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

A. Hessenthaler1*, D. Nordsletten2, O. Röhrle1, J. B. Schroder2, R. D. Falgout3

1Institute of Applied Mechanics (CE), University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany
2Division of Imaging Sciences and Biomedical Engineering, King’s College London, 4th Floor, Lambeth Wing St. Thomas Hospital London, SE1 7EH, UK
3Center 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 ParaFoil, 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.

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
HESSENTHALER ET AL.

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 \( \Gamma^D \) in \( d \) spatial dimensions. Then, \( X \in \Omega(0) \) and \( x \in \Omega \) denote the reference and current position of a material point and \( \nabla_X \) and \( \nabla_x \) denote the respective Lagrangian and Eulerian gradient operators. Further, we define the deformation gradient \( F = \nabla_X x = \nabla_X u + I \) where \( u = u(x, t) = x - X \) is the displacement of
a material point with respect to its position in the reference configuration. The partial derivative operator is denoted as $\partial_t$ 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,

$$\rho \partial_t u - \nabla_x \cdot \sigma = 0$$
$$\nabla_x \cdot \partial_t u = 0$$
$$u(\cdot, t) = 0$$
$$u(\cdot, 0) = 0, \quad v(\cdot, 0) = \hat{v}_0$$

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

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

$$\partial_t u = v$$
$$\rho \partial_t v = \nabla_x \cdot \sigma$$
$$\nabla_x \cdot v = 0$$
$$u(\cdot, t) = 0$$
$$u(\cdot, 0) = 0, \quad v(\cdot, 0) = \hat{v}_0$$

with velocity $v$.

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 $v$ 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

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

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 $[v_i, p_i, u_i]^T = [v(\cdot, t_i), p(\cdot, t_i), u(\cdot, t_i)]^T$. 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 $\partial_t$ in Equation 5 using the midpoint rule,

$$u_i = u_{i-1} + \delta_{N_t} \frac{v_i + v_{i-1}}{2} \quad \text{in } \Omega_0.$$  

(11)

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

$$\rho v_i - \frac{\mu \delta_{N_t}}{2} \nabla X v_i + \delta_{N_t} \nabla X p_i = \rho v_{i-1} + \frac{\mu \delta_{N_t}}{2} \nabla X v_{i-1} + \mu \delta_{N_t} \nabla X u_{i-1} \quad \text{in } \Omega_0,$$

(12)

$$\nabla X \cdot v_i = -\nabla X \cdot v_{i-1} \quad \text{in } \Omega_0,$$

(13)

and compute the displacement $u^i$ according to Equation 11 after each solve.
2.3.2. Scheme II

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

\[ u_i = u_{i-1} + \delta N_i v_i \quad \text{in } \Omega_0. \]  

(14)

Thus, in Scheme II we search the time-discrete \((v_i, p_i)\) for all \(i = 1, \ldots, N_t\), such that,

\[
\begin{align*}
\rho v_i - \mu \delta^2 N_i \nabla_X v_i + \delta N_i \nabla_X p_i &= \rho v_{i-1} + \mu \delta^2 N_i \nabla_X u_{i-1} \\
\nabla_X \cdot v_i &= 0
\end{align*}
\]

in \(\Omega_0\),

(15)

(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 \(\Omega_t\) was discretized using quadrilateral elements, \(\Omega^h_t\). Finite element discretizations were constructed using inf-sup stable \(Q^2 - Q^1\) Taylor-Hood elements for velocity and pressure, \(v^h\) and \(p^h\), and \(Q^2\) elements for displacement, \(u^h\). 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 \(i = 1, \ldots, N_t\) the space-time discrete solution vector \([v_i, p_i]^T\), such that,

\[
\begin{align*}
(\rho M - \mu \delta^2 N_i K)v_i + \delta N_i B^T p_i &= (\rho M + \mu \delta^2 N_i K)v_{i-1} + \mu \delta N_i K u_{i-1} \\
B v_i &= -B v_{i-1}
\end{align*}
\]

in \(\Omega_0\),

(17)

(18)

where \(\rho M\) is the mass matrix and \(K\) and \(B\) refer to the discretized weak form Laplacian and divergence operators \(\nabla^2_X()\) and \(\nabla_X \cdot ()\). Note, after solving for a given \([v_i, p_i]^T\), we can update the displacement \(u_i\) 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,

\[
\begin{bmatrix}
\rho M - \frac{\mu \delta^2 N_i}{2} K & \delta N_i B^T \\
B & 0 & 0 & p_i \\
-\frac{\delta N_i}{2} I & 0 & 0 & I
\end{bmatrix}
\begin{bmatrix}
v_i \\
p_i \\
u_i
\end{bmatrix}
= \begin{bmatrix}
\rho M + \frac{\mu \delta^2 N_i}{2} K & 0 & \mu \delta N_i K \\
-B & 0 & 0 \\
\frac{1}{2} \delta N_i I & 0 & I
\end{bmatrix}
\begin{bmatrix}
v_{i-1} \\
p_{i-1} \\
u_{i-1}
\end{bmatrix}.
\]

(19)

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

\[
D^I := \begin{bmatrix}
\rho M - \frac{\mu \delta^2 N_i}{2} K & \delta N_i B^T \\
B & 0 & 0 \\
-\frac{\delta N_i}{2} I & 0 & I
\end{bmatrix}, \quad C^I := \begin{bmatrix}
\rho M + \frac{\mu \delta^2 N_i}{2} K & 0 & \mu \delta N_i K \\
-B & 0 & 0 \\
\frac{1}{2} \delta N_i I & 0 & I
\end{bmatrix}.
\]

(20)

then \(\Phi^I := [D^I]^{-1} C^I\) and \(s^I_0 := [\hat{v}_0, 0, 0]^T\), the state variables \(s^I_i := [v_i, p_i, u_i]^T\) can be computed by the following equation,

\[
s^I_i = \Phi^I s^I_{i-1}
\]

for \(i = 1, \ldots, N_t\).

(21)

Note, that in the linear elastic incompressible case with constant spatial resolution, the operator \(\Phi^I\) is only dependent on time step size \(\delta N_i\).
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 \([v_i, p_i]^T\), such that,

\[
(\rho M - \mu \delta^2_{N_t} K)v_i + \delta_{N_t} B^T p_i = \rho M v_{i-1} + \mu \delta_{N_t} K u_{i-1} \quad \text{on } \Omega_0,
\]

\[
B v_i = 0 \quad \text{on } \Omega_0,
\]

and update the displacement \( u_i \) from Equation 14 after each solve. In matrix notation, we can write,

\[
\begin{bmatrix}
\rho M - \mu \delta^2_{N_t} K & \delta_{N_t} B^T & 0 \\
B & 0 & 0 \\
-\delta_{N_t} I & 0 & I
\end{bmatrix}
\begin{bmatrix}
v_i \\
p_i \\
u_{i-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\rho M & 0 & \mu \delta_{N_t} K \\
0 & 0 & 0 \\
0 & I & I
\end{bmatrix}
\begin{bmatrix}
v_{i-1} \\
p_{i-1} \\
u_{i-1}
\end{bmatrix}.
\]

With

\[
D_{II}^{II} := \begin{bmatrix}
\rho M - \mu \delta^2_{N_t} K & \delta_{N_t} B^T & 0 \\
B & 0 & 0 \\
-\delta_{N_t} I & 0 & I
\end{bmatrix}, \quad \text{and } C_{II}^{II} := \begin{bmatrix}
\rho M & 0 & \mu \delta_{N_t} K \\
0 & 0 & 0 \\
0 & I & I
\end{bmatrix},
\]

equation 24 can be written as,

\[
s_{II}^{II} = \Phi^I_{II}s_{II-1}^{II}
\]

for \( i = 1, \ldots, N_t \), (26)

where \( \Phi^I_{II} := [D_{II}^{II}]^{-1}C_{II}^{II} \) and \( s_{II}^{II} := [v_i, p_i, u_i]^T \) with \( s_{II}^{0} = \hat{s}_0 := [\hat{v}_0, 0, 0]^T \). Again, the operator \( \Phi^I_{II} \) only depends on time step size \( \delta_{N_t} \).

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,

\[
A^S s^S = \begin{bmatrix}
I \\
-\Phi^S \\
-\Phi^S \\
\vdots \\
-\Phi^S \\
I
\end{bmatrix}
\begin{bmatrix}
s^S_0 \\
s^S_1 \\
s^S_2 \\
\vdots \\
0
\end{bmatrix} = \hat{s},
\]

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 \( \delta^m_{N_l} = c_{m-1}^m \cdot \delta^m_{N_l-1} \) (for \( m = 1, \ldots, m_l - 1 \)), with fine grid time step size \( \delta^m_{N_l} \) and the coarsest grid’s time step size \( \delta^m_{N_{l-1}} \). 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].
Figure 1. Example of an \( m \)-level multigrid hierarchy with \( m = 2 \) time grid levels, respective time points \( t_i \), time step sizes \( \delta_{N^m} \) 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 \( s^I \) obtained from sequential time stepping is denoted as \( s^I_s \), whereas the solution vector obtained with MGRIT is denoted as \( s^I_p \). Similarly, we identify the numerical solution obtained with Scheme II by using the notation \( s^II_s \) and \( s^II_p \).

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 \( \Phi_s^{\delta_m} \) on the fine grid and its coarse grid versions \( \Phi_{\delta_m}^{S_m} \) on grid level \( m \) or similarly, \( \Phi_{\delta_m}^{II} \) on the fine grid and its coarse grid versions \( \Phi_{\delta_m}^{II_m} \) on grid level \( m \). The subscript \( \delta_m \) denotes \( \delta_{N^m} \). In Section 3.3, we will also motivate the design of a hybrid algorithm that employs \( \Phi_{\delta_m}^{S_m} \) on the fine grid and \( \Phi_{\delta_m}^{II_m} \) 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 \( A_{\Delta}^{S_m} \) is equivalent to that defined in (27) except it uses the coarse-grid time-stepper \( \Phi_{\delta_m}^{S_m} \) 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].
The standard Euclidean norm of the MGRIT residual in numerical experiments is computed using

\[ ||r_\Delta||^2 = (w_0 ||r_\Delta^m||_2^2 + w_1 ||r_\Delta^p||_2^2 + w_2 ||r_\Delta^n||_2^2)^{1/2} \]

or

\[ ||r_\Delta||^2 = (w_0 (r_\Delta^m)^T r_\Delta^m + w_1 (r_\Delta^p)^T r_\Delta^p + w_2 (r_\Delta^n)^T r_\Delta^n)^{1/2} \]
where $r^n_\Delta$, $r^n_\$ and $r^n_m$ 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

$$\|r_\Delta\|_{\mathcal{M}} = \|r^n_\Delta\|_{\mathcal{M}^n} + \|r^n_\$\|_{\mathcal{M}^n} + \|r^n_m\|_{\mathcal{M}^m},$$

where $\mathcal{M}$ is a block $N^1 \times N^1$ diagonal matrix with each block corresponding to a spatial mass matrix for variable $x$ (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 $w_2 = 1/c^1_0$. This yields

$$(r^n_\Delta)^T \mathcal{M}^n r^n_\Delta \approx (r^n_\Delta)^T I r^n_\Delta,$$

$$(r^n_m)^T \mathcal{M}^m r^n_m \approx (r^n_m)^T \left(1/c^1_0 I\right) r^n_m,$$

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

$$\|r_\Delta\|_{\mathcal{M}} \approx \left((r^n_\Delta)^T r^n_\$ + \frac{1}{c^1_0} (r^n_m)^T r^n_m \right)^{1/2}.$$  

(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 $\Phi_\Delta^m$ 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 $16 \times 2$ quadrilateral elements and initial conditions are given as $\Phi_0 = [-x^2/640, x^2, (8-x)/640]^T$, 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 $\delta_{N^0} \in \{1, 0.1, 0.01, 0.001\}$ and coarsening factors $c^1_0 \in \{2, 4, 8, 16, 32\}$ and $c^m_{m-1} = 2$ for $2 \leq m \leq m_i$. The stopping criterion on the residual norm is selected as $\|r_\Delta\|_{\mathcal{M}} \leq 10^{-8}/\sqrt{\delta_{N^0}}$ (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.
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 \( \delta_{N_0} \to 0 \). 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 \( \delta_{N_0} = 1 \) irrespectively of the chosen scheme (Scheme I and Scheme II). The effect of numerical damping becomes smaller for smaller \( \delta_{N_0} \), where we note that both schemes reproduce the amplitudes of the beam oscillation with comparable quality for \( \delta_{N_0} = 0.001 \), 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, \( \Phi_{\delta_{N_0}^I} \) and \( \Phi_{\delta_{N_1}^I} \)) 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 \( \delta_{N_0} = 1 \) 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 \( \Phi_{\delta_{N_0}^I} \) and \( \Phi_{\delta_{N_1}^I} \), 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, \( i.e., c_f^{\Phi}, c_f^{\Phi_{FCF}} >> 1 \), confirming experimental observations.
Figure 3. Displacement $u$ 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 $\delta N^0_t \in \{1, 0.1, 0.01, 0.001\}$. Note that numerical damping is reduced with $\delta N^0_t \to 0$.

Figure 4. Mean absolute error of the displacement of the beam’s tip for $t \in [0, 1024]$. Here, Scheme I with $\delta N^0_t = 0.001$ was used as reference.

3.3.2. MGRIT with Scheme II

Using Scheme II as fine- and coarse-grid integrators (that is, $\Phi_{II,0}^{c}$ and $\Phi_{II,1}^{c}$) in a two-level algorithm, we observe good convergence for a range of different coarsening factors $c^f_1$ and for all considered time step sizes, see Figure 6.

We note that for $\delta N^0_t = 1$, both experimental and predicted convergence factors are in excellent agreement and that the predicted values of $c^f_1$ and $c^{ FCC}_f$ are a sharp upper bound despite the approximation of the computed residual norm, see Section 2.5.4. Here, the convergence factors...
Figure 5. Current approximation of the tip displacement for two-grid MGRIT using Scheme I with coarsening factor $c_1^0 = 2$, fine grid time step size $\delta_{N_0} = 1$, $t \in [0, 64]$ 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 $c_1^0$ before decreasing due to the small coarse-grid size $N_t^{m_1-1}$.

On the other hand, for $\delta_{N_0} \in \{0.1, 0.01, 0.001\}$ we observe an increase in predicted and observed convergence factors with increasing coarsening factor $c_1^0$. Again, predicted and observed convergence factors are in excellent agreement for almost all considered cases. Only for $\delta_{N_0} = 0.01$ the maximum observed convergence factor is smaller for the F-relaxation case and approximately 1% larger for (i) F-relaxation and $c_1^0 = 32$ and (ii) FCF-relaxation and $c_1^0 \in \{16, 32\}$. 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 $\delta_{N_0} = 0.001$, 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 $t \in [0, 64]$, $\delta_{N_0} = 0.1$ and $c_1^0 = 2$. Here, no artificial amplification of the amplitude of oscillation is observed as is the case for MGRIT with Scheme I, see Figure 5.
Figure 6. Predicted and measured convergence factor for a two-grid algorithm with various fine grid time step sizes $\delta N_t \in \{1.0, 0.1, 0.01, 0.001\}$ 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 $\delta N_t = 0.1$, $t \in [0, 64]$ 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.
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 $\Phi_{I}^{\delta}$ and $\Phi_{II}^{\delta}$ 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 $c_{m_{l} - 1}^{m_{l} - 1} \in \{4, 8, 16, 32, 64\}$, where we only vary $c_{1}^{0}$ but select $c_{m_{l} - 1}^{m_{l} - 1} = 2$ 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)
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 \( c_{m_l - 1} = \prod_{m_{m_l - 2}}^{m_{m_l}} c_{m} + 1 \). 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 \( \delta_{N_0} N_t \) and \( c_0^{m_l - 1} \) are kept fixed, the mean convergence factor improves with increasing \( m_l \). For example, for \( \delta_{N_0} = 0.1 \) 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 \( \delta_{N_0} = 0.001 \). The standard Euclidean norm is used to measure solution progress. To investigate the dependency of the wall
CONVERGENCE OF THE MGRIT ALGORITHM FOR LINEAR ELASTICITY

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

\[\begin{align*}
u_i &= u_i - 1 + \delta_{N_i} v_i \\
&= u_{i-1} + \frac{\delta_{N_i}}{2} (v_i + v_{i-1}) + \frac{\delta_{N_i}^2}{2} \frac{v_i - v_{i-1}}{\delta_{N_i}},
\end{align*}\]

one may argue that, for \(\delta_{N_i}\), Equation 40 becomes increasingly similar to the time-discrete Equation 11. Further, rewriting Equation 11,

\[\begin{align*}
u_i &= u_i - 1 + \frac{\delta_{N_i}}{2} (v_i + v_{i-1}) = u_{i-1} + \delta_{N_i} v_{i-1} - \frac{\delta_{N_i}}{2} v_{i-1} + \frac{\delta_{N_i}}{2} v_i \\
&= u_{i-1} + \delta_{N_i} v_{i-1} + \frac{\delta_{N_i}^2}{2} \frac{v_i - v_{i-1}}{\delta_{N_i}},
\end{align*}\]

and assuming that the last term in Equation 42 can be neglected for small \(\delta_{N_i}\), 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 \textit{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 \textit{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 \(c_f^F, c_f^{FCF} \gg 1\). 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 \(c_{0i}^{-1}\) 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 \(c_0^F = 16\). Thus, an MGRIT algorithm with improved convergence rates and larger coarsening factors (\textit{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
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) \textit{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