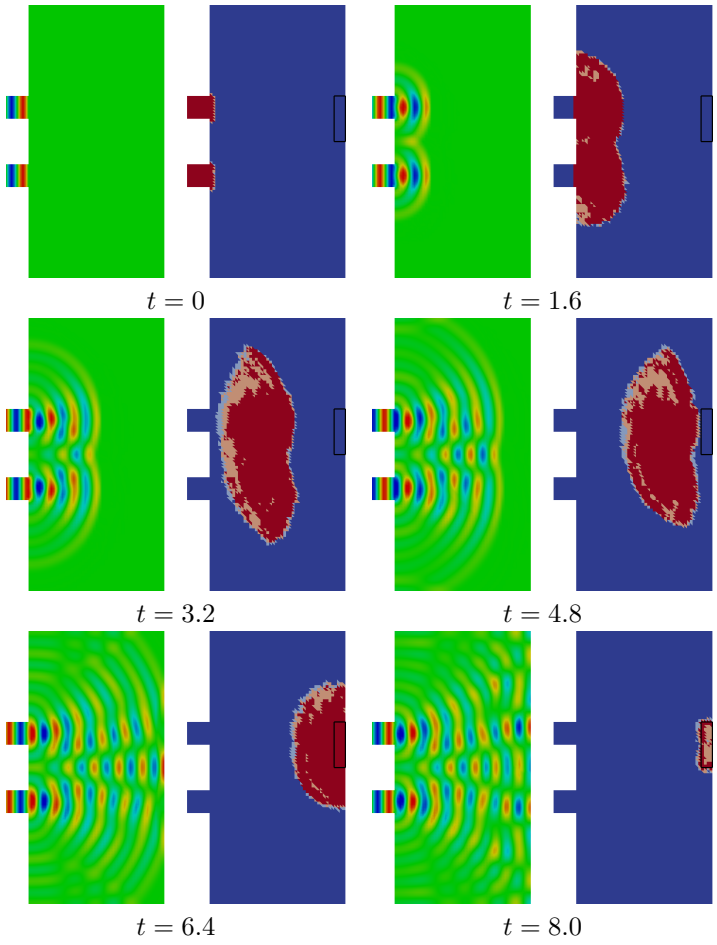


FIGURE 11. Scattered wave solution (left) and used polynomial degrees (right) at different times. Solved on 1024 processes.

All numerical results where computed with 256 or 1024 processes on the ForHLR cluster at KIT, where a node contains two Intel Xeon E5-2670 v2 (2,5 GHz, 10 cores) and 64GB memory.

Acknowledgment. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) through RTG 1294 and CRC 1173.

REFERENCES

[ARW95] U. M. Ascher, S. J. Ruuth, and B. T. R. Wetton. Implicit-explicit methods for time-dependent partial differential equations. *SIAM J. Numer. Anal.*, 32(3):797–823, 1995.

[BR99] W. Bangerth and R. Rannacher. Finite element approximation of the acoustic wave equation: error control and mesh adaptation. *East-West J. Numer. Math.*, 7(4):263–282, 1999.

[BR03] W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. Birkhäuser Verlag, Basel, 2003.

[Bra07] D. Braess. *Finite Elements. Theory, fast solvers, and applications in solid mechanics. 3th ed.* Cambridge University Press, Cambridge, 2007.

[DG14] L. F. Demkowicz and J. Gopalakrishnan. An overview of the discontinuous Petrov–Galerkin method. In *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations*, pages 149–180. Springer, 2014.

[DKT07] M. Dumbser, M. Käser, and E. F. Toro. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – V. Local time stepping and p-adaptivity. *Geophys. J. Int.*, 171:695–717, 2007.

[EDCM14] T. E. Ellis, L. F. Demkowicz, J. L. Chan, and R. D. Moser. Space-time DPG: Designing a method for massively parallel CFD, 2014. ICES REPORT 14-32.

[EKSW14] H. Egger, F. Kretzschmar, S. M. Schnepp, and T. Weiland. A space-time discontinuous Galerkin–Trefftz method for time dependent Maxwell's equations. *arXiv preprint, arXiv:1412.2637 (to appear: SIAM J. Sci. Comput.)*, 2014.

[EM12] M. Emmett and M. L. Minion. Toward an Efficient Parallel in Time Method for Partial Differential Equations. *Comm. Appl. Math. and Comp. Sci.*, 7:105–132, 2012.

[Eva10] L. C. Evans. *Partial differential equations, 2. ed.* American Mathematical Society, Providence, RI, 2010.

[FFK+14] R. D. Falgout, S. Friedhoff, T. V. Kolev, S. P. MacLachlan, and J. B. Schroder. Parallel time integration with multigrid. *SIAM J. Sci. Comput.*, 36(6):C635–C661, 2014.

[Gan15] M. J. Gander. 50 years of time parallel time integration. In T. Carraro, M. Geiger, S. Körkel, and R. Rannacher, editors, *Multiple Shooting and Time Domain Decomposition*. Springer, 2015.

[GHN03] M. J. Gander, L. Halpern, and F. Nataf. Optimal Schwarz waveform relaxation for the one dimensional wave equation. *SIAM J. Numer. Anal.*, 41(5):1643–1681, 2003.

[GN14] M. J. Gander and M. Neumüller. Analysis of a new space-time parallel multigrid algorithm for parabolic problems. *arXiv preprint, arXiv:1411.0519*, 2014.

[GS09] M. J. Grote and D. Schötzau. Optimal error estimates for the fully discrete interior penalty DG method for the wave equation. *J. Sci. Comput.*, 40(1-3):257–272, 2009.

[GV07] M. J. Gander and S. Vandewalle. Analysis of the parareal time-parallel time-integration method. *SIAM J. Sci. Comput.*, 29(2):556–578, 2007.

[HPS+14] M. Hochbruck, T. Pazur, A. Schulz, E. Thawinan, and C. Wieners. Efficient time integration for discontinuous Galerkin approximations of linear wave equations. *ZAMM*, 95(3):237–259, 2014.

[HR03] V. Heuveline and R. Rannacher. Duality-based adaptivity in the hp-finite element method. *J. Numer. Math.*, 11:95–113, 2003.

[HS06] P. Houston and E. Süli. hp-adaptive discontinuous Galerkin finite element methods for first-order hyperbolic problems. *SIAM J. Sci. Comput.*, 23(4):1226–1252, 2006.

[HW08] J. S. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods*. Springer, 2008.

[KB13] U. Köcher and M. Bause. Variational space-time methods for the wave equation. *J. Sci. Comput.*, pages 1–30, 2013.

[KMPS15] F. Kretzschmar, A. Moiola, I. Perugia, and S. M. Schnepp. A priori error analysis of space-time Trefftz discontinuous Galerkin methods for wave problems. *arXiv preprint, arXiv:1501.05253*, 2015.

[LMT01] J.-L. Lions, Y. Maday, and G. Turinici. A "parareal" in time discretization of PDE's. *C. R. Acad. Sci. Paris, Ser. I*, 332(7):661–668, 2001.

[MW11] D. Maurer and C. Wieners. A parallel block LU decomposition method for distributed finite element matrices. *Parallel Comput.*, 37(12):742–758, 2011.

[NPC11] N. C. Nguyen, J. Peraire, and B. Cockburn. High-order implicit hybridizable discontinuous Galerkin methods for acoustics and elastodynamics. *J. Comput. Phys.*, 230(10):3695–3718, 2011.

[OPD05] J. T. Oden, S. Prudhomme, and L. Demkowicz. A posteriori error estimation for acoustic wave propagation problems. *Arch. Comput. Methods Engrg.*, 12(4):343–389, 2005.

[vdVR12] J. J. W. van der Vegt and S. Rhebergen. hp-multigrid as smoother algorithm for higher order discontinuous Galerkin discretizations of advection dominated flows: Part I. Multilevel analysis. *J. Comput. Phys.*, 231(22):7537–7563, 2012.

[Wie10] C. Wieners. A geometric data structure for parallel finite elements and the application to multigrid methods with block smoothing. *Comput. Visual. Sci.*, 13:161–175, 2010.

[WTF14] D. Wang, R. Tezaur, and C. Farhat. A hybrid discontinuous in space and time Galerkin method for wave propagation problems. *Int. J. Numer. Meth. Engng*, 2014.

[WW14] C. Wieners and B. Wohlmuth. Robust operator estimates and the application to substructuring methods for first-order systems. *ESAIM: M^2AN*, 48:161–175, 2014.

[ZMD+11] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V. M. Calo. A class of discontinuous Petrov–Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1D. *J. Comput. Phys.*, 230(7):2406–2432, 2011.