NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2010; **00**:1–17
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

Convergence of the multigrid-reduction-in-time algorithm for the linear elasticity equations

A. Hessenthaler[1]*, D. Nordsletten[2], O. Röhrle[1], J. B. Schroder[3], R. D. Falgout[3]

[1]*Institute of Applied Mechanics (CE), University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany*
[2]*Division of Imaging Sciences and Biomedical Engineering, King's College London, 4th Floor, Lambeth Wing St. Thomas Hospital London, SE1 7EH, UK*
[3]*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551*

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52–07NA27344, LLNL-JRNL-731168.

SUMMARY

This paper presents some recent advances for parallel-in-time methods applied to linear elasticity. With recent computer architecture changes leading to stagnant clock speeds, but ever increasing numbers of cores, future speedups will be available through increased concurrency. Thus, sequential algorithms, such as time stepping, will suffer a bottleneck. This paper explores multigrid reduction in time (MGRIT) for an important application area for many time stepping codes, linear elasticity. Previously, efforts at parallel-in-time for elasticity have experienced difficulties, for example, the beating phenomenon, leading to the current state of no practical parallel-in-time algorithm existing for this application area. This paper proposes some solutions made possible by MGRIT (*e.g.*, slow temporal coarsening and FCF-relaxation) and more importantly, a different formulation of the problem that is more amenable to parallel-in-time methods. Using a recently developed convergence theory for MGRIT and Parareal, we show that the changed formulation of the problem avoids the instability issues and allows reduction of the error using two temporal grids. The paper ends with supporting numerical results showing a practical algorithm. Copyright © 2010 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: multigrid reduction in time (MGRIT); parallel-in-time; linear elasticity; convergence estimate

1. INTRODUCTION

Clock rates of recent computer architectures have become stagnant, while the available number of parallel processors has increased and continues to increase rapidly. Thus, simulation algorithms need to allow greater concurrency to exploit massively parallel hardware and further reduce wall clock time. One severe sequential bottleneck in many parallel application codes is the use of sequential time integration methods. This sequential bottleneck limits parallelism to the spatial component of an investigated space-time problem.

One such research area where sequential time-stepping is almost omnipresent is fluid-structure interaction (FSI) research. Here, the system under consideration studies the interaction between fluids and (often deformable) solid structures or particles, for example, in biomedical or aerospace

*Correspondence to: A. Hessenthaler, Institute of Applied Mechanics (CE), University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany. Email: hessenthaler@mechbau.uni-stuttgart.de

engineering applications [1, 2, 3]. Well-established parallelization techniques, such as spatial domain decomposition methods, provide a straightforward and scalable approach for reducing the wall clock time for many FSI algorithms. However, spatial parallelism saturates when communication tasks become dominant over computation tasks. This often prohibits use of parallel resources beyond an optimal processor count, leaving large numbers of processors unused. On the other hand, parallel-in-time integration methods provide means of introducing an additional layer of parallelism with the potential to greatly enhance efficient and more exhaustive use of parallel resources. Despite its long history [4, 5], parallel-in-time has very little uptake in the large-scale simulation community. Only few works exist that use parallel-in-time ideas in real-world simulations, including reservoir simulation [6], fusion research [7] and numerical weather prediction [8].

In the field of FSI research, limitations to using parallel-in-time methods usually exist because of an instability arising in the parallel-in-time integration of one of its subproblems, that is, dynamic structural mechanics. The instability is known as the beating phenomenon [9, 10]. Although this instability can be overcome by filtering the natural modes of a given structural dynamics problem [10, 11], only small scale parallelism was explored. This can mainly be related to a relatively expensive projection step and the restriction of the method to two time grid levels. Thus, it is of prime importance to obtain a stable and robust parallel-in-time technique that can explore greater parallelism while avoiding the previously mentioned instability. Such a method can then be the foundation of a fast and efficient parallel-in-time method applicable to FSI applications.

The focus of this work is the investigation of convergence of the multigrid reduction in time (MGRIT) method for the second-order partial differential equation governing the dynamic linear-elastic response of an incompressible solid structure. In a special case (two temporal grids with F-relaxation), MGRIT is equivalent [12] with Parareal [13] but has several benefits. For example, MGRIT allows use of FC-relaxation, slow temporal coarsening and is a non-intrusiveness algorithm. Moreover, MGRIT is a true multilevel algorithm. In this paper, we will make use of the recent work of the authors of [12] to use the presented two-grid convergence theory as a design tool for convergent algorithms. The beating phenomenon was described qualitatively in [9, 10]. Here, we will show that it is also founded in the two-grid theory. Furthermore, the same analysis provides a powerful tool for estimating convergence *a priori* that leads to a convergent two-grid MGRIT algorithm for the dynamic second-order elasticity equations and further, to a convergent multigrid MGRIT algorithm.

In Section 2, two different backward Euler time discretization schemes are presented (referred to as Scheme I and Scheme II) for a finite element implementation of the second-order hyperbolic elasticity equations. The two schemes are embedded in the MGRIT algorithm, where we provide a description of the parallel-in-time algorithm and discuss the application of the convergence theory presented in [12]. We discuss the difference between both time discretization schemes in Section 3.2. In Section 3, we relate the theoretical analysis of the two-grid convergence to observations in numerical experiments. Further, we demonstrate how to derive a convergent MGRIT algorithm that does not exhibit the instabilities reported in previous works [9, 10] by using [12] as an algorithm design tool. We then extend the numerical experiments to a multilevel case with FMG-cycles in Section 3.4. The section concludes with preliminary speedup results, emphasizing the potential for parallel speedup when using MGRIT for linear elasticity problems.

2. METHODOLOGY

2.1. Notation

Consider the domain $\Omega = \Omega(t) \subset \mathbb{R}^d \times [0, T]$ with Dirichlet boundary Γ^D in d spatial dimensions. Then, $\boldsymbol{X} \in \Omega(0)$ and $\boldsymbol{x} \in \Omega$ denote the reference and current position of a material point and $\nabla_{\boldsymbol{X}}$ and $\nabla_{\boldsymbol{x}}$ denote the respective Lagrangian and Eulerian gradient operators. Further, we define the deformation gradient $\boldsymbol{F} = \nabla_{\boldsymbol{X}} \boldsymbol{x} = \nabla_{\boldsymbol{X}} \boldsymbol{u} + \boldsymbol{I}$ where $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{x} - \boldsymbol{X}$ is the displacement of

DOI: 10.1002/nla

a material point with respect to its position in the reference configuration. The partial derivative operator is denoted as ∂_t and $\partial_{tt} = \partial_t \partial_t$.

2.2. Governing equations

Consider the governing equations for the dynamic and linear elastic response of an incompressible solid structure with given initial data and Dirichlet boundary condition data,

$$\rho \partial_{tt} \boldsymbol{u} - \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\sigma} = \boldsymbol{0} \qquad \text{in } \Omega, \tag{1}$$
$$\nabla_{\boldsymbol{x}} \cdot \partial_t \boldsymbol{u} = 0 \qquad \text{in } \Omega, \tag{2}$$
$$\boldsymbol{u}(\cdot, t) = \boldsymbol{0} \qquad \text{on } \Gamma^D, \tag{3}$$
$$\boldsymbol{u}(\cdot, 0) = \boldsymbol{0}, \quad \boldsymbol{v}(\cdot, 0) = \hat{\boldsymbol{v}}_0 \qquad \text{in } \Omega(0), \tag{4}$$

with density ρ, Cauchy stress tensor $\boldsymbol{\sigma}(\boldsymbol{u}, p) = \mu(\boldsymbol{F} - \boldsymbol{I}) - p\boldsymbol{I}$, material stiffness parameter μ, the hydrostatic pressure variable p and initial velocity $\hat{\boldsymbol{v}}_0$.

Equation 1 can be transformed to a system of first-order equations,

$$\partial_t \boldsymbol{u} = \boldsymbol{v} \qquad \text{in } \Omega, \tag{5}$$
$$\rho \partial_t \boldsymbol{v} = \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\sigma} \qquad \text{in } \Omega, \tag{6}$$
$$\nabla_{\boldsymbol{x}} \cdot \boldsymbol{v} = 0 \qquad \text{in } \Omega, \tag{7}$$
$$\boldsymbol{u}(\cdot, t) = \boldsymbol{0} \qquad \text{on } \Gamma^D, \tag{8}$$
$$\boldsymbol{u}(\cdot, 0) = \boldsymbol{0}, \quad \boldsymbol{v}(\cdot, 0) = \hat{\boldsymbol{v}}_0 \qquad \text{in } \Omega(0), \tag{9}$$

with velocity $\boldsymbol{v}$.

To reduce the complexity of the computational model, Equation 5 is eliminated from the system of equations in the following by including it implicitly. That is, we solve for velocity $\boldsymbol{v}$ and hydrostatic pressure p and update the displacement variable based on the solution for the velocity variable.

Further, we note that all quantities are computed on the reference domain $\Omega(0)$. That is, linear elastic response is assumed and higher order effects of the deforming domain are neglected.

2.3. Time discretization

We decompose the temporal domain [0, T] by $N_t + 1$ equidistant time points, such that

$$t_i = i \cdot \delta_{N_t}, \quad \text{with } i = 0, \ldots, N_t, \tag{10}$$

with time step size $\delta_{N_t} = T/N_t$, initial time $t_0 = 0$ and final time $t_{N_t} = T$.

Then, velocity, pressure and displacement at time point t_i are denoted as $[\boldsymbol{v}_i, p_i, \boldsymbol{u}_i]^T = [\boldsymbol{v}(\cdot, t_i), p(\cdot, t_i), \boldsymbol{u}(\cdot, t_i)]^T$. We now introduce the two discretization schemes (Scheme I and Scheme II) examined here.

2.3.1. Scheme I

Firstly, we approximate the partial derivative operator ∂_t in Equation 5 using the midpoint rule,

$$\boldsymbol{u}_i = \boldsymbol{u}_{i-1} + \delta_{N_t} \frac{\boldsymbol{v}_i + \boldsymbol{v}_{i-1}}{2} \qquad \text{in } \Omega_0. \tag{11}$$

The partial derivative operator in Equation 6 is discretized using the backward Euler scheme where we substitute Equation 11. Thus, we search $(\boldsymbol{v}_i, p_i)$ for all $i = 1, \ldots, N_t$, such that,

$$\rho \boldsymbol{v}_i - \frac{\mu \delta_{N_t}^2}{2} \nabla_{\boldsymbol{X}}^2 \boldsymbol{v}_i + \delta_{N_t} \nabla_{\boldsymbol{X}} p_i = \rho \boldsymbol{v}_{i-1} + \frac{\mu \delta_{N_t}^2}{2} \nabla_{\boldsymbol{X}}^2 \boldsymbol{v}_{i-1} + \mu \delta_{N_t} \nabla_{\boldsymbol{X}}^2 \boldsymbol{u}_{i-1} \qquad \text{in } \Omega_0, \tag{12}$$
$$\nabla_{\boldsymbol{X}} \cdot \boldsymbol{v}_i = -\nabla_{\boldsymbol{X}} \cdot \boldsymbol{v}_{i-1} \qquad \text{in } \Omega_0, \tag{13}$$

and compute the displacement $\boldsymbol{u}^i$ according to Equation 11 after each solve.

Prepared using nlaauth.cls
DOI: 10.1002/nla

2.3.2. *Scheme II*

We can obtain a slightly modified scheme by approximating the partial derivative operator in Equation 5 as,

$$\boldsymbol{u}_i = \boldsymbol{u}_{i-1} + \delta_{N_t} \boldsymbol{v}_i \qquad \text{in } \Omega_0. \tag{14}$$

Thus, in Scheme II we search the time-discrete $(\boldsymbol{v}_i, p_i)$ for all $i = 1, \dots, N_t$, such that,

$$\rho \boldsymbol{v}_i - \mu \delta_{N_t}^2 \nabla_{\boldsymbol{X}}^2 \boldsymbol{v}_i + \delta_{N_t} \nabla_{\boldsymbol{X}} p_i = \rho \boldsymbol{v}_{i-1} + \mu \delta_{N_t} \nabla_{\boldsymbol{X}}^2 \boldsymbol{u}_{i-1} \qquad \text{in } \Omega_0, \tag{15}$$

$$\nabla_{\boldsymbol{X}} \cdot \boldsymbol{v}_i = 0 \qquad \text{in } \Omega_0, \tag{16}$$

The modification in Equation 14 is motivated by the observation, that Scheme I does not yield a convergent MGRIT algorithm and exhibits the same instability as described in previous works [9, 10], see Section 3.3. Scheme II does not suffer from this instability, however, does not reproduce amplitudes of oscillation as well as Scheme I (see Section 3.2).

2.4. *Space discretization*

The domain Ω_i was discretized using quadrilateral elements, Ω_i^h. Finite element discretizations were constructed using inf-sup stable $\mathbb{Q}^2 - \mathbb{Q}^1$ Taylor-Hood elements for velocity and pressure, $\boldsymbol{v}_i^h$ and p_i^h, and $\mathbb{Q}^2$ elements for displacement, $\boldsymbol{u}_i^h$. The superscript h denotes the space-discretized version of the domain and state variables. As we do not consider spatial refinement or coarsening, we omit the superscript h for the remainder of this work.

2.4.1. *Scheme I*

Spatially discretizing Equation 12 and Equation 13 leads to the problem of seeking for each $i = 1, \dots, N_t$ the space-time discrete solution vector $[\boldsymbol{v}_i, \boldsymbol{p}_i]^T$, such that,

$$(\rho \boldsymbol{M} - \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K}) \boldsymbol{v}_i + \delta_{N_t} \boldsymbol{B}^T \boldsymbol{p}_i = (\rho \boldsymbol{M} + \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K}) \boldsymbol{v}_{i-1} + \mu \delta_{N_t} \boldsymbol{K} \boldsymbol{u}_{i-1} \qquad \text{in } \Omega_0, \tag{17}$$

$$\boldsymbol{B} \boldsymbol{v}_i = -\boldsymbol{B} \boldsymbol{v}_{i-1} \qquad \text{in } \Omega_0, \tag{18}$$

where $\boldsymbol{M}$ is the mass matrix and $\boldsymbol{K}$ and $\boldsymbol{B}$ refer to the discretized weak form Laplacian and divergence operators $\nabla_{\boldsymbol{X}}^2()$ and $\nabla_{\boldsymbol{X}} \cdot ()$. Note, after solving for a given $[\boldsymbol{v}_i, \boldsymbol{p}_i]^T$, we can update the displacement $\boldsymbol{u}_i$ from Equation 11. Writing Equation 17 and Equation 18 in matrix form and including the update given in Equation 11 yields the following linear system,

$$\begin{bmatrix} \rho \boldsymbol{M} - \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K} & \delta_{N_t} \boldsymbol{B}^T & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ -\frac{\delta_{N_t}}{2} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{p}_i \\ \boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} \rho \boldsymbol{M} + \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K} & \boldsymbol{0} & \mu \delta_{N_t} \boldsymbol{K} \\ -\boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ \frac{1}{2} \delta_{N_t} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{i-1} \\ \boldsymbol{p}_{i-1} \\ \boldsymbol{u}_{i-1} \end{bmatrix}. \tag{19}$$

Letting the linear operators on the left and right hand sides be denoted,

$$\boldsymbol{D}^I := \begin{bmatrix} \rho \boldsymbol{M} - \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K} & \delta_{N_t} \boldsymbol{B}^T & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ -\frac{\delta_{N_t}}{2} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \quad \boldsymbol{C}^I := \begin{bmatrix} \rho \boldsymbol{M} + \frac{\mu \delta_{N_t}^2}{2} \boldsymbol{K} & \boldsymbol{0} & \mu \delta_{N_t} \boldsymbol{K} \\ -\boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ \frac{1}{2} \delta_{N_t} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \tag{20}$$

then $\boldsymbol{\Phi}^I := [\boldsymbol{D}^I]^{-1} \boldsymbol{C}^I$ and $\boldsymbol{s}_0^I = \hat{\boldsymbol{s}}_0 := [\hat{\boldsymbol{v}}_0, \boldsymbol{0}, \boldsymbol{0}]^T$, the state variables $\boldsymbol{s}_i^I := [\boldsymbol{v}_i, \boldsymbol{p}_i, \boldsymbol{u}_i]^T$ can be computed by the following equation,

$$\boldsymbol{s}_i^I = \boldsymbol{\Phi}^I \boldsymbol{s}_{i-1}^I \qquad \text{for } i = 1, \dots, N_t. \tag{21}$$

Note, that in the linear elastic incompressible case with constant spatial resolution, the operator $\boldsymbol{\Phi}^I$ is only dependent on time step size δ_{N_t}.

DOI: 10.1002/nla

2.4.2. *Scheme II*

In a similar way, Scheme II can be written as: we seek for each $i = 1, \ldots, N_t$ the space-time discrete $[\boldsymbol{v}_i, \boldsymbol{p}_i]^T$, such that,

$$(\rho \boldsymbol{M} - \mu \delta_{N_t}^2 \boldsymbol{K})\boldsymbol{v}_i + \delta_{N_t} \boldsymbol{B}^T \boldsymbol{p}_i = \rho \boldsymbol{M} \boldsymbol{v}_{i-1} + \mu \delta_{N_t} \boldsymbol{K} \boldsymbol{u}_{i-1} \qquad \text{on } \Omega_0, \tag{22}$$

$$\boldsymbol{B}\boldsymbol{v}_i = 0 \qquad \text{on } \Omega_0, \tag{23}$$

and update the displacement $\boldsymbol{u}_i$ from Equation 14 after each solve. In matrix notation, we can write,

$$\begin{bmatrix} \rho \boldsymbol{M} - \mu \delta_{N_t}^2 \boldsymbol{K} & \delta_{N_t} \boldsymbol{B}^T & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ -\delta_{N_t} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{p}_i \\ \boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} \rho \boldsymbol{M} & \boldsymbol{0} & \mu \delta_{N_t} \boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{i-1} \\ \boldsymbol{p}_{i-1} \\ \boldsymbol{u}_{i-1} \end{bmatrix}. \tag{24}$$

With

$$\boldsymbol{D}^{II} := \begin{bmatrix} \rho \boldsymbol{M} - \mu \delta_{N_t}^2 \boldsymbol{K} & \delta_{N_t} \boldsymbol{B}^T & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ -\delta_{N_t} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{C}^{II} := \begin{bmatrix} \rho \boldsymbol{M} & \boldsymbol{0} & \mu \delta_{N_t} \boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \tag{25}$$

equation 24 can be written as,

$$\boldsymbol{s}_i^{II} = \boldsymbol{\Phi}^{II} \boldsymbol{s}_{i-1}^{II} \qquad \text{for } i = 1, \ldots, N_t, \tag{26}$$

where $\boldsymbol{\Phi}^{II} := [\boldsymbol{D}^{II}]^{-1} \boldsymbol{C}^{II}$ and $\boldsymbol{s}_i^{II} := [\boldsymbol{v}_i, \boldsymbol{p}_i, \boldsymbol{u}_i]^T$ with $\boldsymbol{s}_0^{II} = \hat{\boldsymbol{s}}_0 := [\hat{\boldsymbol{v}}_0, 0, \boldsymbol{0}]^T$. Again, the operator $\boldsymbol{\Phi}^{II}$ only depends on time step size δ_{N_t}.

2.5. *Multigrid reduction in time (MGRIT) algorithm*

Based on Equation 21 and Equation 26, the global space-time problem can be written in the linear form,

$$\boldsymbol{A}^S \boldsymbol{s}^S = \begin{bmatrix} \boldsymbol{I} & & & & \\ -\boldsymbol{\Phi}^S & \boldsymbol{I} & & & \\ & -\boldsymbol{\Phi}^S & \boldsymbol{I} & & \\ & & \ddots & \ddots & \\ & & & -\boldsymbol{\Phi}^S & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_0^S \\ \boldsymbol{s}_1^S \\ \boldsymbol{s}_2^S \\ \vdots \\ \boldsymbol{s}_{N_t}^S \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{s}}_0 \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix} = \hat{\boldsymbol{s}}, \tag{27}$$

with $S \in \{I, II\}$. A traditional time stepping method would solve Equation 27 in a block-forward fashion, whereas the MGRIT algorithm solves Equation 27 iteratively. Both algorithms are $O(N_t)$, however, the constant is bigger for MGRIT [15, 12]. On the other hand, MGRIT enables parallelism in the temporal domain in contrast to the sequential nature of traditional time stepping. The parallelism is achieved by introducing a time grid hierarchy and applying multigrid techniques in the temporal domain. The coarser time grids provide error corrections to the finest time grid, thus accelerating convergence to the solution; while a relaxation process on each time grid reduces the error that cannot be adequately reduced on coarser grids. These two complementary processes, relaxation and coarse-grid correction, form the core of multigrid methods.

Figure 1 illustrates an example of an m_l-level time grid hierarchy with respective time points, time step sizes and coarsening factors. Here, the maximum number of time grid levels is given as $m_l = 2$ (with levels $m = 0, \ldots, m_l - 1$). Each level with $m < m_l - 1$ is composed of F- and C-points. F-points only exist on level m and the C-points compose the coarser grid $m + 1$. The coarsening factor c_{m-1}^m (for $m = 1, \ldots, m_l - 1$) defines a coarsening of a given level $m - 1$, such that the time step size on level m is $\delta_{N_t^m} = c_{m-1}^m \cdot \delta_{N_t^{m-1}}$ (for $m = 1, \ldots, m_l - 1$), with fine grid time step size $\delta_{N_t^0}$ and the coarsest grid's time step size $\delta_{N_t^{m_l-1}}$. Similarly, $c_0^{m_l-1}$ denotes the total coarsening from level 0 to level m_l. Further, F-relaxation means an update of the F-points based on the previous C-points and C-relaxation refers to an update of a C-point based on the previous F-point. For more details, see [15].

Prepared using nlaauth.cls
DOI: 10.1002/nla

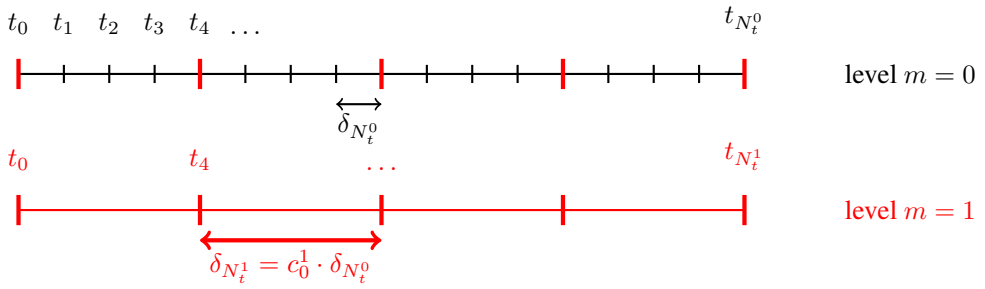


Figure 1. Example of an m_l-level multigrid hierarchy with $m_l = 2$ time grid levels, respective time points t_i, time step sizes $\delta_{N_t^m}$ and coarsening factors c_{m-1}^m.

For the remainder of this work, we distinguish between the numerical space-time solution obtained with Scheme I with sequential time stepping with a subscript s (sequential) and those quantities obtained from using MGRIT with a subscript p (parallel), *e.g.*, the solution vector $\boldsymbol{s}^I$ obtained from sequential time stepping is denoted as $\boldsymbol{s}_s^I$, whereas the solution vector obtained with MGRIT is denoted as $\boldsymbol{s}_p^I$. Similarly, we identify the numerical solution obtained with Scheme II by using the notation $\boldsymbol{s}_s^{II}$ and $\boldsymbol{s}_p^{II}$.

Further, we consider two-grid algorithms using V-cycles with various numbers of FC-relaxation steps and multilevel algorithms using full multigrid (FMG) cycles (*i.e.*, F-cycles) as described in the following. Unless noted otherwise, we either use the one-step integrator $\boldsymbol{\Phi}_{\delta_t^0}^I$ on the fine grid and its coarse grid versions $\boldsymbol{\Phi}_{\delta_t^m}^I$ on grid level m or similarly, $\boldsymbol{\Phi}_{\delta_t^0}^{II}$ on the fine grid and its coarse grid versions $\boldsymbol{\Phi}_{\delta_t^m}^{II}$ on grid level m. The subscript δ_t^m denotes $\delta_{N_t^m}$. In Section 3.3, we will also motivate the design of a hybrid algorithm that employs $\boldsymbol{\Phi}_{\delta_t^0}^I$ on the fine grid and $\boldsymbol{\Phi}_{\delta_t^m}^{II}$ on the coarse grid in a two-level time grid hierarchy.

2.5.1. *Two-level algorithm using V-cycles*

Firstly, a two-grid version of the MGRIT algorithm is considered. In this version, r FC-relaxations ($r = 0, 1, 2$) are applied on the fine grid for Scheme I and Scheme II. For $r = 0$, MGRIT is equivalent to Parareal [13]. To restrict the variables from the fine to the coarse grid, a restriction operator $R(\cdot)$ is chosen that purely injects the respective fine grid quantities to the coarse grid. To transfer the coarse grid quantities to the fine grid, an ideal interpolation is employed [15]. The pseudo-code of the linear two-grid version of the MGRIT algorithm used here is described in Algorithm 1. [†] The coarse-grid operator $\boldsymbol{A}_\Delta^S$ is equivalent to that defined in (27) except it uses the coarse-grid time-stepper $\boldsymbol{\Phi}_{\delta_t^{m+1}}^S$ and is defined over the smaller coarse time-grid depicted in Figure 1.

The two-grid algorithm with F- and FCF- relaxation will be the basis for predicting and analyzing convergence of the proposed algorithm and for comparing the observed convergence factors with the theoretical upper bounds derived from [12]. The presented analysis will then prove and emphasize the relevance of [12] for the design of convergent parallel-in-time algorithms and motivate the selection of Scheme II to be included in a multilevel hierarchy. In this setting, we also investigate the benefits of switching from V-cycles to F-cycles.

2.5.2. *Multilevel algorithm using FMG-cycles*

In addition to the two-level algorithms, we consider m_l-grid algorithms ($m_l > 2$) with FMG-cycles. A FMG-cycle is achieved by applying Step 4 of Algorithm 1 recursively and performing one V-cycle as post-relaxation at each level [16].

[†]The nonlinear version of MGRIT is used in practice and is equivalent to the linear version for linear problems [14].

DOI: 10.1002/nla

Algorithm 1 Pseudo-code for MGRIT algorithm with V-cycles and $S \in \{I, II\}$, adapted from [12].

1: **repeat**
2: Relax on $\boldsymbol{A}^S \boldsymbol{s}^S = \hat{\boldsymbol{s}}$ (r times FC-relaxation, one F-relaxation) using $\boldsymbol{\Phi}^S_{\delta_t^m}$. ▷ In parallel
3: Compute coarse-grid residual $\boldsymbol{r}_\Delta = R(\hat{\boldsymbol{s}} - \boldsymbol{A}^S \boldsymbol{s}^S)$.
4: Solve the coarse grid correction problem $\boldsymbol{A}^S_\Delta \boldsymbol{e}^S_\Delta = \boldsymbol{r}_\Delta$ using $\boldsymbol{\Phi}^S_{\delta_t^{m+1}}$. ▷ Apply recursively
5: Correct the solution at the fine-grid C-points with $\boldsymbol{e}^S_\Delta$. ▷ In parallel
6: **until** norm of residual is small enough
7: Update the solution at the F-points with $\boldsymbol{\Phi}^S_{\delta_t^0}$. ▷ In parallel

2.5.3. *Theoretical two-grid reduction of the MGRIT residual norm per iteration in the two-grid case*

The theoretical two-grid reduction rate estimates of the global space-time error per MGRIT iteration are based on the two-grid cases discussed in [12]. Firstly, we note that $\boldsymbol{\Phi}^S_{\delta_t^0}$ (and the coarse-grid version $\boldsymbol{\Phi}^S_{\delta_t^1}$) have the sparsity pattern,

$$\begin{bmatrix} [\boldsymbol{\Phi}^S_{\delta_t^m}]_{11} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{12} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{13} \\ [\boldsymbol{\Phi}^S_{\delta_t^m}]_{21} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{22} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{23} \\ [\boldsymbol{\Phi}^S_{\delta_t^m}]_{31} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{32} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{33} \end{bmatrix} = \begin{bmatrix} [\boldsymbol{\Phi}^S_{\delta_t^m}]_{11} & \boldsymbol{0} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{13} \\ [\boldsymbol{\Phi}^S_{\delta_t^m}]_{21} & \boldsymbol{0} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{23} \\ [\boldsymbol{\Phi}^S_{\delta_t^m}]_{31} & \boldsymbol{0} & [\boldsymbol{\Phi}^S_{\delta_t^m}]_{33} \end{bmatrix}, \tag{28}$$

with $S \in \{I, II\}$ and $m = 0, 1$. This highlights that the current state vector $\boldsymbol{s}^S_i$ is not dependent on the previous pressure value (*i.e.*, the pressure variable and its associated rows and columns in $\boldsymbol{\Phi}$ can be ignored without affecting $\boldsymbol{u}$ or $\boldsymbol{v}$.) Thus, we eliminate the rows and columns related to the pressure variable and proceed with the analysis by simultaneously diagonalizing the time stepping matrix,

$$\boldsymbol{T}^{-1} \begin{bmatrix} [\boldsymbol{\Phi}^S_{\delta_t^0}]_{11} & [\boldsymbol{\Phi}^S_{\delta_t^0}]_{13} \\ [\boldsymbol{\Phi}^S_{\delta_t^0}]_{31} & [\boldsymbol{\Phi}^S_{\delta_t^0}]_{33} \end{bmatrix} \boldsymbol{T} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots), \tag{29}$$

$$\boldsymbol{T}^{-1} \begin{bmatrix} [\boldsymbol{\Phi}^S_{\delta_t^1}]_{11} & [\boldsymbol{\Phi}^S_{\delta_t^1}]_{13} \\ [\boldsymbol{\Phi}^S_{\delta_t^1}]_{31} & [\boldsymbol{\Phi}^S_{\delta_t^1}]_{33} \end{bmatrix} \boldsymbol{T} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots). \tag{30}$$

Then, the authors of [12] prove that the global space-time error vector at the C-points can be reduced in the mass matrix norm for the two-grid case with F- and FCF-relaxation by a convergence factor of at least,

$$c_f^F = \max_\omega \left\{ |\lambda_\omega^{c_0^1} - \gamma_\omega| \frac{1 - |\gamma_\omega|^{N_t^1}}{1 - |\gamma_\omega|} \right\}, \tag{31}$$

$$c_f^{FCF} = \max_\omega \left\{ |\lambda_\omega^{c_0^1} - \gamma_\omega| \frac{1 - |\gamma_\omega|^{N_t^1 - 1}}{1 - |\gamma_\omega|} |\lambda_\omega|^{c_0^1} \right\}, \tag{32}$$

$$\tag{33}$$

respectively.

2.5.4. *Computation of MGRIT residual norm in experiment*

The standard Euclidean norm of the MGRIT residual in numerical experiments is computed using the coarse-grid residual $\boldsymbol{r}_\Delta$ (see Step 3 of Algorithm 1),

$$\begin{aligned} \|\boldsymbol{r}_\Delta\|_2^2 &= \left(w_0 \|\boldsymbol{r}^{\boldsymbol{v}}_\Delta\|_2^2 + w_1 \|\boldsymbol{r}^p_\Delta\|_2^2 + w_2 \|\boldsymbol{r}^{\boldsymbol{u}}_\Delta\|_2^2 \right)^{1/2} \\ &= \left(w_0 (\boldsymbol{r}^{\boldsymbol{v}}_\Delta)^T \boldsymbol{r}^{\boldsymbol{v}}_\Delta + w_1 (\boldsymbol{r}^p_\Delta)^T \boldsymbol{r}^p_\Delta + w_2 (\boldsymbol{r}^{\boldsymbol{u}}_\Delta)^T \boldsymbol{r}^{\boldsymbol{u}}_\Delta \right)^{1/2} \end{aligned} \tag{34}$$

Prepared using nlaauth.cls
DOI: 10.1002/nla

where $\boldsymbol{r}_\Delta^{\boldsymbol{v}}$, $\boldsymbol{r}_\Delta^{p}$ and $\boldsymbol{r}_\Delta^{\boldsymbol{u}}$ denote components in the residual vector corresponding to velocity, pressure and displacement. Further, the weights w_0, w_1 and w_3 are equal to 1 in the standard case.

However, the predicted decrease of the residual is measured in the mass matrix norm

$$\|\boldsymbol{r}_\Delta\|_{\mathfrak{M}}^2 = \|\boldsymbol{r}_\Delta^{\boldsymbol{v}}\|_{\mathfrak{M}^{\boldsymbol{v}}}^2 + \|\boldsymbol{r}_\Delta^{\boldsymbol{p}}\|_{\mathfrak{M}^{\boldsymbol{p}}}^2 + \|\boldsymbol{r}_\Delta^{\boldsymbol{u}}\|_{\mathfrak{M}^{\boldsymbol{u}}}^2, \tag{35}$$

where $\mathfrak{M}^{\boldsymbol{x}}$ is a block $N_t^1 \times N_t^1$ diagonal matrix with each block corresponding to a spatial mass matrix for variable $\boldsymbol{x}$ (see Section 2.5.3 and [12]). Thus, a modification of measuring the observed reduction of the residual norm per MGRIT iteration is advisable, and hence, proposed: instead of computing the computationally more demanding mass matrix norm and the less accurate standard Euclidean norm, the following weights to improve the measurement of experimental convergence factors are chosen: $w_0 = 1$, $w_1 = 0$ and $w_2 = 1/c_0^1$. This yields

$$(\boldsymbol{r}_\Delta^{\boldsymbol{v}})^T \mathfrak{M}^{\boldsymbol{v}} \boldsymbol{r}_\Delta^{\boldsymbol{v}} \approx (\boldsymbol{r}_\Delta^{\boldsymbol{v}})^T \boldsymbol{I} \boldsymbol{r}_\Delta^{\boldsymbol{v}}, \tag{36}$$

$$(\boldsymbol{r}_\Delta^{\boldsymbol{u}})^T \mathfrak{M}^{\boldsymbol{u}} \boldsymbol{r}_\Delta^{\boldsymbol{u}} \approx (\boldsymbol{r}_\Delta^{\boldsymbol{u}})^T \left(\frac{1}{c_0^1}\boldsymbol{I}\right) \boldsymbol{r}_\Delta^{\boldsymbol{u}}, \tag{37}$$

Thus, the approximate value of the mass matrix norm of the residual at the C-points is given by,

$$\|\boldsymbol{r}_\Delta\|_{\mathfrak{M}} \approx \left((\boldsymbol{r}_\Delta^{\boldsymbol{v}})^T \boldsymbol{r}_\Delta^{\boldsymbol{v}} + \frac{1}{c_0^1} (\boldsymbol{r}_\Delta^{\boldsymbol{u}})^T \boldsymbol{r}_\Delta^{\boldsymbol{u}} \right)^{1/2}. \tag{38}$$

It is important to note that this proposition only changes how the solution progress (*i.e.*, reduction of residual norm) is measured. But it does neither affect the coarse-grid update nor does it change the numerical solution.

2.6. *Implementation details*

Scheme I and Scheme II were implemented in the finite element software tool CHeart [17]. Wrapper routines were written to incorporate the MGRIT algorithm into CHeart using the open-source library XBraid [18], a non-intrusive implementation of the MGRIT algorithm.

Here, we introduced separate MPI groups and communicators to maintain the capability of CHeart to parallelize in the spatial domain by using domain decomposition methods while enabling independent parallelization in the temporal domain. That is, one can parallelize in space, time or in space-time.

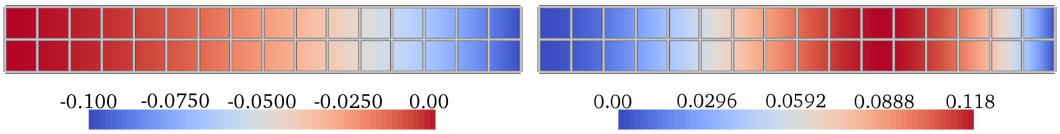
Furthermore, a given relaxation step is performed even if the initial spatial residual satisfies the convergence criteria for the spatial problem. Moreover, the XBraid option to skip work on the first down-cycle is used. Note, due to the linearity of the problem, the matrix $\boldsymbol{\Phi}_{\delta_t^m}^S$ is only computed once for each time step size (*i.e.*, time grid level). This significantly reduces computational work compared to re-computing the matrix for each time step.

3. RESULTS

In the following, the methods described in Section 2 are used for a linear beam problem. For this purpose, consider the domain $\Omega(0) = [0, 8] \times [0, 1]$, final time $T = 1024$ and material parameters $\mu = \rho = 1$. The spatial discretization uses 16 x 2 quadrilateral elements and initial conditions are given as $\hat{\boldsymbol{v}}_0 = [-x^2/640, x^2 \cdot (8 - x)/640]^T$, see Figure 2. The initial condition is also used as initial guess at all time steps for MGRIT. Further, no displacement boundary conditions are prescribed at $\Gamma^D = \Omega|_{x=0}$.

We consider fine grid time step sizes $\delta_{N_t^0} \in \{1, 0.1, 0.01, 0.001\}$ and coarsening factors $c_0^1 \in \{2, 4, 8, 16, 32\}$ and $c_{m-1}^m = 2$ for $2 \leq m \leq m_l$. The stopping criterion on the residual norm is selected as $\|\boldsymbol{r}_\Delta\|_{\mathfrak{M}} \leq 10^{-8}/\sqrt{\delta_{N_t^0}}$ (see Equation 38) with the maximum number of MGRIT iterations set to 60 iterations. Unless noted otherwise, reported experimental convergence factors are the global maximum values.

DOI: 10.1002/nla



(a) Initial velocity in x-direction. (b) Initial velocity in y-direction.

Figure 2. Discretization using 16 x 2 quadrilateral elements and initial velocity.

3.1. Numerical solution of beam oscillation

The initial velocity distribution over the cantilever beam length (see Figure 2) causes the free end to first deflect in negative x- and positive y-direction, see Figure 3. The elastic stresses cause the beam to decelerate and move downward, passing its initial position and deflecting in negative y-direction. The beam deformation follows a repeatable deflection pattern.

3.2. Effect of time step size on amplitude of oscillation

Backward Euler time integration schemes introduce artificial numerical diffusion. Its effect on the sum of kinematic and potential energy of the system over time depends on the time step size where we expect energy conservation in the asymptotic limit $\delta_{N_t^0} \to 0$. The artificial damping of the system causes the amplitude of oscillation to become smaller over time. The backward Euler time integration scheme exhibits quick damping for $\delta_{N_t^0} = 1$ irrespectively of the chosen scheme (Scheme I and Scheme II). The effect of numerical damping becomes smaller for smaller $\delta_{N_t^0}$, where we note that both schemes reproduce the amplitudes of the beam oscillation with comparable quality for $\delta_{N_t^0} = 0.001$, see Figure 4.

It is important to note, that the varying magnitude of numerical damping is also present with respect to varying time step sizes in an MGRIT time grid hierarchy. Nevertheless, MGRIT converges to the same solution (within the selected solver tolerance) as sequential time stepping on the fine grid. Thus, the converged numerical solution obtained with the MGRIT algorithm suffers from the same amount of numerical damping as the numerical solution from sequential time stepping on the fine grid.

3.3. Convergence in the two-grid case

3.3.1. MGRIT with Scheme I

If Scheme I is employed as a one-step integrator on the fine and coarse grid (that is, $\mathbf{\Phi}^I_{\delta_t^0}$ and $\mathbf{\Phi}^I_{\delta_t^1}$) in a two-level algorithm, we observe divergence in the numerical experiments for all considered test cases.

To qualitatively investigate what leads to divergence of MGRIT in conjunction with Scheme I, we track the current position of the tip of the cantilever beam (initial coordinate $[8, 0.5]^T$ at $t = 0$) over time with smaller time horizon $t \in [0, 64]$. The time step size is $\delta_{N_t^0} = 1$ with coarsening factor $c_0^1 = 2$. FCF-relaxation is employed. Figure 5 shows the current approximation of the tip's displacement with respect to the initial position over time for a number of algorithmic steps, for example, after FCF-relaxation, after restriction, after the coarse-grid solve, *etc.* (where we extract the values on return from applying $\mathbf{\Phi}^I_{\delta_t^0}$ and $\mathbf{\Phi}^I_{\delta_t^1}$, respectively). The data in Figure 5 highlight how the current approximation of the cantilever's tip first improves. Though, already during the first MGRIT iteration an instability is introduced by the coarse-grid update which is then amplified in subsequent steps. The observed phenomena are neither physical nor part of the mathematical model, however, are in line with observations in literature, see [9, 10]. Previously this has been a limiting factor for parallel-in-time integration methods of the dynamic elasticity equation.

On the other hand, divergence of the numerical algorithm is reflected in very large theoretical convergence bounds, *i.e.*, $c_f^F, c_f^{FCF} >> 1$, confirming experimental observations.

DOI: 10.1002/nla

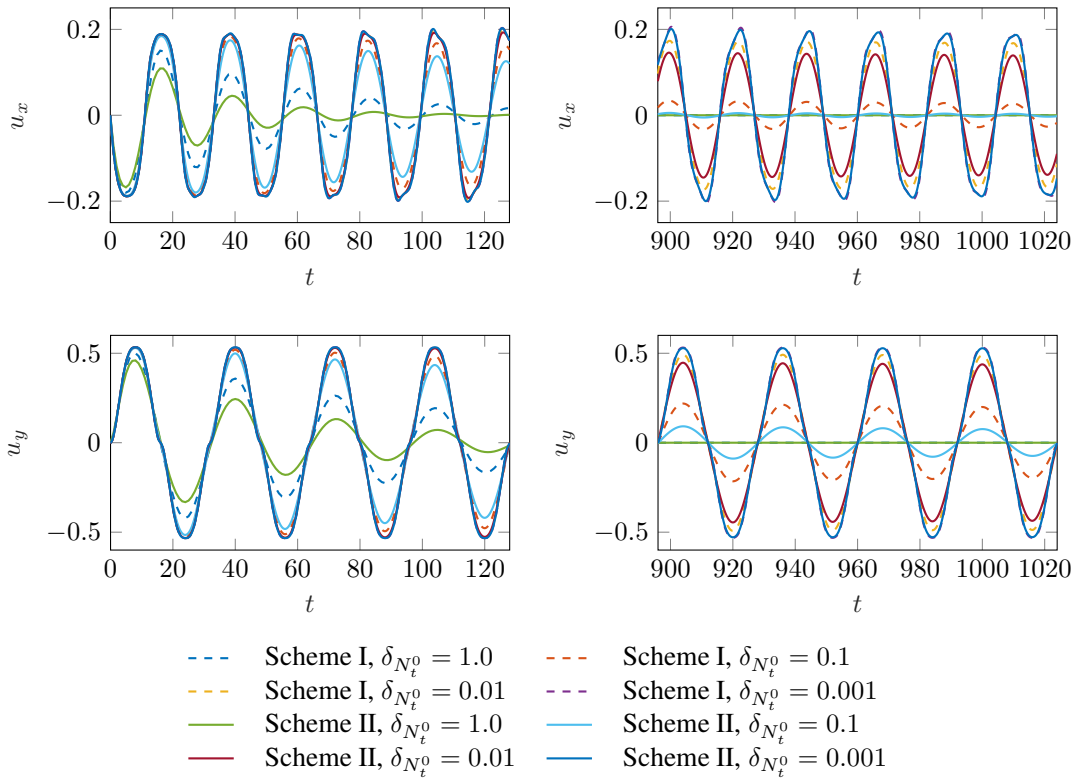


Figure 3. Displacement $\boldsymbol{u}$ of the tip of the cantilever beam (initial position $[8, 0.5]^T$) in x- and y-direction for $t \in [0, 128]$ and $[896, 1024]$ for Scheme I and Scheme II and time step sizes $\delta_{N_t^0} \in \{1, 0.1, 0.01, 0.001\}$. Note that numerical damping is reduced with $\delta_{N_t^0} \to 0$.

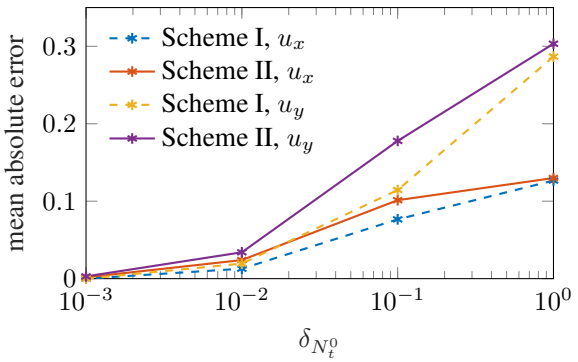


Figure 4. Mean absolute error of the displacement of the beam's tip for $t \in [0, 1024]$. Here, Scheme I with $\delta_{N_t^0} = 0.001$ was used as reference.

3.3.2. MGRIT with Scheme II

Using Scheme II as fine- and coarse-grid integrators (that is, $\boldsymbol{\Phi}_{\delta_t^0}^{II}$ and $\boldsymbol{\Phi}_{\delta_t^1}^{II}$) in a two-level algorithm, we observe good convergence for a range of different coarsening factors c_0^1 and for all considered time step sizes, see Figure 6.

We note that for $\delta_{N_t^0} = 1$, both experimental and predicted convergence factors are in excellent agreement and that the predicted values of c_f^F and c_f^{FCF} are a sharp upper bound despite the approximation of the computed residual norm, see Section 2.5.4. Here, the convergence factors

DOI: 10.1002/nla

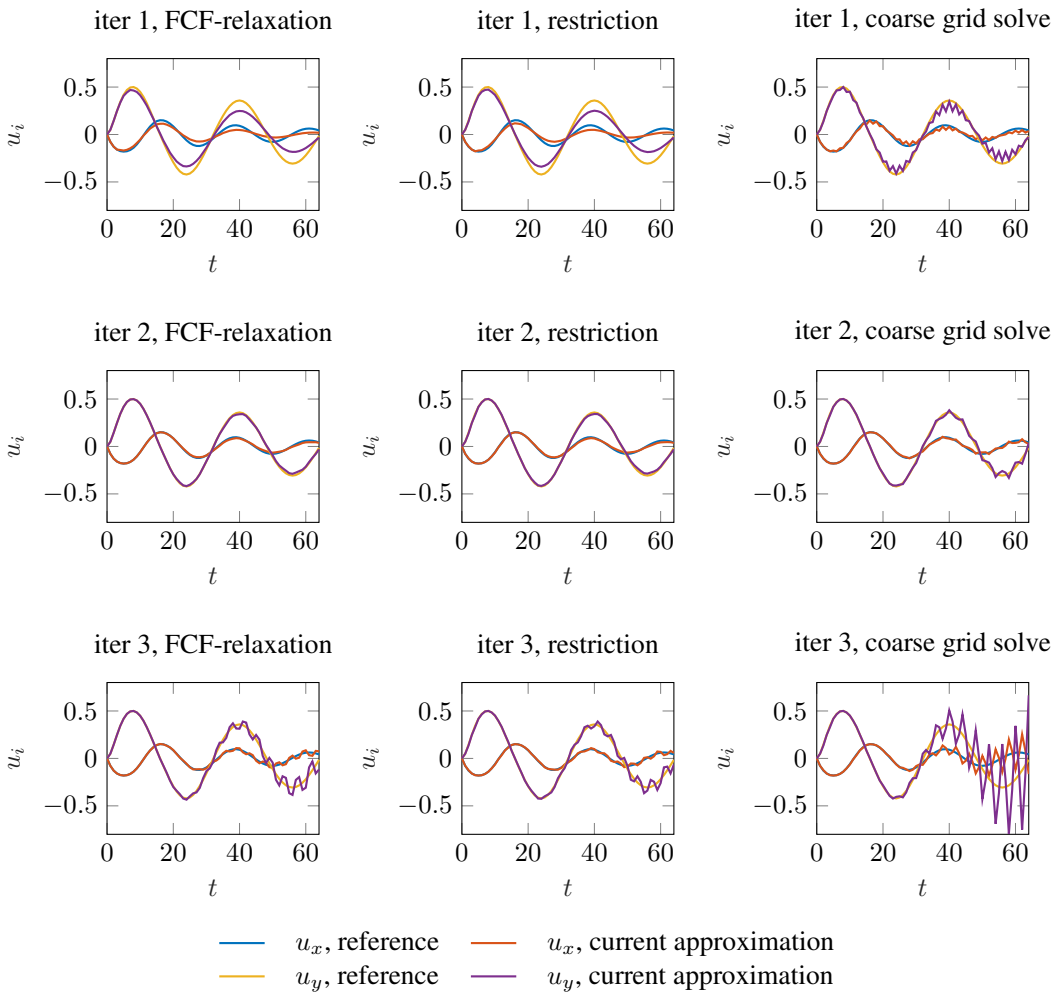


Figure 5. Current approximation of the tip displacement for two-grid MGRIT using Scheme I with coarsening factor $c_0^1 = 2$, fine grid time step size $\delta_{N_t^0} = 1$, $t \in [0, 64]$ and FCF-relaxation compared with a reference solution from sequential time stepping. Note, how the coarse grid update introduces an instability which is amplified in subsequent steps.

first increase with increasing coarsening factor c_0^1 before decreasing due to the small coarse-grid size $N_t^{m_l-1}$.

On the other hand, for $\delta_{N_t^0} \in \{0.1, 0.01, 0.001\}$ we observe an increase in predicted and observed convergence factors with increasing coarsening factor c_0^1. Again, predicted and observed convergence factors are in excellent agreement for almost all considered cases. Only for $\delta_{N_t^0} = 0.01$ the maximum observed convergence factor is smaller for the F-relaxation case and approximately 1% larger for (i) F-relaxation and $c_0^1 = 32$ and (ii) FCF-relaxation and $c_0^1 \in \{16, 32\}$. Again, this is likely due to the approximate residual norm computation.

Further, we note that additional relaxation steps can be beneficial for small fine grid time step sizes, whereas the effect is negligible for $\delta_{N_t^0} = 0.001$, thus suggesting that relaxation can be omitted for small fine grid time step sizes to reduce computational work without sacrificing convergence.

Finally, we note that in all considered cases we do not observe any instability as described in Section 3.3.1 and previous works [9, 10]. For example, Figure 7 illustrates the position of the tip of the cantilever after the first three MGRIT iteration with $t \in [0, 64]$, $\delta_{N_t^0} = 0.1$ and $c_0^1 = 2$. Here, no artificial amplification of the amplitude of oscillation is observed as is the case for MGRIT with Scheme I, see Figure 5.

Prepared using nlaauth.cls
DOI: 10.1002/nla

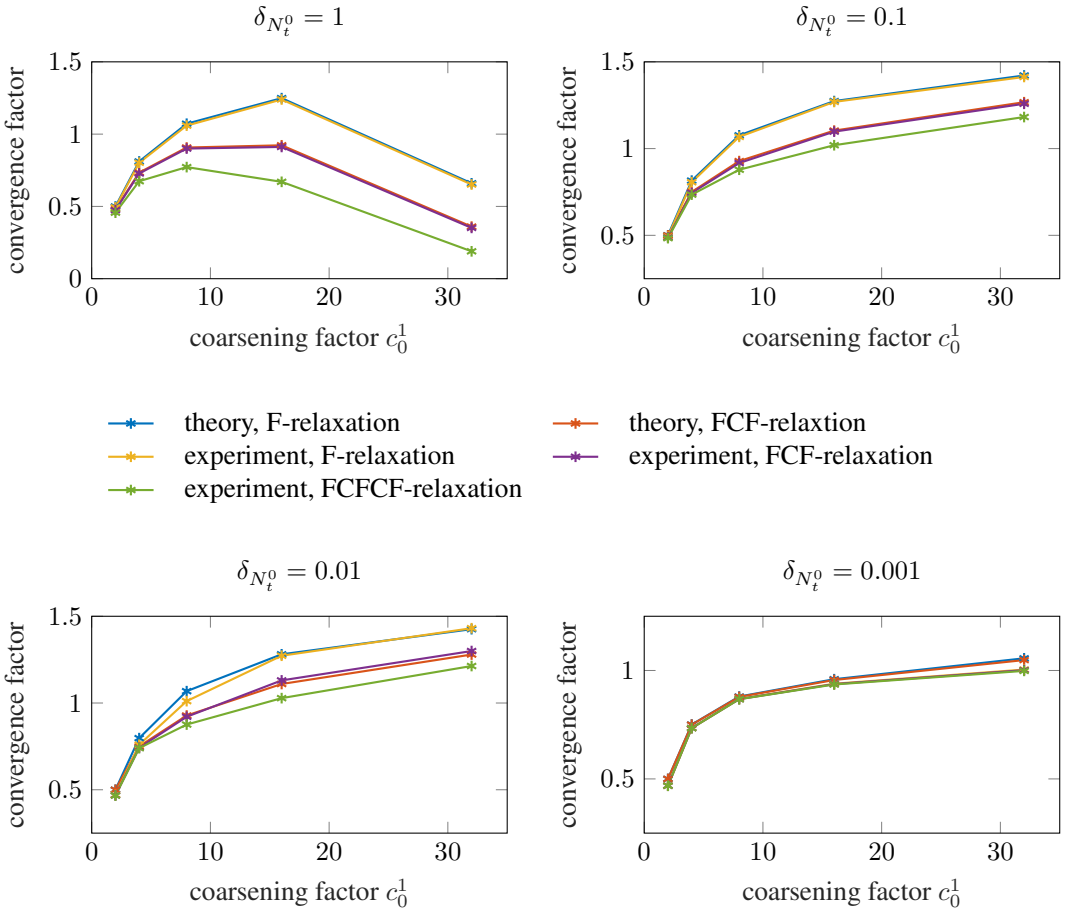


Figure 6. Predicted and measured convergence factor for a two-grid algorithm with various fine grid time step sizes $\delta_{N_t^0} \in \{1.0, 0.1, 0.01, 0.001\}$ and coarsening factors $c_0^1 \in \{2, 4, 8, 16, 32\}$.

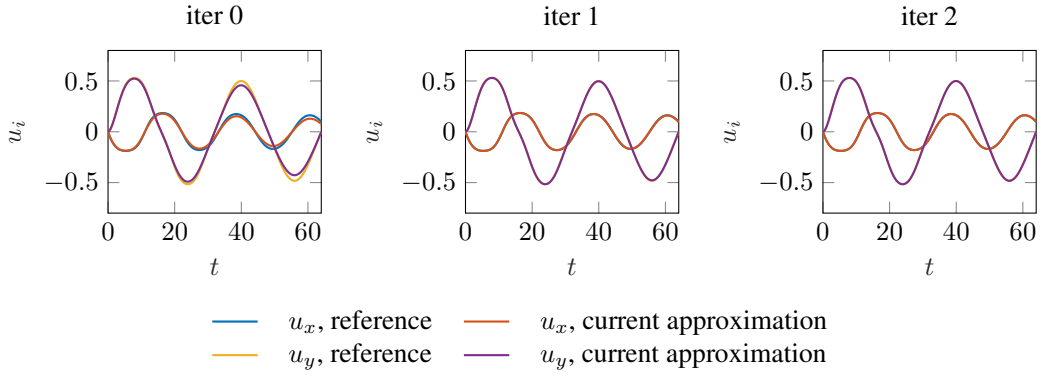


Figure 7. Current approximation of the tip displacement for two-grid MGRIT using Scheme II with coarsening factor $c_0^1 = 2$, fine grid time step size $\delta_{N_t^0} = 0.1$, $t \in [0, 64]$ and FCF-relaxation compared with reference solution from sequential time stepping. No instability is observed with Scheme II in contrast to using Scheme I, see Figure 5.

The results in this section highlight the benefit of using theoretical upper bounds as given in [12] as tool to estimate experimental convergence *a priori* and to design convergent MGRIT algorithms with guaranteed worst case convergence factors.

DOI: 10.1002/nla

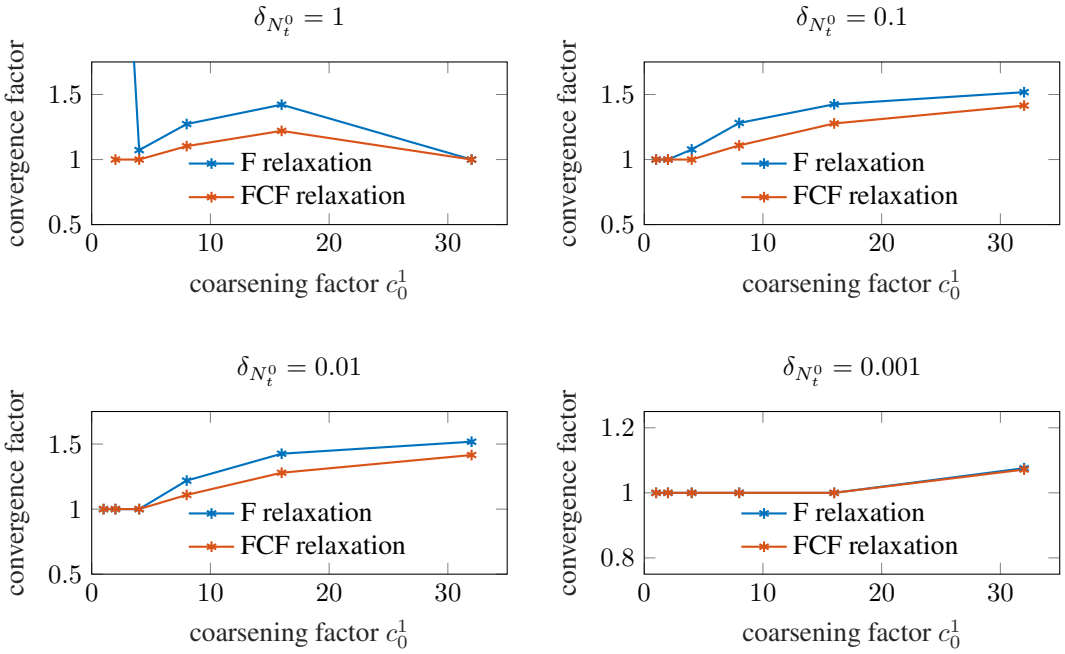


Figure 8. Predicted worst-case convergence for MGRIT with the hybrid approach. The hybrid approach avoids the instability observed for MGRIT with Scheme I, however, cannot yield guaranteed convergence in the two-grid case.

3.3.3. MGRIT with hybrid scheme

Section 3.2 highlights that for large time step sizes, Scheme I can better reproduce the amplitude of oscillation compared to Scheme II. Though, Scheme I with MGRIT does not yield a convergent parallel-in-time algorithm. On the other hand, Scheme II with MGRIT does not exhibit the same instability as Scheme I with MGRIT, see Section 3.3.2. Additionally, MGRIT yields the same solution as sequential time stepping on the fine grid within the selected solver tolerance, as was noted earlier. These aspects motivate the design of a hybrid algorithm with the aim to preserve amplitudes as good as Scheme I but with the MGRIT convergence properties of Scheme II. Thus, the hybrid scheme employs $\mathbf{\Phi}^I_{\delta_t^0}$ and $\mathbf{\Phi}^{II}_{\delta_t^1}$ on the fine and coarse grid, respectively.

Although MGRIT with the hybrid scheme does not suffer from the same instability as MGRIT with Scheme I, convergence deteriorates. The predicted upper bound for the convergence factor either indicates divergence with $c_f^F, c_f^{FCF} > 1$, in the worst case, or a stalling solver with $c_f^F, c_f^{FCF} = 1$. This is illustrated in Figure 8, where one can observe that convergence cannot be guaranteed for all fine grid time step sizes. Thus, MGRIT with the hybrid scheme has no significance for designing robust parallel-in-time solvers in the two-grid case.

3.4. Convergence in the multigrid case

Since a convergent algorithm could only be achieved for MGRIT with Scheme II in the two-grid case, only this case is considered in the following to investigate convergence in the multigrid case.

3.4.1. MGRIT with Scheme II

To explore convergence in the multigrid case, we employ multilevel hierarchies with $m_l \in \{3, 4, 5\}$ for $\delta_{N_t^0} \in \{1, 0.1, 0.01\}$ and $m_l \in \{3, 4, 5, 6\}$ for $\delta_{N_t^0} = 0.001$. Further, we consider combined coarsening factors of $c_0^{m_l - 1} \in \{4, 8, 16, 32, 64\}$, where we only vary c_0^1 but select $c_{m-1}^m = 2$ for $m = 2, \ldots, m_l - 1$. The MGRIT algorithm is started with a forward solve on the coarsest time grid (*i.e.* use of the XBraid skip-first-down option), performs an initial V-cycle and full multigrid (FMG)

DOI: 10.1002/nla

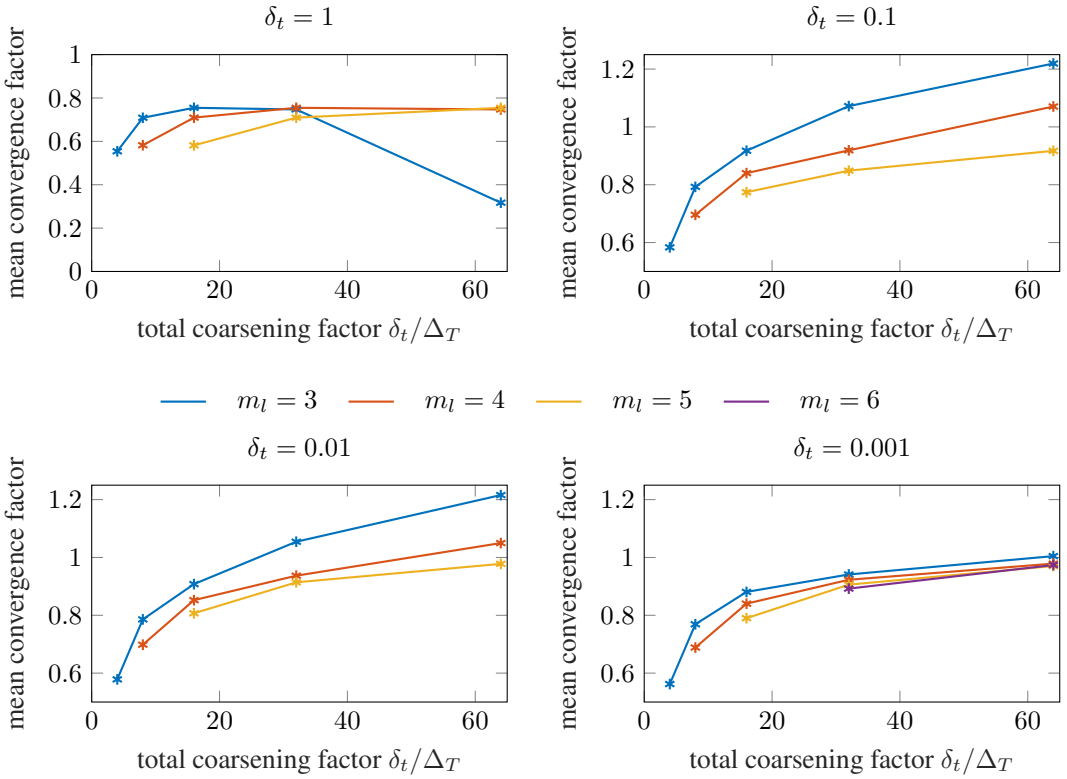


Figure 9. Measured mean convergence factor for m_l-grid MGRIT algorithm with F-cycles and Scheme II for fine grid time step sizes $\delta_{N_t^0} \in \{1, 0.1, 0.01, 0.001\}$. Note, that the x-axis corresponds to the combined coarsening factor $c_0^{m_l-1}$.

cycles for all following MGRIT iterations. Scheme II is employed as one-step integrator on all grid levels and one V-cycle is performed as post-relaxation at each FMG level. To provide a better indicator for overall performance of the MGRIT algorithm, we report the mean of the experimental convergence factor over all MGRIT iterations.

Figure 9 reports mean experimental convergence factors over the combined coarsening factor $c_0^{m_l-1} = \prod_{m=0}^{m=m_l-2} c_m^{m+1}$. As illustrated, the best convergence factor available for a particular combined coarsening factor is available through the use of FMG cycles with more levels and slower coarsening, between levels, as opposed to FMG cycles with more aggressive coarsening between levels. Observed convergence factors are generally significantly smaller than in the two-grid case. Thus, the use of FMG-cycles allows more aggressive coarsening and yields more potential for parallelism in the temporal domain.

If $\delta_{N_t^0}$ and $c_0^{m_l-1}$ are kept fixed, the mean convergence factor improves with increasing m_l. For example, for $\delta_{N_t^0} = 0.1$ and $c_0^{m_l-1} = 64$ the mean convergence factor is approximately 0.92 for $m_l = 5$ but 1.07 for $m_l = 4$ and 1.22 for $m_l = 3$. Thus, one can obtain a moderately convergent instead of a slowly divergent algorithm simply by introducing an additional intermediate time grid level but with the same fine and coarsest grid size.

3.4.2. Timing results

In this section, we present preliminary speedup results for an MGRIT algorithm that employs Scheme II and $m_l = 5$ time grid levels with a coarsening of two between all time grid levels. Similar to Section 3.4.1, an initial V-cycle is performed before switching to FMG-cycles with one V-cycle as post-relaxation at each FMG-level. The time step size is selected as $\delta_{N_t^0} = 0.001$. The standard Euclidean norm is used to measure solution progress. To investigate the dependency of the wall

DOI: 10.1002/nla

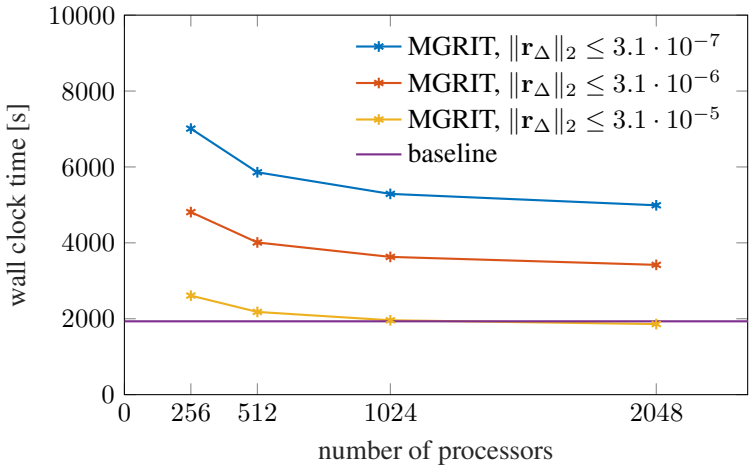


Figure 10. Wall clock time for MGRIT with Scheme II and FMG-cycles with $m_l = 5$ and $c^m_{m-1} = 2$ for all m. Note, that even for small processor counts a marginal speedup can be achieved with relaxed convergence criteria besides the relatively large coarse grid size with $c_0^{m_l-1} = 16$.

clock time for MGRIT on the convergence criterion, we vary the tolerance between $3.1 \cdot 10^{-7}$ and $3.1 \cdot 10^{-5}$.

All reported timing results were obtained on CAB (Intel Xeon E5-2670 architecture, 16 cores 2.6GHz and 32GB RAM per node) at Lawrence Livermore National Laboratory, CA, USA. Data export was switched off and status messages were restricted to a bare minimum to ensure that reported wall clock times are almost entirely dominated by computation and communcation costs but not data I/O tasks.

The elapsed wall clock time for the sequential time stepping algorithm (using Scheme II) was 1932 seconds, which is established as baseline in Figure 10. As illustrated in Figure 10, the elapsed wall clock time for the five-grid MGRIT algorithm drops from approximately 7010 seconds to 4990 seconds when using 2048 processors instead of 256 processors for parallelizing in the temporal domain. By relaxing the convergence criterion for the time-parallel algorithm, the wall clock time of the time-parallel algorithm drops below the wall clock time for the sequential algorithm with a marginal speedup. Here, it is important to note that although the achieved speedup is almost negligible, one can expect to improve the parallel speedup, for example, by selecting a more efficient cycling and relaxation strategy (*e.g.*, cheaper MGRIT iterations and improved reduction of residual norm per MGRIT iteration) and a larger coarsening factor $c_0^{m_l-1}$ in combination with growing processor counts.

4. DISCUSSION

In the previous sections, two different backward Euler time discretization schemes were presented and investigated. It was shown, that MGRIT with Scheme I exhibits strong instabilities parallel-in-time. The observations are in line with previous works in this field [9, 10]. Although MGRIT with Scheme II uses only a slightly modified time discretization, a convergent scheme can be derived for a range of coarsening factors, both for two-grid V-cycles and multigrid FMG-cycles.

Intuitively, the modification in Scheme II changes how the discretized version of the stress tensor in Equation 6 is evaluated (compare Equation 12 and Equation 15), such that it is evaluated at given discrete time points across all levels in the time grid hierarchy in a consistent manner, see Section 2.5. Here, we note that both time discretizations are consistent and converge to the same

DOI: 10.1002/nla

numerical solution for decreasing time step sizes. By rewriting the time-discrete Equation 14,

$$\boldsymbol{u}_i = \boldsymbol{u}_{i-1} + \delta_{N_t}\boldsymbol{v}_i \tag{39}$$

$$= \boldsymbol{u}_{i-1} + \frac{\delta_{N_t}}{2}(\boldsymbol{v}_i + \boldsymbol{v}_{i-1}) + \frac{\delta_{N_t}^2}{2}(\frac{\boldsymbol{v}_i - \boldsymbol{v}_{i-1}}{\delta_{N_t}}), \tag{40}$$

one may argue that, for δ_{N_t}, Equation 40 becomes increasingly similar to the time-discrete Equation 11. Further, rewriting Equation 11,

$$\boldsymbol{u}_i = \boldsymbol{u}_{i-1} + \frac{\delta_{N_t}}{2}(\boldsymbol{v}_i + \boldsymbol{v}_{i-1}) = \boldsymbol{u}_{i-1} + \delta_{N_t}\boldsymbol{v}_{i-1} - \frac{\delta_{N_t}}{2}\boldsymbol{v}_{i-1} + \frac{\delta_{N_t}}{2}\boldsymbol{v}_i \tag{41}$$

$$= \boldsymbol{u}_{i-1} + \delta_{N_t}\boldsymbol{v}_{i-1} + \frac{\delta_{N_t}^2}{2}(\frac{\boldsymbol{v}_i - \boldsymbol{v}_{i-1}}{\delta_{N_t}}), \tag{42}$$

and assuming that the last term in Equation 42 can be neglected for small δ_{N_t}, one would effectively approximately discretize Equation 5 by an explicit step. This might be a hint for explaining the observed instability of MGRIT with Scheme I. On the other hand, the two-grid theory presented in [12], when used as an *a priori* tool for estimating convergence of MGRIT with Scheme I, provides a more mathematical and clear pathway for relating observed divergence in numerical experiments with the employed time-discretization.

In the case of using MGRIT with Scheme II, convergence was predicted for two-grid algorithms with F- and FCF-relaxation and a range of coarsening factors. The predictions matched quite closely with observations in the numerical experiments, supporting the use of the analysis presented in [12] as a powerful tool to design convergent algorithms *a priori*.

Two-grid results also motivated the design of a hybrid algorithm with the use of Scheme I on the fine grid and Scheme II on the coarse grid. Although, we were not able to derive a convergent hybrid scheme in this work, we showed that the use of Scheme II on the coarse grid removes the instability observed for MGRIT with Scheme I on both time grid levels. Further, Figure 8 illustrates that the predicted worst-case convergence factors are close to one, whereas for MGRIT with Scheme I on all levels yielded immediate and strong divergence with estimates $c_f^F, c_f^{FCF} >> 1$. Thus, hybrid schemes might be a promising approach to combine the advantages of different time discretization schemes.

Although the analysis presented in [12] is restricted to the two-grid case, the multilevel algorithm using MGRIT with Scheme II (presented in Section 3.4.1) was a direct result from the design of the convergent two-grid algorithm using MGRIT with Scheme II. Here, the use of FMG-cycles was beneficial to accelerate convergence in the true multilevel case, enabling larger combined coarsening factors $c_0^{m_l-1}$ compared to the two-grid case.

Further, preliminary timing results for the time-parallel algorithm were presented in Section 3.4.2. While only a marginal speedup could be achieved in the presented case for MGRIT with Scheme II and a five-grid hierarchy, results are promising. That is, reported wall clock times are almost entirely dominated by computation and communication. However, it is important to note that data export, for example, is a completely serial process for sequential time-stepping while it is parallel for the MGRIT algorithm. Thus, for practical applications that include data I/O tasks, larger speedups can be expected. Furthermore, the reported timing results were also restricted to a relatively small combined coarsening factor $c_0^4 = 16$. Thus, an MGRIT algorithm with improved convergence rates and larger coarsening factors (*i.e.*, smaller coarse-grid size) will exhibit much better speedup over sequential time-stepping.

5. CONCLUSION

In this work, we provided an analysis of the convergence of the multigrid reduction in time algorithm for the dynamic linear elasticity equations. It was shown that the two-grid convergence theory presented in [12] provides a mathematical explanation for the instability observed in previous

DOI: 10.1002/nla

works [9, 10]. Using the two-grid convergence theory as a design tool to estimate convergence of a modified backward Euler scheme (MGRIT with Scheme II) *a priori*, we were able to obtain a convergent parallel-in-time algorithm for a range of coarsening factors, advancing the application of parallel-in-time methods for second-order hyperbolic equations. In this study, the predicted theoretical convergence bounds were in excellent agreement with the worst-case convergence factors in numerical experiments. We then extended the algorithm to the multilevel case, noting that FMG-cycles and slow temporal coarsening can improve convergence compared to the two-grid algorithm.

In future work, we will investigate other cycling strategies (*e.g.* switching from FMG- to V-cycles after a number of initial iterates) to improve and accelerate convergence. Moreover, spatial coarsening will be investigated. Further, we aim to generalize the framework to the efficiently solve nonlinear hyperelastic stress-strain relationships with MGRIT to tackle fluid-structure interaction applications, particularly in the field of biomedical engineering.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement No. 306757 (LEAD).

REFERENCES

1. McCormick M, Nordsletten DA, Kay D, Smith NP. Simulating left ventricular fluid–solid mechanics through the cardiac cycle under LVAD support. J Comput Phys 2013; 244:80–96.
2. Hessenthaler A, Röhrle O, Nordsletten D. Validation of a non-conforming monolithic fluid-structure interaction method using phase-contrast MRI. Int J Numer Meth Bio 2016.
3. Bai Y, Sun D, Lin J, Kennedy D, Williams F. Numerical aerodynamic simulations of a NACA airfoil using CFD with block-iterative coupling and turbulence modelling. Int J Comput Fluid D 2012; 26(2):119–132.
4. Nievergelt J. Parallel methods for integrating ordinary differential equations. Commun ACM 1964/ 7(12):731–733.
5. Gander M. 50 years of time parallel time integration (In: Multiple Shooting and Time Domain Decomposition Methods). Springer 2015.
6. Garrido I, Espedal MS, Fladmark GE. A convergent algorithm for time parallelization applied to reservoir simulation (In: Domain Decomposition Methods in Science and Engineering). Springer 2005.
7. Samaddar D, Casper TA, Kim SH, Berry LA, Elwasif WR, Batchelor DB, Houlberg WA. Time parallelization of advanced operation scenario simulations of ITER plasma. J Phys Conf Ser 2013; 410(1):012032.
8. Schreiber M, Peixoto PS, Haut T, Wingate B. Beyond spatial scalability limitations with a massively parallel method for linear oscillatory problems. Int J High Perform C 2017.
9. Farhat C, Chandesris M. Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications. Int J Numer Meth Eng 2003; 58(9):1397–1434.
10. Farhat C, Cortial CJ, Dastillung C, Bavestrello H. Time-parallel implicit integrators for the near-real-time prediction of linear structural dynamic responses. Int J Numer Meth Eng 2006; 67(5):697–724.
11. Cortial J, Farhat C. A time-parallel implicit method for accelerating the solution of non-linear structural dynamics problems. Int J Numer Meth Eng 2009; 77(4):451–470.
12. Dobrev VA, Kolev TZ, Petersson NA, Schroder JB. Two-level convergence theory for multigrid reduction in time (MGRIT). SIAM J Sci Comput 2016 (submitted).
13. Gander MJ, Vandewalle S. Analysis of the parareal time-parallel time-integration method. SIAM J Sci Comput 2007; 29(2):556–578.
14. Falgout RD, Manteuffel TA, O'Neill B, Schroder JB. Multigrid Reduction in Time for Nonlinear Parabolic Problems: A Case Study. SIAM J Sci Comput 2016 (to appear).
15. Falgout RD, Friedhoff S, Kolev TZ, MacLachlan SP, Schroder JB. Parallel time integration with multigrid. SIAM J Sci Comput 2014; 36(6):C635–C661.
16. Trottenberg U, Oosterlee C, Schüller A. Multigrid. Academic Press, London, UK, 2001.
17. Lee J, Cookson A, Roy I, Kerfoot E, Asner L, Vigueras G, Sochi T, Deparis S, Michler C, Smith NP, Nordsletten DA. Multiphysics Computational Modeling in CHeart. SIAM J Sci Comput 2016; 38(3):C150–C178.
18. XBraid: Parallel multigrid in time. http://llnl.gov/casc/xbraid.

DOI: 10.1002/nla