MULTILEVEL CONVERGENCE ANALYSIS OF
MULTIGRID-REDUCTION-IN-TIME

ANDREAS HESSENTHALER†, BEN S. SOUTHWORTH‡, DAVID NORDSLETTERN§,
OLIVER RÖHRLE†, ROBERT D. FALGOUT¶, AND JACOB B. SCHRODER∥

Abstract. This paper presents a multilevel convergence framework for multigrid-reduction-in-time (MGRIT) as a generalization of previous two-grid estimates. The framework provides a priori upper bounds on the convergence of MGRIT V- and F-cycles with different relaxation schemes by deriving the respective residual and error propagation operators. The residual and error operators are functions of the time stepping operator, analyzed directly and bounded in norm, both numerically and analytically. We present various upper bounds of different computational cost and varying sharpness. These upper bounds are complemented by proposing analytic formulae for the approximate convergence factor of V-cycle algorithms that take the number of fine grid time points, the temporal coarsening factors, and the eigenvalues of the time stepping operator as parameters.

The paper concludes with supporting numerical investigations of parabolic and hyperbolic model problems. We assess the sharpness of the bounds and the quality of the approximate convergence factors. Observations from these numerical investigations demonstrate the value of the proposed multilevel convergence framework for the design of a convergent MGRIT algorithm. We further highlight that observations in the literature are captured by the theory, including that two-level Parareal and multilevel MGRIT with F-relaxation do not yield scalable algorithms, that iteration counts increase with increasing numbers of levels, and the benefit of a stronger relaxation scheme. The theory also proves in a two-level setting, and confirms for multilevel, that L-stable Runge-Kutta schemes are more amendable to parallel-in-time integration with MGRIT than A-stable Runge-Kutta schemes.

Key words. multilevel convergence theory, multigrid-reduction-in-time (MGRIT), parallel-in-time, multigrid

1. Introduction. Modern computer architectures enable massively parallel computations for systems under numerical investigation. While clock rates of recent high-performance computing resources have largely become stagnant, increased concurrency continues to reduce the time-to-solution, allowing for increased complexity of the computational model and accuracy of the computed quantities.

Spatial domain decomposition (DD) methods are a wide-spread class of parallelization techniques to exploit parallelism in numerical simulations. Many DD methods are straightforward to implement and scalable in parallel up to the point that communication tasks become dominant over computation tasks. Thus, spatial parallelism may saturate without exploiting the full potential of the available hardware.

Parallel-in-time methods [33, 16] increase the amount of parallelism that can be exploited by introducing parallelism in the temporal domain. Many such methods exist, including waveform relaxation [30, 45, 26, 44], space-time multigrid [25], parallel implicit time-integrator [11, 12, 4], revisionsit integral deferred correction [3], spectral deferred correction [7, 43, 34], Parareal [29] and multigrid-reduction-in-time [13, 8].
These methods were developed for various application areas with varying degree of intrusiveness, ease of implementation, level of parallelism, and potential for speedup. For an extensive review see [16].

Multigrid-reduction-in-time (MGRIT) is a recently developed, non-intrusive method that is a true multilevel algorithm. MGRIT is an $O(N)$ iterative solver that introduces parallelism in the temporal domain by employing a parallel, iterative coarse-grid correction scheme based on multigrid reduction. It has been explored for various application areas, such as, the numerical solution of parabolic and hyperbolic partial differential equations (PDEs) [13, 10, 27, 23], investigations of power systems [28, 39], solving adjoint problems [18] and optimization problems [19], and neural network training [37].

Two-level convergence theory for MGRIT was developed in [5], based on the assumptions of time integration on a uniform time grid and simultaneous diagonalization of time-stepping operators (see Section 2.2), and the theoretical bounds were shown to be quite accurate when compared with observed convergence. Southworth [42] generalized this framework, deriving necessary and sufficient conditions and tight two-level convergence bounds for general two-level MGRIT for linear PDEs on a uniform time grid. Some extensions to the case of non-uniform time grids are also provided in [42, 47]. However, no work has been done on convergence theory for the general multilevel case. In practice, the selection of an appropriate cycling strategy and relaxation scheme is crucial for scalable multilevel performance and, ultimately, for achieving parallel speedup. A strong framework for multilevel convergence of MGRIT can help guide these decisions in a rigorous, a priori manner.

In this work, we generalize previous two-level convergence theory of MGRIT to the multilevel setting and derive bounds on convergence. The upper bounds on residual and error propagation derived here are able to analyze multilevel performance of MGRIT a priori, both numerically and analytically, and with varying degree of sharpness and computational cost. Furthermore, we present analytic formulae for approximating the convergence factor of MGRIT algorithms that employ the most common cycling strategy and two different relaxation schemes. We then demonstrate the sharpness of the derived theoretical bounds for parabolic (including new analysis for the anisotropic diffusion equation) and hyperbolic model PDEs in Section 5 and highlight the benefits of the theory presented in this work.

This paper lays the groundwork for future in-depth examination and understanding of MGRIT and will help guide future development and improvement of MGRIT to explore ideas such as coarsening in integration order as opposed to step size ($p$-MGRIT), weighted relaxation schemes, etc. Furthermore, we provide a parallel C++ implementation of all derived bounds and approximate convergence factors\(^1\).

\section{Multigrid-reduction-in-time (MGRIT).}

For linear problems, sequential time stepping can be written as,

\begin{equation}
  u_n = \Phi_n u_{n-1} + g_n, \quad \text{for } n = 1, \ldots, N_t - 1,
\end{equation}

with state vector $u_n \in \mathbb{R}^{N_x}$ at time $t_n \in (0, T]$, initial condition $u_0$ at time $t_0 = 0$, and forcing function $g_n$. Here, $N_x$ refers to the number of degrees of freedom at one point in time, and $N_t$ refers to the number of time points\(^2\). For the theoretical analysis, we consider equidistant time points $\delta t_n = t_n - t_{n-1} = \delta t$ and a time-independent

\(^1\)Implementation publicly available at: github.com/XBraid/XBraid-convergence-est

\(^2\)Note that in contrast to [5], we include the initial time point.
one-step\textsuperscript{3} integrator $\Phi_n = \Phi$.

In matrix form, (2.1) can be written as,

$$
(2.2) \quad Au = \begin{bmatrix}
I & -\Phi & I & \cdots \\
-\Phi & I & -\Phi & I \\
& \ddots & \ddots & \ddots \\
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
\vdots \\
\end{bmatrix} = f,
$$

where sequential time-stepping is identified as a block-forward solve of (2.2).

Multigrid-reduction-in-time (MGRIT) [13, 8] solves (2.2) iteratively and, like sequential time-stepping, is an $O(N_t)$ method. MGRIT introduces a hierarchy of $n_\ell$ time grids to achieve parallelism in the temporal domain by employing a coarse-grid correction scheme based on multigrid reduction. The fine grid (referred to as level $\ell = 0$) is composed of all time points $t_n$ ($n = 0, \ldots, N_t - 1$) and the coarser grids (referred to as levels $\ell = 1, \ldots, n_\ell - 1$) are derived from a uniform coarsening of the fine grid, see Figure 1. The temporal coarsening factors are denoted as $m_\ell \in \mathbb{N}$ (for $\ell = 0, \ldots, n_\ell - 2$), such that the number of points on each time grid level are given as,

$$
(2.3) \quad N_\ell = \frac{N_{\ell-1} - 1}{m_{\ell-1}} + 1, \quad \text{for } \ell = 1, \ldots, n_\ell - 1.
$$

On each time grid level $\ell$, the time points are partitioned into F-points (black) and C-points (red), whereas C-points compose the next coarser grid level $\ell + 1$.

Fig. 1: Two-grid hierarchy: time points $t_n$, fine-/coarse-grid step sizes $\delta_0$ and $\delta_1$, coarsening factor $m_0 = 4$.

\textbf{2.1. MGRIT Operators.} MGRIT approximates the exact coarse-grid time-stepping operator\textsuperscript{5} on level $\ell$ by introducing,

$$
(2.4) \quad \Phi_\ell \approx \Phi_{\ell-1}^{m_{\ell-1}}, \quad \text{for } \ell = 1, \ldots, n_\ell - 1,
$$

and we write,

$$
(2.5) \quad A_\ell = \begin{bmatrix}
I & -\Phi_\ell & I & \cdots \\
-\Phi_\ell & I & -\Phi_\ell & I \\
& \ddots & \ddots & \ddots \\
\end{bmatrix} \in \mathbb{R}^{N_\ell \times N_\ell \times N_\ell}, \quad \text{for } \ell = 1, \ldots, n_\ell - 1.
$$

\textsuperscript{3}Multistep time integration schemes can be addressed in a similar way; see [9].

\textsuperscript{4}Note that $m_\ell = 1$ for some or all $\ell$ is a valid choice, e.g., for a p-multigrid-like approach.

\textsuperscript{5}Time-stepping on the coarse-grid is referred to as exact, if it yields the same solution as sequential time-stepping on the fine-grid.
MGRIT constructs coarse-grids from a Schur complement decomposition of (2.5), relative to the F/C-splitting from Figure 1. This Schur complement can be defined with ideal restriction and interpolation operators \([5]\), \(R_\ell \in \mathbb{R}^{N_x N_{\ell+1} \times N_x N_\ell}\) (level \(\ell\) to \(\ell + 1\)) and \(P_\ell \in \mathbb{R}^{N_x N_\ell \times N_x N_{\ell+1}}\) (level \(\ell + 1\) to \(\ell\)),

\[
R_\ell = \begin{bmatrix}
I & \Phi_{\ell}^{m_\ell - 1} & \Phi_{\ell}^{m_\ell - 2} & \ldots & \Phi_{\ell} & I \\
& \ddots & & & & \Phi_{\ell}^{m_\ell - 1} & \Phi_{\ell}^{m_\ell - 2} & \ldots & \Phi_{\ell} & I \\
\end{bmatrix}, \tag{2.6}
\]

\[
P_\ell = \begin{bmatrix}
I & \Phi_{\ell}^{m_\ell - 1} \\
& \ddots & & & & \Phi_{\ell} & I \\
\end{bmatrix}, \tag{2.7}
\]

for \(\ell = 0, \ldots, n_\ell - 2\), and operator \(S_\ell \in \mathbb{R}^{N_x N_\ell \times (N_\ell - N_{\ell+1})}\),

\[
S_\ell = \begin{bmatrix}
0 & I & I & \cdots & I & 0 \\
I & 0 & I & \cdots & 0 & I \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & I & 0
\end{bmatrix}, \quad \text{for } \ell = 1, \ldots, n_\ell - 2.
\]

The number of block columns in \(S_\ell\) corresponds to the total number of F-points on level \(\ell\). If the operator \(S_\ell\) is applied to the right of \(A_\ell\), the result \(A_\ell S_\ell\) is composed of all block rows in \(A_\ell\), and all block columns in \(A_\ell\) that correspond to F-points on level \(\ell\) (zeroing out the respective C-point block columns). Thus, \(S_\ell^T A_\ell S_\ell\) is composed of all block rows and block columns in \(A_\ell\) that correspond to F-points.

We note, that \(R_\ell A_\ell S_\ell = 0\), so that we can write the coarse-grid operator as \(R_\ell A_\ell P_\ell = R_\ell I_\ell P_\ell\) \([5]\), where restriction by injection is given by the operator,

\[
R_\ell = \begin{bmatrix}
I & 0 & 0 & \cdots & I \\
& \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & I & 0 \\
\end{bmatrix}, \quad \text{for } \ell = 1, \ldots, n_\ell - 2.
\]

Here, \(R_\ell \in \mathbb{R}^{N_x N_{\ell+1} \times N_x N_\ell}\) has the same structure as \(R_\ell\), with all blocks \(\Phi_{\ell}^d\) (for \(d = 1, \ldots, m_\ell - 1\)) set to zero. Thus, \(R_\ell\) restricts the C-points from level \(\ell\) to level
\( \ell + 1 \), omitting the respective F-points. The number of block rows corresponds to the total number of C-points on level \( \ell \), i.e., \( N_{\ell+1} \). Note, that the inverse of \( A_\ell \) is given as,

\[
A_\ell^{-1} = \begin{bmatrix}
I & & & \\
\Phi_\ell & I & & \\
\Phi_\ell^2 & \Phi_\ell & I & \\
\vdots & \vdots & \ddots & \ddots \\
\Phi_\ell^{N_\ell-1} & \Phi_\ell^{N_\ell-2} & \cdots & I
\end{bmatrix}, \quad \text{for } \ell = 1, \ldots, n_\ell - 1.
\]

MGRIT uses relaxation steps on each time grid to reduce the error that cannot be adequately reduced on coarser grids. F-relaxation is block Jacobi applied to the F-point block rows of \( A_\ell \) and updates all F-points based on the previous C-point, which is defined by an application of the idempotent operator \( F_\ell = P_\ell R_{\ell t} = I - S_\ell (S_\ell^T A_\ell S_\ell)^{-1} S_\ell^T \). On the other hand, define \( T_\ell = R_{\ell t}^T \), C-relaxation updates a C-point based on the previous F-point (block Jacobi applied to the C-point block rows of \( A_\ell \)), which is an application of \( C_\ell = I - T_\ell (T_\ell^T A_\ell T_\ell)^{-1} T_\ell^T A_\ell \). Thus, FCF-relaxation (subsequent F-, C- and F-relaxation steps) is given as \( F_\ell C_\ell F_\ell = P_\ell (I - R_\ell A_\ell P_\ell) R_{\ell t} \).

2.2. Simultaneous diagonalization of \( \{\Phi_\ell\} \). Let \( \Phi_\ell \) denote the time-stepping operator on level \( \ell \). Primary results in this paper rest on the assumption that \( \Phi_\ell \) is diagonalizable with the same set of eigenvectors, for all levels \( \ell = 0, \ldots, n_\ell - 1 \). This is equivalent to saying that the set \( \{\Phi_\ell\} \) commutes, that is, \( \Phi_\ell \Phi_j = \Phi_j \Phi_\ell \) for all \( i, j \), and that \( \Phi_\ell \) is diagonalizable for all \( \ell \). The concept of simultaneous diagonalization (albeit, in the Fourier basis) was introduced in [14], and was modified and used as the basis for improved two-grid convergence bounds developed in [5].

An important question is then when such an assumption is valid. In the context of time integration, suppose we use a method of lines approach, where \( \mathcal{L} \) is a time-independent operator (such as a spatial discretization) that is propagated through time by operators \( \{\Phi_\ell\} \). In most cases, the assumption on diagonalizability holds if \( \mathcal{L} \) is diagonalizable. Indeed, for nearly all standard time-integration routines, including all single-step Runge-Kutta-type methods, \( \Phi_\ell \) is some rational function of \( \mathcal{L} \). Noting that the polynomial basis \( c_0 \mathcal{L}, I \pm c_1 \mathcal{L}, \text{ and } (I \pm c_2 \mathcal{L})^{-1} \) all commute, for arbitrary constants \( c_0, c_1, \) and \( c_2 \), if \( \mathcal{L} \) is diagonalizable, any rational combination of these polynomials is simultaneously diagonalizable.

Under the assumption of simultaneous diagonalizability, certain bounds can be derived on functions of \( \{\Phi_\ell\} \). Let \( \Phi_\ell = U \Lambda_\ell U^{-1} \), where \( (\Lambda_\ell)_{ii} = \lambda_{\ell, i} \) are diagonal matrices consisting of the eigenvalues of \( \Phi_\ell \) for \( i = 1, \ldots, N_\ell \), and columns of \( U \) are the corresponding eigenvectors. Let \( \tilde{U} \) denote a block-diagonal matrix with diagonal blocks given by \( U \), and suppose \( \mathcal{A} \) is some matrix operator, where each entry is a rational function of time-stepping operators in \( \{\Phi_\ell\} \) (such as, for example, the error and residual-propagation operators of MGRIT that are derived in Section 3). Then, as in [42],

\[
\|\mathcal{A}(\Phi_0, \ldots, \Phi_{n_\ell-1})\|_{(\tilde{U} \tilde{U}^*)^{-1}} = \sup_i \|\mathcal{A}(\lambda_{0, i}, \ldots, \lambda_{n_\ell-1, i})\|.
\]

Thus, the \((\tilde{U} \tilde{U}^*)^{-1}\)-norm of \( \mathcal{A}(\Phi_0, \ldots, \Phi_{n_\ell-1}) \) can be computed by maximizing the \( \ell^2 \)-norm of \( \mathcal{A} \) over eigenvalues of \( \{\Phi_\ell\} \). In the case that \( \{\Phi_\ell\} \) are normal matrices, \( U \) is unitary and the \((\tilde{U} \tilde{U}^*)^{-1}\)-norm reduces to the standard Euclidean 2-norm. More
generally, we have bounds on the $\ell^2$-norm of $\mathcal{A}(\Phi_0, \ldots, \Phi_{n-1})$,
$$\frac{1}{\pi \rho(U)} \left( \sup_i \|\mathcal{A}(\lambda_0, \ldots, \lambda_{n-1}, i)^k\| \right) \leq \|\mathcal{A}(\Phi_0, \ldots, \Phi_{n-1})^k\| \leq \kappa(U) \left( \sup_i \|\mathcal{A}(\lambda_0, \ldots, \lambda_{n-1}, i)^k\| \right),$$
where $\kappa(U)$ denotes the matrix condition number of $U$.

Here, we are interested in $\mathcal{A}$ corresponding to the error- and residual-propagation operators of MGRIT, denoted $\mathcal{E}$ and $\mathcal{R}$. Convergence of MGRIT requires
$$\|\mathcal{E}^k\|, \|\mathcal{R}^k\| \to 0,$$
with iteration $k$; to that end, bounding $\sup_i \|\mathcal{E}(\lambda_0, \ldots, \lambda_{n-1}, i)^k\| < 1$ for all $i$ provides necessary and sufficient conditions for $\|\mathcal{E}(\Phi_0, \ldots, \Phi_{n-1})^k\| \to 0$ with $k$ (eventually), and similarly for $\mathcal{R}(\Phi_0, \ldots, \Phi_{n-1})$.

### 2.2.1. General Runge-Kutta schemes
This section considers operators $\{\Phi_t\}$ in the case of general single-step, multi-stage schemes, most commonly presented as Runge-Kutta methods. In particular, a closed form for eigenvalues of $\Phi_t$ is derived as a function of eigenvalues of $\mathcal{L}$, for arbitrary Runge-Kutta methods.

Let $\mathfrak{A} = a_k (i, k = 1, \ldots, s), \hspace{0.5cm} \mathbf{b}^T = [b_1, \ldots, b_s], \hspace{0.5cm} \mathbf{c} = [c_1, \ldots, c_s]^T$ denote the Butcher Tableau,
$$\begin{array}{c|c}
\mathfrak{A} & \mathbf{b}^T \\
\hline
\mathbf{c} & \\
\end{array}$$
for an arbitrary $s$-stage Runge-Kutta scheme and let $\mathcal{L}$ denote the time-independent operator being propagated through time. Then, the linear system corresponding to time integration (2.2) can be represented in the following block form:

\[
\begin{bmatrix}
I_{n\times n} & 0 & 0 & \cdots & 0 \\
-I_{n\times n} & I_{n\times n} & 0 & \cdots & 0 \\
-I_{n\times n} & -a_1 \mathcal{L} & I_{n\times n} & \cdots & 0 \\
& \vdots & \ddots & \ddots & \vdots \\
& & & -a_s \mathcal{L} & 0 \\
\end{bmatrix}
\begin{bmatrix}
\kappa_0(n-1) \\
\kappa_1(n) \\
\vdots \\
\kappa_s(n) \\
\end{bmatrix} = \begin{bmatrix}
f_0 \\
f_1 \\
\vdots \\
\tilde{f}_{s+1} \\
\end{bmatrix}
\]

Here, $u_i$ is the solution at time $t_i$ and $f$ is some vector right-hand side that depends on the right-hand side of the problem and the vector $\mathbf{c}$ from the Butcher Tableau. The inner $s \times s$ block in Equation (2.8) corresponds to the inner stages of the Runge-Kutta method. Such a structure can also be condensed to the convenient form

\[
\begin{bmatrix}
\mathbf{I}_{n\times n} & -\mathcal{L} \mathfrak{A} \otimes \mathcal{L} \\
-\mathcal{L} \mathfrak{A} \otimes \mathcal{L} & \mathbf{I}_{n\times n} \\
-\mathcal{L} \mathfrak{A} \otimes \mathcal{L} & \mathbf{I}_{n\times n} \\
\end{bmatrix}
\begin{bmatrix}
\mathbf{u}_0 \\
\mathbf{u}_1 \\
\mathbf{u}_2 \\
\end{bmatrix} = \tilde{f}.
\]

This yields a lower triangular matrix with overlapping $3 \times 3$ blocks. Note that for explicit Runge-Kutta (ERK) schemes, $\mathfrak{A}$ is strictly lower triangular; for diagonally implicit Runge-Kutta (DIRK) methods, $\mathfrak{A}$ is lower triangular; and for singly diagonally

A similar modified norm also occurs in the case of integrating in time with a mass matrix [5].

\[6\]
implicit Runge-Kutta (SDIRK) methods, $\mathfrak{A}$ is lower triangular with unit diagonal. Let $L = U D U^{-1}$, where $D_{ii} = \xi_i$ is a diagonal matrix containing the eigenvalues of $L$. Simple elimination of blocks through forward substitution yields a linear system as in (2.2), where
\[
\Phi = I + \delta t b^T \otimes I \left( I - \delta \mathfrak{A} \otimes D \right)^{-1} \left( 1_n \otimes D \right)
= U \left( I + \delta t b^T \otimes I \left( I - \delta \mathfrak{A} \otimes D \right)^{-1} \left( 1_n \otimes D \right) \right) U^{-1}.
\]

Then, error and residual propagation of MGRIT can be bounded in norm based on the eigenvalues of $\Phi$, which are given by the eigenvalues of $I + \delta t b^T \otimes I \left( I - \delta \mathfrak{A} \otimes D \right)^{-1} \left( 1_n \otimes D \right)$. This matrix can be reordered by eigenvalue index to be block diagonal in the eigenvalues of $L$. If $\xi_i$ is the $i^{th}$ eigenvalue of $L$, the $i^{th}$ eigenvalue of $\Phi$ is then given by
\[
\lambda_i = 1 + \delta t \xi_i b^T (I - \delta t \xi_i \mathfrak{A})^{-1}.
\]

Note that Equation (2.9) is exactly equivalent to the stability function for a Runge-Kutta time-integration scheme [22, 2].

**Remark 2.1.** A similar result was presented in [5], demonstrating that the assumption on simultaneous diagonalization holds for any $\Phi$ that can each be written as a rational function of the form
\[
\Phi(L) = \left( I + \sum_{\nu} \beta_{\nu}(\delta t L)^{\nu} \right)^{-1} \left( I + \sum_{\nu} \alpha_{\nu}(\delta t L)^{\nu} \right).
\]
In the case of scalar ODEs, the relation between (2.9) and a rational function $R(z) = P(z)/Q(z)$ is well known [2]. For an operator on the space of diagonalizable matrices, this equivalence holds, but is more subtle. Here, we present material in a way that leads directly to the stability function (2.9), which is generally derived and available for existing Runge-Kutta schemes.

**2.3. Two-level results, and why multilevel is harder.** For multigrid-type algorithms, it is generally the case that two-level convergence rates provide a lower bound on attainable multilevel convergence rates; that is, a multilevel algorithm will typically observe worse convergence than its two-level counterpart. Indeed, more expensive multilevel cycling strategies such as W-cycles or F-cycles are used specifically to solve the coarse-grid operator more accurately, better approximating a two-grid method. The non-Galerkin coarse grid used in MGRIT makes this relationship more complicated, and it is not clear that two-grid convergence provides a lower bound on multilevel in all cases. However, in practice, it is consistently the case that two-level convergence is better than multilevel. To that end, a two-level method which converges every iteration is a heuristic necessary condition for multilevel convergence.

A key observation in convergence of two-level MGRIT is that error and residual propagation take a block outer product form [42]. For example, error and residual propagation for $k$ iterations of two-level MGRIT with $F$-relaxation are given by [42]
\[
E_k^F = P_{ideal} \left[ 0 \ (I - A_1^{-1}(R_0 A_0 P_0)^k) \right], \quad R_k^F = \left[ (I - (R_0 A_0 P_0) A_1^{-1})^k \right] R_{ideal}.
\]
Here, $R_{ideal}$ denotes the ideal restriction operator, corresponding to $R_\ell$ in (2.6) with columns reordered such that the last $N_1$ columns are the identity, and similarly $P_{ideal}$
corresponds to (2.7) with rows reordered such that the last \( N \) rows are the identity. There are two ways to interpret this result. In [5], it is recognized that C-point error and residual propagation is governed by \( I - (R_0 A_0 P_0) A_1^{-1} \) on all iterations. Because the final step in any MGRIT iteration is a direct solve over F-points (F-relaxation), if error at C-points is zero, then it will be at F-points as well. Thus, bounding \( \|I - (R_0 A_0 P_0) A_1^{-1}\| < 1 \) is a sufficient condition for convergence in any norm \( \| \cdot \| \).

In [42], it is recognized that the factors of \( P_{\text{ideal}} \) and \( R_{\text{ideal}} \) only affect F-point error and residual on the last and first iteration, respectively. Convergence on C- and F-points of all other iterations is governed by \( I - (R_0 A_0 P_0) A_1^{-1} \). This led to the development of necessary and sufficient conditions for two-level convergence as well as tight bounds in norm of \( I - (R_0 A_0 P_0) A_1^{-1} \) based on a “temporal approximation property” (TAP), where the TAP is defined as follows.

**Definition 2.2 (Temporal approximation property).** Let \( \Phi_0 \) denote a fine-grid time-stepping operator and \( \Phi_1 \) denote a coarse-grid time-stepping operator, for all time points, with coarsening factor \( m_0 \). Then, \( \Phi_0 \) satisfies an F-relaxation temporal approximation property (F-TAP), with respect to \( \Phi_1 \), with constant \( \varphi_F \), if, for all vectors \( \mathbf{v} \),

\[
\| (\Phi_1 - \Phi_0^{m_0}) \mathbf{v} \| \leq \varphi_F \left[ \min_{x \in [0,2\pi]} \| (I - e^{ix} \Phi_1) \mathbf{v} \| \right].
\]

Similarly, \( \Phi_0 \) satisfies an FCF-relaxation temporal approximation property (FCF-TAP), with respect to \( \Phi_1 \), with constant \( \varphi_{FCF} \), if, for all vectors \( \mathbf{v} \),

\[
\| (\Phi_1 - \Phi_0^{m_0}) \mathbf{v} \| \leq \varphi_{FCF} \left[ \min_{x \in [0,2\pi]} \| \Phi_0^{-m_0} (I - e^{ix} \Phi_1) \mathbf{v} \| \right].
\]

Note that for real-valued operators, \( \min_{x} \) is given at \( x = 0 \), in which case \( e^{ix} = 1 \).

The results of [42] effectively state that satisfying an F-TAP with constant \( \varphi_F < 1 \) is a necessary and sufficient condition for convergence of two-level MGRIT with F-relaxation in the \( \ell^2 \)- or \( A^* A \)-norm. Furthermore, \( \varphi_F \) gives tight bounds (to \( O(1/N_1) \)) on convergence for all but the last (first) iteration in the \( \ell^2 \)-norm (\( A^* A \)-norm) (or, bounds on C-point error and residual for all iterations). Similarly, satisfying an FCF-TAP with constant \( \varphi_{FCF} < 1 \) is a necessary and sufficient condition for convergence of two-level MGRIT with FCF-relaxation in the \( \ell^2 \)- or \( A^* A \)-norm, with tight bounds given by \( \varphi_{FCF} \). If \( \Phi_0 \) and \( \Phi_1 \) are simultaneously diagonalizable with eigenvectors \( U \), this result simplifies to saying that worst-case convergence of two-level MGRIT with F- and FCF-relaxation, respectively, in the \( (U U^*)^{-1} \)-norm, for all but the first iteration (or, for convergence of C-points on all iterations), is given by

\[
(2.10) \quad \sup_i \frac{|\lambda_{1,i} - \lambda_{0,i}^{m_0}|}{1 - |\lambda_{1,i}|} = \varphi_F < 1, \quad \sup_i \frac{|\lambda_{0,i}^{m_0}| |\lambda_{1,i} - \lambda_{0,i}^{m_0}|}{1 - |\lambda_{1,i}|} = \varphi_{FCF} < 1.
\]

If \( U \) is unitary, then \( (U U^*)^{-1} \)-norm is just the \( \ell^2 \)-norm. Note that these are exactly the bounds obtained in [5] for two-level MGRIT; that is, the upper bounds derived in [5] are tight.

Although results obtained in [42, 5] are tight for all but one iteration, or tight with respect to C-points on all iterations, in the multilevel setting we are interested

\footnote{Note that residual and error propagation are similar, where \( \| E \|_{A^* A} = \| R \|. \)
in convergence over all points for a single iteration. For example, consider a three-level MGRIT V-cycle, with levels 0, 1, and 2. On level 1, a single iteration of two-level MGRIT is applied as an approximate residual correction for level 0. Suppose conditions in [42, 5] are satisfied, ensuring a decrease in C-point error, but a possible increase in F-point error on level 1. If the total error on level 1 has increased, then a correction is interpolated to level 0 that is a worse approximation to the exact residual correction, compared with no correction (corresponding to the zero initial guess used for coarse-grid correction in multigrid). In general, if divergent behavior is observed for iterations in the middle of the hierarchy, it is likely the case that the whole multilevel scheme will diverge. Given that multilevel convergence is usually worse than two-level in practice, this motivates a stronger two-grid result that can ensure two-grid convergence for all points on all iterations.

The following theorem introduces stronger variations in the TAP that provide necessary conditions for a two-level method with F- and FCF-relaxation to converge every iteration on C-points and F-points. Corollary 2.4 strengthens this result in the case of simultaneous diagonalization of \( \Phi_0 \) and \( \Phi_1 \), deriving necessary and sufficient conditions for convergence, with tight bounds in norm.

**Theorem 2.3.** Let \( E_F \) and \( E_{FCF} \) denote error propagation of two-level MGRIT with F-relaxation and FCF-relaxation, respectively. Define

\[
W_F := \sqrt{\sum_{k=0}^{h-1} \Phi_0^k(\Phi_0^k)^*},
\]

\[
W_{FCF} := \sqrt{\sum_{k=0}^{h-1} \Phi_0^k(\Phi_0^k)^*},
\]

and \( \varphi_F^{A^*A} \) and \( \varphi_{FCF}^{A^*A} \) as the minimum constants such that, for all \( \mathbf{v} \),

\[
\|(\Phi_1 - \Phi_0^m_0)\mathbf{v}\| \leq \varphi_F^{A^*A} \left[ \min_{x \in [0, 2\pi]} \left\| W_F^{-1}(I - e^{ix}\Phi_1)\mathbf{v}\right\| + O(1/\sqrt{N_1}) \right],
\]

\[
\|(\Phi_1 - \Phi_0^m_0)\mathbf{v}\| \leq \varphi_{FCF}^{A^*A} \left[ \min_{x \in [0, 2\pi]} \left\| W_{FCF}^{-1}(I - e^{ix}\Phi_1)\mathbf{v}\right\| + O(1/\sqrt{N_1}) \right].
\]

Then,

\[
\|E_F\|_{A^*A} \geq \varphi_F^{A^*A}, \quad \|E_{FCF}\|_{A^*A} \geq \varphi_{FCF}^{A^*A}.
\]

In addition, assume that \( \Phi_0 \) and \( \Phi_1 \) commute, and define \( \varphi_F^2 \) and \( \varphi_{FCF}^2 \) as the minimum constants such that, for all \( \mathbf{v} \),

\[
\|W_F(\Phi_1 - \Phi_0^m_0)\mathbf{v}\| \leq \varphi_F^2 \left[ \min_{x \in [0, 2\pi]} \left\| (I - e^{ix}\Phi_1)\mathbf{v}\right\| + O(1/\sqrt{N_1}) \right],
\]

\[
\|W_F(\Phi_1 - \Phi_0^m_0)\mathbf{v}\| \leq \varphi_{FCF}^2 \left[ \min_{x \in [0, 2\pi]} \left\| \Phi_0^{-m_0}(I - e^{ix}\Phi_1)\mathbf{v}\right\| + O(1/\sqrt{N_1}) \right].
\]

Let \( E_F \) and \( E_{FCF} \) denote error propagation of two-level MGRIT with F-relaxation and FCF-relaxation, respectively. Then,

\[
\|E_F\| \geq \varphi_F^2, \quad \|E_{FCF}\| \geq \varphi_{FCF}^2.
\]

**Proof.** See Appendix A.

---

\(^8\)This assumption is not technically necessary, but simplifies the proof and is sufficient for this paper. For a more detailed discussion, see [42].
Corollary 2.4. Assume that $\Phi_0$ and $\Phi_1$ are simultaneously diagonalizable with eigenvectors $U$ and eigenvalues $\{\lambda_{0,i}\}_i$ and $\{\lambda_{1,i}\}_i$, respectively. Let $\tilde{U}$ denote a block-diagonal matrix with diagonal blocks given by $U$. Then,

$$
\|R_F\|_{(\tilde{U}\tilde{U}^*)^{-1}} = \|E_F\|_{(\tilde{U}\tilde{U}^*)^{-1}} = \sqrt{1 - \frac{|\lambda_{0,i}|^{2m}_0}{1 - |\lambda_{0,i}|^2} \left(1 - |\lambda_{1,i}|\right) + O(1/N_1)},
$$

(2.11)

$$
\|R_{FCF}\|_{(\tilde{U}\tilde{U}^*)^{-1}} = \|E_{FCF}\|_{(\tilde{U}\tilde{U}^*)^{-1}} = \sqrt{1 - \frac{|\lambda_{0,i}|^{2m}_0}{1 - |\lambda_{0,i}|^2} |\lambda_{1,i} - \lambda_{0,i}^{m}_0|} \left(1 - |\lambda_{1,i}|\right) + O(1/N_1).
$$

Proof. See Appendix A.

Note in Theorem 2.3 and Corollary 2.4, there is an additional scaling compared with results obtained on two-grid convergence for all but the first iteration [42], which makes the convergence factors larger. For example, in error propagation in Theorem 2.3, the left-hand side of the F-TAP and FCF-TAP has an additional factor of $W$ compared with results in [42]. Because $\|W_Fv\| > \|v\|$ for all $v$, worst-case convergence factors on the first iteration are necessarily larger than those obtained for other iterations (based on just the F-TAP and FCF-TAP). How much larger depends on the error mode and the coarsening factor, but this immediately gives one reason that multilevel convergence is more difficult than two-level. Figure 2 demonstrates the impact of this additional scaling by plotting two-level MGRIT convergence as a function of $\delta t$ times the spatial eigenvalues in the complex plane.

There are a few interesting points to note from Figure 2. First, when considering integration schemes to use with Parareal or MGRIT, it is clear that, for example, not all second-order schemes are equal. The second and third column in Figure 2 both correspond to second-order, implicit Runge-Kutta schemes, but the L-stable scheme (third column) offers a much larger region of convergence in the complex plane. This corresponds, in particular, to converging for large $\delta t$, which is particularly important in the multilevel setting, where the coarsest grids take very large time steps. It is also worth noting that having a stable integration scheme on both levels (regions of stability left of the red lines) does not indicate a convergent method. Second, the single-iteration bounds shown in the second row, and increase in convergence over general two-grid convergence in the third row, help demonstrate why multilevel convergence is more difficult than two-level. For example, consider the backward Euler case (first column) for a spatial discretization with purely imaginary eigenvalues. As seen in the first row, previous theory [42] guarantees convergence on C-points for all iterations/all points on all but the first iteration. However, we can see in the second row that error over all points for a single iteration will likely be divergent (Corollary 2.4), making the scheme unlikely to be successful in the multilevel setting.

Looking at the difference between single-iteration two-grid convergence and iterations 2, 3, ... gives insight as to why the multilevel setting is more difficult than two-level. Furthermore, in most cases, two-level convergence rates are overly optimistic estimates for multilevel, meaning the modified TAP in Theorem 2.3 and Corollary 2.4 is likely a necessary condition for multilevel convergence. However, despite strong heuristics on multilevel convergence, the connection between two-level convergence rates and multilevel is not rigorous. This motivates the rest of this paper, where multilevel error- and residual-propagation operators are analytically derived, and upper bounds on $\ell^2$-convergence placed directly on these operators.

3. Multilevel residual and error propagation. Denote residual- and error-propagation operators of MGRIT by $R^{nt}$ and $E^{nt}$. As noted in [42], residual and
Fig. 2: Two-grid MGRIT with F-relaxation convergence and $m_0 = 4$ for A- and L-stable SDIRK schemes of order 1 and 2, as a function of spatial eigenvalues $\{\xi_i\}$ in the complex plane. Red lines indicate the stability region of the integration scheme (stable left of the line). The top row shows two-grid convergence rates for all but one iteration (2.10), with the green line marking the boundary of convergence. Similarly, the second row shows single-iteration two-grid convergence (2.11), with blue line marking the boundary of convergence. The final row shows the difference in convergence between single-iteration and further two-grid iterations.

Error propagation are formally similar, that is,

$R_n^m = A_0 E_n^m A_0^{-1} = A_0 (I - M^{-1} A_0) A_0^{-1} = I - A_0 M^{-1},$  \hspace{1cm} (3.1)

where $M^{-1}$ denotes the MGRIT preconditioner for $A_0^{-1}$. Noting that there is a closed form for $A^{-1}$, it follows that if the error-propagation operator of a particular MGRIT
algorithm is known, the residual propagation operator can be easily found by the relation in (3.1), and vice-versa. In this section, we derive the error-propagation operator for generalized two-level MGRIT and multilevel (V-cycle) MGRIT with F- and FCF-relaxation, using the following convention for sums and products: \( \sum_{i=a}^{b} f_i = 0 \) and \( \prod_{i=a}^{b} f_i = 1 \) for \( b < a \). We further write, e.g., \( \mathcal{E}_{rFCF}^{n_r=2} \) to refer to the two-grid error propagation operator and distinguish the notation from a power of \( \mathcal{E} \), and similarly for other numbers of levels \( n_r \).

### 3.1. Two-level MGRIT with rFCF-relaxation

Here, we generalize the two-level error propagator, as given in [5], to two-level MGRIT with rFCF-relaxation. rFCF-relaxation refers to F-relaxation followed by FCF-relaxation steps. A similar result can be found in [17], where MGRIT was interpreted as Parareal with overlap in time.

The error propagator for an exact iterative two-grid method with rFCF-relaxation and \( r \geq 0 \) is given as,

\[
0 = I - A_0^{-1}A_0 = (I - P_0(R_0A_0P_0)^{-1}R_0A_0)(F_0C_0)^{r}F_0
= (I - P_0(R_0A_0P_0)^{-1}R_0A_0)P_0(I - R_0A_0P_0)^{r}R_{t_0},
\]

and MGRIT approximates the coarse-grid operator as \( A_1 \approx R_0A_0P_0 \).

**Lemma 3.1.** The error propagator of two-level MGRIT with rFCF-relaxation and \( r \geq 0 \) is given as,

\[
\mathcal{E}_{rFCF}^{n_r=2} = (I - P_0A_1^{-1}R_0A_0)P_0(I - R_0A_0P_0)^{r}R_{t_0}.
\]

**Proof.** This follows by substituting the coarse-grid operator \( A_1 \approx R_0A_0P_0 \) in Equation (3.2). \( \square \)

### 3.2. Multilevel V-cycles with F-relaxation

The error propagator of a multilevel V-cycle method with F-relaxation can be derived from the error propagator of the exact two-level method on level \( \ell \),

\[
0 = I - A_{\ell}^{-1}A_{\ell} = (I - P_{\ell}(R_{\ell}A_{\ell}P_{\ell})^{-1}R_{\ell}A_{\ell})(I - S_{\ell}(S_{\ell}^{T}A_{\ell}S_{\ell})^{-1}S_{\ell}^{T}A_{\ell})
= I - (P_{\ell}(R_{\ell}A_{\ell}P_{\ell})^{-1}R_{\ell} + S_{\ell}(S_{\ell}^{T}A_{\ell}S_{\ell})^{-1}S_{\ell}^{T})A_{\ell}
\]

and noting that,

\[
A_{\ell}^{-1} = P_{\ell}(R_{\ell}A_{\ell}P_{\ell})^{-1}R_{\ell} + S_{\ell}(S_{\ell}^{T}A_{\ell}S_{\ell})^{-1}S_{\ell}^{T}.
\]

**Lemma 3.2.** The error propagator of a multilevel V-cycle method with F-relaxation is given as,

\[
\mathcal{E}_{F}^{n_r} = P_0R_{t_0} - \left( \prod_{k=0}^{n_{t-2}} P_k \right) A_{n_{t}-1}^{-1} \left( \prod_{k=n_{t}-2}^{0} R_k \right) A_0P_0R_{t_0}
-
\sum_{i=0}^{n_{t}-3} \left( \prod_{k=0}^{i} P_k \right) S_{i+1}(S_{i+1}^{T}A_{i+1}S_{i+1})^{-1}S_{i+1}^{T} \left( \prod_{k=i}^{0} R_k \right) A_0P_0R_{t_0},
\]

for \( n_r \geq 2 \) levels.

**Proof.** See Appendix B. \( \square \)
3.3. Multilevel V-cycles with FCF-relaxation. The error propagator of a multilevel V-cycle method with FCF-relaxation can be derived from the error propagator of the exact two-level method on level $\ell$, 

$$0 = I - A^{-1}_\ell A_t = (I - P_t(R_t A_t P_t)^{-1} R_t A_t) F_t C_t F_t$$

$$= I - P_t(R_t A_t P_t)^{-1} R_t A_t - S_t(S_t^T A_t S_t)^{-1} S_t^T A_t - T_t(T_t^T A_t T_t)^{-1} T_t^T A_t$$

$$+ S_t(S_t^T A_t S_t)^{-1} S_t^T A_t T_t(T_t^T A_t T_t)^{-1} T_t^T A_t + P_t(R_t A_t P_t)^{-1} R_t A_t T_t(T_t^T A_t T_t)^{-1} T_t^T A_t,$$

and noting that,

$$A_t^{-1} = T_t(T_t^T A_t T_t)^{-1} T_t^T + [S_t(S_t^T A_t S_t)^{-1} S_t^T + P_t(R_t A_t P_t)^{-1} R_t] [I - A_t T_t(T_t^T A_t T_t)^{-1} T_t^T].$$

Lemma 3.3. The error propagator of a multilevel V-cycle method with FCF-relaxation is given as,

$$(3.6) \quad \mathcal{E}_t^{FCF} = P_0(I - (T_0^T A_0 T_0)^{-1} R_{l_0} A_0 P_0) R_{l_0} - \left( \prod_{k=0}^{n_t-2} P_k \right) A_{n_t-1}^{-1} \left( \prod_{k=0}^{n_t-2} R_k \left[ I - A_k T_k(T_k^T A_k T_k)^{-1} T_k^T \right] \right) A_0 P_0 R_{l_0}$$

$$- \sum_{i=1}^{n_t-2} \left( \prod_{k=0}^{i-1} P_k \right) S_i(S_i^T A_i S_i)^{-1} S_i^T \left[ I - A_i T_i(T_i^T A_i T_i)^{-1} T_i^T \right]$$

$$+ T_i(T_i^T A_i T_i)^{-1} T_i^T \left( \prod_{k=0}^{i-1} R_k \left[ I - A_k T_k(T_k^T A_k T_k)^{-1} T_k^T \right] \right) A_0 P_0 R_{l_0},$$

with $n_t \geq 2$ levels.

Proof. The proof is analogous to the proof of Lemma 3.2. \qed

3.4. Multilevel F-cycles with rFCF-relaxation. Following [20] (see p. 53), error propagation of MGRIT for a multilevel F-cycle with rFCF-relaxation can be defined recursively,

$$(3.7) \quad \mathcal{F}_F^{n_t} = \mathcal{F}_F^{n_t} = \mathcal{M}_F^{n_t} \mathcal{F}_F^{n_t}, \quad \text{for } n_t \geq 2,$$

with

$$M_{F,\ell-1}^{n_t} = P_{\ell-1} (I - (I - M^{n_t} F_{\ell-1}) A_{\ell-1} R_{\ell-1} A_{\ell-1} P_{\ell-1}) (I - R_{\ell-1} A_{\ell-1} P_{\ell-1})^T R_{l_{\ell-1}},$$

$$M_{F,\ell-1}^{n_t} = P_{\ell-1} (I - (I - M^{n_t} V_{\ell-1}) A_{\ell-1} R_{\ell-1} A_{\ell-1} P_{\ell-1}) (I - R_{\ell-1} A_{\ell-1} P_{\ell-1})^T R_{l_{\ell-1}},$$

for $l = 1, \ldots, n_t - 2$, and,

$$M_{F,\ell-2}^{n_t} = M_{F,\ell-2}^{n_t} = P_{\ell-2} (I - A_{\ell-1}^{-1} R_{\ell-1} A_{\ell-2} P_{\ell-2} A_{\ell-2} P_{\ell-2}) (I - R_{\ell-2} A_{\ell-2} P_{\ell-2})^T R_{l_{\ell-2}}.$$

It is easy to verify, that for $n_t = 2$, the recursive formulae result in 

$$\mathcal{F}_F^{n_t=2} = \mathcal{E}_F^{n_t=2}.$$

For $n_t = 3$ and $r = 0$, we can write,

$$\mathcal{F}_F^{n_t=3} = \mathcal{E}_F^{n_t=2} + P_1 (I - A_2^{-1} R_1 A_1 P_1)^2 R_1 A_1^{-1} R_0 A_0 P_0 R_{l_0}.$$

However, it is not straightforward to convert the recursive definition in (3.7) into a summation similar to (3.5) or (3.3), for arbitrary $n_t$. Yet, this formula is still useful for numerically computing bounds of $\mathcal{F}_F^{n_t}$, hence, we include it.

To refer to F-cycle error propagation, we use the notation $\mathcal{F}_F^{n_t}$ instead of $\mathcal{E}_F^{n_t}$. 

13
4. Bounds for MGRIT residual and error propagation. Following the work in [5], we assume that operators \( \Phi_\ell, \ell = 0, \ldots, n_\ell - 1 \), are all diagonalized by the same eigenvectors,

\[
U \Phi_\ell U^{-1} = \text{diag}(\lambda_{\ell,1}, \ldots, \lambda_{\ell,N_\ell}),
\]

where \( U = (\mathbf{u}_1, \ldots, \mathbf{u}_{N_\ell}) \). We also assume that \( \Phi_\ell \) are stable time stepping operators, that is, \( \|\Phi_\ell\| < 1 \), which implies \( |\lambda_{\ell,k}| < 1 \) for all \( \ell = 0, \ldots, n_\ell - 1 \) and \( k = 1, \ldots, N_\ell \). To simplify notation in the following derivations, we use \( \Phi_\ell \) to denote the diagonalized time stepping operator moving forward. Results then follow in a \((UU^*)^{-1}\)-norm, which (as discussed in Section 2.3) is equivalent to the \( \ell^2 \)-norm if \( \Phi_\ell \) is normal (and, thus, \( U \) is unitary).

For ease of presentation and because many of the derivations are fairly involved, but repetitive, results in this section are stated without proof. For proofs, we refer the interested reader to Appendix C.

4.1. Residual and error on level 0 and level 1. It is typically difficult or impossible in practical applications to precisely measure the error propagation of an iterative method or the error itself. It is, however, possible to measure the residual, and stopping criteria for iterative methods are often based on a residual tolerance. In the case of MGRIT, there is a nice relation between error and residual propagation. From Corollary 2.4, the norm of residual- and error-propagation operators are equal in the \((UU^*)^{-1}\)-norm (recall, \( U \) is a block diagonal matrix of eigenvectors, \( U \)).

\[\text{If } \{\Phi_\ell\} \text{ are normal operators, they are diagonalizable by unitary transformation, in which case } UU^* = I, \text{ and error and residual propagation are equal in the } \ell^2 \text{-norm.}\]

Similar to Section 3, let \( \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} \) be the error propagator on level 0. We further refer to \( \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} \) as the error propagator on level 1, that denotes the operator that acts on the error at C-points on level 0, i.e., all points on level 1. In the two-grid setting, we also refer to \( \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} \) as the coarse-grid error propagator.

To quantify how fast MGRIT converges in the worst case, we can bound the convergence factor of the fast grid residual [5] \( r_{i+1} \) at iteration \( i + 1, i \in \mathbb{N}_0 \), by the norm of the error propagator on level 1 (in the unitary case),

\[
\|r_{i+1}\|_2 / \|r_i\|_2 = \|A_1 \mathbf{e}^\Delta_{i+1}\|_2 / \|A_1 \mathbf{e}^\Delta_i\|_2 \leq \|A_1 \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} A_1^{-1}\|_2 = \|\mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}}\|_2,
\]

where \( \mathbf{e}^\Delta_{i+1} \) is the error on level 1 or equivalently, error at C-points on level 0. With,

\[\mathbf{e}^\Delta_{i+1} = \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} A_1^{-1} R_{I_0} \mathbf{e}_i, \quad \Leftrightarrow \quad P_0 \mathbf{e}^\Delta_{i+1} = P_0 \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} R_{I_0} \mathbf{e}_i,
\]

we can identify, \( \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} = R_{I_0} \mathcal{E}^{n_{r_{FCF}}} P_0 \), which is a generalization of the approach in [5], where the operators \( P_0 \) and \( R_{I_0} \) are pulled out to the left and right of the error propagator. Thus, in general we analyze the error propagator on level 1 to bound residual propagation on level 0, as given in (4.1).

This raises the question of how the error develops at the F-points on the fine grid. Considering error propagation on level 0 over \( k \) iterations,

\[\mathbf{e}_{i+1} = \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} \mathbf{e}_i = \ldots = (\mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}})^{i+1} \mathbf{e}_0 = \left( P_0 \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} R_{I_0} \right)^{i+1} \mathbf{e}_0 = P_0 \left( \mathcal{E}^{n_{r_{FCF}}}_{r_{FCF}} \right)^{i+1} R_{I_0} \mathbf{e}_0,
\]

we find that error propagation at the F-points of the fine grid can be bounded by error propagation at the respective C-points times a constant.

---

\[\text{Although Corollary 2.4 specifically addresses two-grid bounds, equality of error and residual propagation in the } (UU^*)^{-1}\text{-norm follows if } \Phi_\ell \text{ is simultaneously diagonalizable for all levels } \ell.\]
Lemma 4.1. Error propagation on level 0 for an MGRIT V-cycle method can be bounded by error propagation on level 1,

\[ \| E_{rFCF}^{n_0 \Delta} \|_2 \leq \sqrt{m_0} \| E_{rFCF}^{n_1 \Delta} \|_2, \]

with temporal coarsening factor \( m_0 \) on level 0.

Remark 4.2. It is clear from Lemma 4.1, that if the error at C-points on level 0 converges, then the error at F-points on level 1 converges as well. This is the basis for the theory developed in the rest of this work, where convergence is attained by bounding \( E_{rFCF}^{n_0 \Delta} \) in norm.

Lemma 4.1 is intuitive in the sense that the fine grid error propagation is a direct result of the level 1 error propagation; it is simply ideal interpolation applied to the level 1 error; that is, the operator \( P_0 \) propagates the error at the C-points on level 0 to the subsequent F-points. A similar result was presented in [42] for two-level convergence of Parareal and MGRIT.

Based on the formulae derived in Section 3, we can construct residual and error propagators numerically and bound the worst case convergence factor of MGRIT from above via

\[ c_f = \max_i \| r_{i+1} \|_2 / \| r_i \|_2 \leq \| E_{rFCF}^{n_0 \Delta} \|_2, \]

which corresponds to the maximum singular value of \( E_{rFCF}^{n_0 \Delta} \). In practice, the dimension of \( E_{rFCF}^{n_0 \Delta} \) grows with the problem size in space and time, \( N_x \) and \( N_0 \). Similarly, \( E_{rFCF}^{n_0 \Delta} \) grows with \( N_x \) and \( N_1 \). Depending on the available resources, numerical construction and investigation of these operators may be limited by memory consumption and/or compute time. To that end, it is desirable to derive further cheaper upper bounds that enable fast assessment of MGRIT convergence for larger space-time problem sizes. In the following, we present several bounds and approximate convergence factors for fine-grid residual propagation and error propagation on level 1.

4.2. Upper bound using inequality. One straightforward way to reduce computational cost by roughly one order of magnitude is bounding the \( \ell^2 \)-norm of the error propagator on level 1 using the well-known inequality [24],

\[ \| E_{rFCF}^{n_0 \Delta} \|_2 \leq \| E_{rFCF}^{n_0 \Delta} \|_1 / \| E_{rFCF}^{n_0 \Delta} \|_\infty. \]

In [5], this was used to develop an upper-bound on two-grid convergence, which was proven to be sharp in [42]. This section extends this approach to three and four grid levels based on analytic formulae. Although the sharpness of (4.3) suffers as the number of levels increases (see Section 5), we show that it is still reasonably sharp and provides a useful tool to analyze MGRIT convergence a priori.

4.2.1. Two-level MGRIT with \( rFCF \)-relaxation. The coarse-grid error propagator follows from Equation (3.3) with \( n_f = 2 \) (see Appendix C.2),

\[ E_{rFCF}^{n_0 \Delta} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ (\Phi_{0m_0} - \Phi_0) \Phi_{0m_0} \\ \Phi_{0m_0} - \Phi_0 \\ \Phi_{0m_0} - \Phi_0 \\ \Phi_{0m_0} - \Phi_0 \\ \Phi_{0m_0} - \Phi_0 \end{bmatrix}, \]

where the first \( r + 1 \) rows and last \( r + 1 \) columns are zero.
Theorem 4.3. Let $\Phi_\ell$ be simultaneously diagonalizable by the same unitary transformation $X$, with eigenvalues $\lambda_{\ell,k} |\lambda_{\ell,k}| < 1$. Then, the worst case convergence factor of the fine-grid residual of two-level MGRIT with rFCF-relaxation is bounded by,

$$c_f \leq \max_{1 \leq k \leq N_x} |\lambda_{0,k} - \lambda_{1,k}| \|\lambda_{0,k}^{-1} \frac{1 - |\lambda_{1,k}|^{N_1 - 1 - r}}{1 - |\lambda_{1,k}|}.$$

Remark 4.4. The cases of F- and FCF-relaxation (i.e., $r = 0$ and $r = 1$), yield the result in [5],

$$\|E_{n=2,\Delta}\|_2 \leq \max_{1 \leq k \leq N_x} |\lambda_{0,k} - \lambda_{1,k}| \|\lambda_{0,k}^{-1} \frac{1 - |\lambda_{1,k}|^{N_1 - 2}}{1 - |\lambda_{1,k}|}.$$

Remark 4.5. In [42], it was shown that the bound in Theorem 4.3 is exact to $O(1/N_1)$ for F- and FCF-relaxation.

An interesting observation of (4.4) is the fact that the coarse-grid error propagator is nilpotent and that each block can be diagonalized by the same unitary transformation. This implies that we can re-order the rows and columns of the coarse-grid error propagator, yielding a block diagonal form with lower triangular nilpotent blocks.

Lemma 4.6. Let $\Phi_0$ and $\Phi_1$ be simultaneously diagonalizable by the same unitary transformation $X$, with eigenvalues $\lambda_{\ell,k} |\lambda_{\ell,k}| < 1$. Then, the $\ell^2$-norm of the coarse-grid error propagator of two-level MGRIT with rFCF-relaxation can be computed as,

$$\|E_{\ell=2,\Delta}\|_2 = \sup_{1 \leq k \leq N_x} \|\tilde{E}_{\ell=2,\Delta}(k)\|_2,$$

with the coarse-grid error propagator $\tilde{E}_{\ell=2,\Delta}(k)$ for a single spatial mode $k$ with $1 \leq k \leq N_x$.

Proof. This follows from the discussion above and the fact that the spectral norm of a block diagonal operator with lower triangular blocks can be computed as the supremum of the spectral norm of all lower triangular blocks. See also [5], Remark 3.1.

Remark 4.7. Lemma 4.6 implies that computing a bound of the form of (4.2) can be parallelized over the number of spatial modes. Thus, the time complexity of evaluating (4.2) is $O(N_x N_1^3/p)$ with $1 \leq p \leq N_x$ parallel processors.

Remark 4.8. The result in Lemma 4.6 is not limited to $n_\ell = 2$ and can be applied to all subsequent convergence results.

4.2.2. Three-level V-cycles with F-relaxation. Evaluating the error propagator in Equation (3.5) for a three-level V-cycle with F-relaxation on level 1 (see Equation (C.5)) and comparison with the two-level error propagator for F-relaxation in [5] highlights a slight complication: In general, the maximum absolute column sum (and similarly, for the maximum absolute row sum) is no longer given by the first column. Instead, the maximum absolute column sum is given by the maximum of the first $m_1$ absolute column sums, corresponding to the first CF-interval (first C-point and first $m_1 - 1$ F-points) on level 1. This structure arises because of the recursive partitioning of time points into F- and C-points on each level.
Theorem 4.9. Let \( \Phi_f \) be simultaneously diagonalizable by the same unitary transformation \( X \), with eigenvalues \( \lambda_{\ell,k}, \quad |\lambda_{\ell,k}| < 1 \). Then, the worst case convergence factor of three-level MGRIT with \( F \)-relaxation is bounded by,

\[
c_f \leq \sqrt{\|E^{\ell=3,\Delta}_F\|_1 \|E^{\ell=3,\Delta}_F\|_\infty},
\]

and \( \|E^{\ell=3,\Delta}_F\|_1 \) and \( \|E^{\ell=3,\Delta}_F\|_\infty \) are given analytically as,

\[
\|E^{\ell=3,\Delta}_F\|_1 = \max_{1 \leq s \leq N_x} \left| \lambda_{2,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \right| \left( \lambda_{2,k} |N_2^{-2} - \frac{1-|\lambda_{2,k}|}{1-|\lambda_{1,k}|} |N_2^{-2} |1 - |\lambda_{1,k}|^{m_1} \right) + \left| \lambda_{1,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \right| |1 - |\lambda_{1,k}|^{m_1} |1 - |\lambda_{1,k}| \right|
\]

and

\[
\|E^{\ell=3,\Delta}_F\|_\infty = \max_{1 \leq s \leq N_x} \left| \lambda_{2,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \right| \left( \lambda_{2,k} |N_2^{-2} - \frac{1-|\lambda_{2,k}|}{1-|\lambda_{1,k}|} |N_2^{-2} |1 - |\lambda_{1,k}|^{m_1} \right) + \left| \lambda_{1,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \right| |1 - |\lambda_{1,k}|^{m_1} |1 - |\lambda_{1,k}| \right|_{j=1,\ldots,m_1-1}.
\]

4.2.3. Four-level V-cycles with F-relaxation. Evaluating the error propagator in Equation (3.5) for a four-level V-cycle with \( F \)-relaxation on level 1 yields the following result.

Theorem 4.10. Let \( \Phi_f \) be simultaneously diagonalizable by the same unitary transformation \( X \), with eigenvalues \( \lambda_{\ell,k}, \quad |\lambda_{\ell,k}| < 1 \). Then, the worst case convergence factor of four-level MGRIT with \( F \)-relaxation is bounded by,

\[
c_f \leq \sqrt{\|E^{\ell=4,\Delta}_F\|_1 \|E^{\ell=4,\Delta}_F\|_\infty},
\]

and \( \|E^{\ell=4,\Delta}_F\|_1 \) and \( \|E^{\ell=4,\Delta}_F\|_\infty \) are given analytically as,

\[
\|E^{\ell=4,\Delta}_F\|_1 = \max_{1 \leq s \leq N_x} s_d^{\text{col}}(k), \quad \|E^{\ell=4,\Delta}_F\|_\infty = \max_{1 \leq s \leq N_x} s_d^{\text{row}}(k),
\]

where the column and row sums, \( s_d^{\text{col}} \) and \( s_d^{\text{row}} \) (row and column subscripts \( d \)), are defined as follows. The absolute column sums of the first CF-interval on level 1 are given as,

\[
s_0^{\text{col}}(k) = |\lambda_{3,k} |N_2^{-2} |\lambda_{0,k} + \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \lambda_{2,k}^{m_2-1} | + |\lambda_{3,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \lambda_{2,k}^{m_2-1} | \left( 1 - |\lambda_{3,k}|^{m_2} \right) + |\lambda_{2,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} \lambda_{2,k}^{m_2-1} | \left( 1 - |\lambda_{3,k}| |N_2^{-2} - \frac{1-|\lambda_{2,k}|}{1-|\lambda_{1,k}|} |N_2^{-2} \right)
\]

and corresponding to the first C-point on level 1. Next,

\[
s_{m_1(m_2-j)}^{\text{col}}(k) = |\lambda_{2,k} - \lambda_{0,k}^{m_0} \lambda_{1,k}^{m_1-1} | \left( \sum_{p=0}^{m_2-2} |\lambda_{2,k}|^{p} \right) \left( \frac{1-|\lambda_{2,k}|}{1-|\lambda_{1,k}|} \right) + |\lambda_{1,k} - \lambda_{0,k}^{m_0} \lambda_{2,k}^{m_1-1} | \left( \frac{1-|\lambda_{1,k}|^{m_1}}{1-|\lambda_{1,k}|} \right),
\]
+ |λ_{2,k}|^{j-1}|λ_{2,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}| \left[ |λ_{3,k}|^{N_3-2} + \frac{1-|λ_{3,k}|^{N_3-2}}{1-|λ_{2,k}|} \right] \left[ |λ_{3,k}|^{N_3-2} - \frac{1-|λ_{3,k}|^{N_3-2}}{1-|λ_{2,k}|} \right]

for j = 1, \ldots, m_2 - 1, corresponding to the interior level 2 C-points of the first CF-interval on level 1. Lastly,

s^{col}_{m_1(m_2-j)-r-1}(k) = |λ_{1,k} - λ_{0,k}^{m_0}| \left[ \left( \sum_{q=0}^{m_1-1} |λ_{1,k}|^q \right) \left( \sum_{p=0}^{j-1} |λ_{2,k}|^p \right) + \left( \sum_{q=0}^{r-1} |λ_{1,k}|^q \right) \right]

+ |λ_{2,k}|^j |λ_{1,k} - λ_{0,k}^{m_0}| \left[ |λ_{3,k}|^{N_3-2} - \frac{1-|λ_{3,k}|^{N_3-2}}{1-|λ_{2,k}|} \right]

for j = 0, \ldots, m_2 - 1 and r = 0, \ldots, m_1 - 2, corresponding to the level 2 F-points of the first CF-interval on level 1.

The absolute row sums of the last FC-interval on level 1 are given as,

s^{row}_{N_1-1}(k) = \frac{1 - |λ_{3,k}|^{N_3-1}}{1 - |λ_{3,k}|} \left[ |λ_{1,k} - λ_{0,k}^{m_0}| \frac{1 - |λ_{2,k}|^{m_2}}{1 - |λ_{1,k}|} + |λ_{2,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}| \right]

+ |λ_{2,k}|^j \frac{1 - |λ_{3,k}|^{N_3-2}}{1 - |λ_{3,k}|} \left[ |λ_{1,k} - λ_{0,k}^{m_0}| \frac{1 - |λ_{2,k}|^{m_2}}{1 - |λ_{1,k}|} + |λ_{2,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}| \right]

+ |λ_{2,k}| \left( \sum_{q=0}^{j-1} |λ_{2,k}|^q \right) \left[ |λ_{3,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}λ_{2,k}^{m_2-1}| \right]

+ |λ_{1,k}| \left( \sum_{q=0}^{r-1} |λ_{1,k}|^q \right) \left[ |λ_{3,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}λ_{2,k}^{m_2-1}| \right]

for j = 1, \ldots, m_2 - 1, corresponding to the interior C-points of the last FC-interval on level 1. Lastly,

s^{row}_{N_1-1-m_1m_2+r+jm_1}(k) = \frac{1 - |λ_{3,k}|^{N_3-1}}{1 - |λ_{3,k}|} \left[ |λ_{1,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}| \frac{1 - |λ_{2,k}|^{m_2}}{1 - |λ_{3,k}|} + |λ_{3,k} - λ_{0,k}^{m_0}λ_{1,k}^{m_1-1}λ_{2,k}^{m_2-1}| \right]

for j = 0, \ldots, m_2 - 1 and r = 1, \ldots, m_1 - 1, corresponding to the F-points of the last FC-interval on level 1.

Remark 4.11. We note, that evaluating the 2m_1m_2 analytic formulae in Theorem 4.10 significantly reduce the time complexity of evaluating Equation (4.7) compared to constructing $c_F^{n_2=4,\Delta}$ numerically and computing $\|c_F^{n_2=4,\Delta}\|_1$ and $\|c_F^{n_2=4,\Delta}\|_\infty$.

4.2.4. Three-level V-cycles with FCF-relaxation. Following the same approach as in the previous sections, we can find the following result for three-level V-cycles with FCF-relaxation.
Theorem 4.12. Let $\Phi_t$ be simultaneously diagonalizable by the same unitary transformation $X$, with eigenvalues $\lambda_{t,k}$, $|\lambda_{t,k}| < 1$. Then, the worst case convergence factor of three-level MGRIT with FCF-relaxation is bounded by,

$$c_f \leq \sqrt{\|E^{n_1=3,\Delta}_F\|_1 \|E^{n_1=3,\Delta}_{FCF}\|_\infty},$$

and $\|E^{n_1=3,\Delta}_F\|_1$ and $\|E^{n_1=3,\Delta}_{FCF}\|_\infty$ are given analytically as,

$$\|E^{n_1=3,\Delta}_{FCF}\|_1 = \max_{1 \leq k \leq N_x} s^\text{col}_d(k), \quad \|E^{n_1=3,\Delta}_{FCF}\|_\infty = \max_{0 \leq d \leq m_1 m_2 - 1} s^\text{row}_d(k).$$

The absolute column sums of the first CF-interval on level 1 are given as,

$$s^\text{col}_0(k) = |\lambda_{0,k}|^{m_0} |\lambda_{1,k} - \lambda_{0,k}| \left| |\lambda_{1,k}|^{m_1 - 1} \left( |\lambda_{2,k}|^{N_2 - 3} + \frac{1 - |\lambda_{2,k}|^{N_2 - 3}}{1 - |\lambda_{1,k}|} \right) \right|,$$

corresponding to the last F-point on level 1. Next,

$$s^\text{col}_{m_1 - 2}(k) = |\lambda_{0,k}|^{m_0} |\lambda_{1,k} - \lambda_{0,k}| \left| |\lambda_{1,k}|^{m_1 - 1} \left( |\lambda_{2,k}|^{N_2 - 3} + \frac{1 - |\lambda_{2,k}|^{N_2 - 3}}{1 - |\lambda_{1,k}|} \right) \right|,$$

corresponding to the first C-point on level 1 if $m_1 = 2$, or the penultimate F-point on level 1 if $m_1 > 2$. Lastly, if $m_1 > 2$,

$$s^\text{col}_{m_1 - 2 - j}(k) = |\lambda_{0,k}|^{m_0} |\lambda_{1,k} - \lambda_{0,k}| \left| |\lambda_{1,k}|^{m_1 - 1} \left( |\lambda_{2,k}|^{N_2 - 2} + \frac{1 - |\lambda_{2,k}|^{N_2 - 2}}{1 - |\lambda_{1,k}|} \right) \right|,$$

for $j = 1, \ldots, m_1 - 2$, corresponding to the first C-point and the following F-points on level 1,

The absolute row sums of the last FC-interval on level 1 are given as,

$$s^\text{row}_{N_1 - 1}(k) = |\lambda_{0,k}|^{m_0} |\lambda_{1,k} - \lambda_{0,k}| \left| |\lambda_{1,k}|^{m_1 - 1} \left( |\lambda_{2,k}|^{N_2 - 2} + \frac{1 - |\lambda_{2,k}|^{N_2 - 2}}{1 - |\lambda_{1,k}|} \right) \right|,$$

corresponding to the last C-point on level 1, and,

$$s^\text{row}_{N_1 - m_1 + j}(k) = |\lambda_{0,k}|^{m_0} \left| |\lambda_{1,k} - \lambda_{0,k}|^{j+2} \left( |\lambda_{2,k} - \lambda_{0,k}^{m_1 - 1} | + |\lambda_{1,k} - \lambda_{0,k}|^{j} \right) \right|,$$

for $j = 0, \ldots, m_1 - 2$, corresponding to the preceding F-points on level 1.

4.3. Approximate convergence factor of multilevel V-cycle algorithm. In Section 4.2, we have presented analytic formulae as exact representations of the inequality bound (4.3). These a priori convergence bounds reduce memory consumption and computational cost significantly. It is, however, increasingly difficult to derive such analytic formulae for larger numbers of levels. Thus, we propose an analytic approximate convergence factor for multilevel V-cycles with F- and FCF-relaxation, that takes the eigenvalues $\lambda_{t,k}$ of the time stepping operator $\Phi_t$, the number of time points...
transformation factor of multilevel MGRIT V-cycles with F-relaxation is given as, analytic formulae in Section 4.2 in an approximate manner. The proposed approximate convergence factors are based on approximating the inequality bound (4.3), and therefore, are expected to be a conservative upper bound in a large number of cases. More specifically, the multilevel formulae generalize the analytic formulae in Section 4.2 in an approximate manner.

First, we present the approximate convergence factor for multilevel V-cycles with F-relaxation.

Approximation 1. Let $\Phi_\ell$ be simultaneously diagonalizable by the same unitary transformation $X$, with eigenvalues $\lambda_{\ell,k}$, $|\lambda_{\ell,k}| < 1$. Then, the approximate convergence factor of multilevel MGRIT V-cycles with F-relaxation is given as,

$$c_\ell \approx \max_{1 \leq k \leq N_x} \sqrt{s_{0}^{\text{col}}(k, n_\ell) s_{N_{\ell}-1}^{\text{col}}(k, n_\ell) \approx \sqrt{\|E^\Delta_{F_{\ell}}\|_1 \|E^\Delta_{F_{\ell}}\|_\infty},$$

with approximate maximum absolute column and row sum,

$$s_{0}^{\text{col}}(k, n_\ell) \approx \sum_{\ell=1}^{n_\ell-1} |\lambda_{\ell,k} - \lambda_{0,k}^m | \left( \prod_{p=1}^{\ell-1} \lambda_{p,k}^{m_p-1} \right) \left( \prod_{q=1}^{\ell} \frac{1 - |\lambda_{q,k}|^{m_q-1}}{1 - |\lambda_{q,k}|} \right),$$

$$s_{N_{\ell}-1}^{\text{row}}(k, n_\ell) \approx \sum_{\ell=1}^{n_\ell-1} |\lambda_{\ell,k} - \lambda_{0,k}^m | \left( \prod_{p=1}^{\ell-1} \lambda_{p,k}^{m_p-1} \right) \left( \prod_{q=\ell}^{n_\ell-1} \frac{1 - |\lambda_{q,k}|^{m_q}}{1 - |\lambda_{q,k}|} \right),$$

with $m_{\ell} = [m_0, \ldots, m_{n_\ell-2}, N_{n_{\ell-1}} - 1]^T$. Furthermore, in many cases, $\|E^\Delta_{F_{\ell}}\|_2 \leq c_\ell$, because the analytic formulae approximate the right-hand-side of Equation (4.3) (compare Equation (4.8)).

A similar result can be formulated for multilevel V-cycles with FCF-relaxation.

Approximation 2. Let $\Phi_\ell$ be simultaneously diagonalizable by the same unitary transformation $X$, with eigenvalues $\lambda_{\ell,k}$, $|\lambda_{\ell,k}| < 1$. Then, the approximate convergence factor of multilevel MGRIT V-cycles with FCF-relaxation is given as,

$$c_\ell \approx \max_{1 \leq k \leq N_x} \sqrt{s_{0}^{\text{col}}(k, n_\ell) s_{N_{\ell}-1}^{\text{col}}(k, n_\ell) \approx \sqrt{\|E^\Delta_{FCF_{\ell}}\|_1 \|E^\Delta_{FCF_{\ell}}\|_\infty},$$

with approximate maximum absolute column and row sum,

$$s_{0}^{\text{col}}(k, n_\ell) \approx (n_{t} > 2) \cdot |\lambda_{0,k}^m | \left| \lambda_{1,k} - \lambda_{0,k}^m \right| \frac{1 - |\lambda_{1,k}|^{m_1}}{1 - |\lambda_{1,k}|} \frac{1 - |\lambda_{0,k}|^{m_0}}{1 - |\lambda_{0,k}|} \left( \prod_{p=1}^{n_{t}-1} |\lambda_{p,k}^{m_p} | \left( \prod_{q=1}^{n_{t}-2} \frac{1 - |\lambda_{q,k}|^{m_q}}{1 - |\lambda_{q,k}|} \right) \right),$$

$$s_{N_{\ell}-1}^{\text{row}}(k, n_\ell) \approx (n_{t} > 2) \cdot |\lambda_{0,k}^m | \left| \lambda_{1,k} - \lambda_{0,k}^m \right| \frac{1 - |\lambda_{1,k}|^{m_1}}{1 - |\lambda_{1,k}|} \frac{1 - |\lambda_{0,k}|^{m_0}}{1 - |\lambda_{0,k}|} \left( \prod_{p=1}^{n_{t}-1} |\lambda_{p,k}^{m_p} | \left( \prod_{q=1}^{n_{t}-2} \frac{1 - |\lambda_{q,k}|^{m_q}}{1 - |\lambda_{q,k}|} \right) \right).$$

\(^{11}\)The generalization of Lemma 4.6 implies that time complexity is in fact $O(N_{t}/p)$ with $1 \leq p \leq N_x$ parallel processors.
Furthermore, in many cases, \( \|\mathbf{e}_{FCF}^{n+1}\|_2 \leq \hat{c}_F \), because the analytic formulae approximate the right-hand-side of Equation (4.3) (compare Equation (4.9)).

5. Numerical results. In this section, we evaluate the derived bounds and approximate convergence factors for various model problems. In particular, we assess how sharp the various upper bounds are and how much sharpness is sacrificed by employing a bound that is cheaper to compute numerically. For all results, we consider Runge-Kutta time-integration schemes [21, 22] of orders 1-4 (Butcher tableaux provided in Appendix D). In [5], it was noted that in the two-level setting, L-stable schemes seem to be better suited for parallel-in-time integration than A-stable schemes. Here, we review this observation in the multilevel setting. We further investigate the difference between V- and F-cycle convergence, as well as the effect of F- and FCF-relaxation.

For all cases, the number of time grids varies between two and six levels. The fine grid is composed of \( N_0 = 1025 \) time points and the temporal coarsening factor is \( m_\ell = 2 \) between all levels. The spatial domain is two-dimensional and discretized using 11 nodes in each coordinate direction (grid spacing \( \delta_x \)). Derived bounds and approximate convergence factors are compared with the maximum observed convergence factor in numerical simulations, based on the \( L^2 \)-norm of the residual \( \|r_i\|_2 \) at iteration \( i \) (see Equation (4.1)),

\[
\max_i \|r_{i+1}\|_2 / \|r_i\|_2.
\]

All test cases are implemented in C++, using the open-source libraries Armadillo [35, 36] and XBraid [46], a non-intrusive implementation of the MGRIT algorithm. The absolute stopping tolerance for MGRIT is selected as \( \|r_i\|_2 < 10^{-11} \) and the initial global space-time guess is random.

5.1. Diffusion equation. Consider the general time-dependent diffusion equation in two spatial dimensions over domain \( \mathbf{x} \in \Omega = (0, 2\pi) \times (0, 2\pi) \),

\[
\partial_t u = \nabla \cdot [K \nabla u] \quad \text{for} \quad \mathbf{x} \in \Omega, \ t \in (0, 2\pi],
\]

with homogeneous boundary and discontinuous initial condition (see Figure 3),

\[
\begin{align*}
  u(\mathbf{x}, \cdot) &= 0 \quad \text{for} \ \mathbf{x} \in \partial \Omega, \\
  u(\cdot, 0) &= 1 - \max \left\{ \text{sign}(\{4 - (x_1 - \pi + 1)^2 - 4(x_2 - \pi)^2\}^2 + 1.2(1 + \pi - x_1)^3 - 10), 1 \right\} \\
  &\quad \text{for} \ \mathbf{x} \in \Omega \cup \partial \Omega,
\end{align*}
\]

for a scalar solution \( u(\mathbf{x}, t) \) and boundary \( \partial \Omega \). Here, \( K = \text{diag}(k_1, k_2) = \text{const} \) is the grid-aligned conductivity tensor. If \( k_1 = k_2 \), the problem is isotropic, while if \( k_1 \ll k_2 \) or \( k_2 \ll k_1 \), the problem is anisotropic. The spatial problem is discretized using second-order centered finite differences, in which case the time-stepping operators \( \Phi_\ell \) are unitarily diagonalizable.

5.1.1. Isotropic diffusion. First, we consider the isotropic case with \( k_1 = k_2 = 10 \). The CFL number on each level,

\[
\text{CFL}_\ell = 2\pi/(N_\ell - 1) \left( k_1/\delta_x^2 + k_2/\delta_y^2 \right) = 4\pi k_1/\delta_x^2(N_\ell - 1),
\]

where \( \delta_x = \delta_y = \delta \) and \( N_\ell \) is the number of levels. The CFL numbers on each level are

\[
\text{CFL}_\ell = \frac{2\pi}{(N_\ell - 1)\left(\frac{k_1}{\delta_x^2} + \frac{k_2}{\delta_y^2}\right)} = \frac{4\pi k_1}{\delta_x^2(N_\ell - 1)}.
\]

In this section, we evaluate the derived bounds and approximate convergence factors for various model problems. In particular, we assess how sharp the various upper bounds are and how much sharpness is sacrificed by employing a bound that is cheaper to compute numerically. For all results, we consider Runge-Kutta time-integration schemes [21, 22] of orders 1-4 (Butcher tableaux provided in Appendix D). In [5], it was noted that in the two-level setting, L-stable schemes seem to be better suited for parallel-in-time integration than A-stable schemes. Here, we review this observation in the multilevel setting. We further investigate the difference between V- and F-cycle convergence, as well as the effect of F- and FCF-relaxation.

For all cases, the number of time grids varies between two and six levels. The fine grid is composed of \( N_0 = 1025 \) time points and the temporal coarsening factor is \( m_\ell = 2 \) between all levels. The spatial domain is two-dimensional and discretized using 11 nodes in each coordinate direction (grid spacing \( \delta_x \)). Derived bounds and approximate convergence factors are compared with the maximum observed convergence factor in numerical simulations, based on the \( L^2 \)-norm of the residual \( \|r_i\|_2 \) at iteration \( i \) (see Equation (4.1)),

\[
\max_i \|r_{i+1}\|_2 / \|r_i\|_2.
\]

All test cases are implemented in C++, using the open-source libraries Armadillo [35, 36] and XBraid [46], a non-intrusive implementation of the MGRIT algorithm. The absolute stopping tolerance for MGRIT is selected as \( \|r_i\|_2 < 10^{-11} \) and the initial global space-time guess is random.

5.1. Diffusion equation. Consider the general time-dependent diffusion equation in two spatial dimensions over domain \( \mathbf{x} \in \Omega = (0, 2\pi) \times (0, 2\pi) \),

\[
\partial_t u = \nabla \cdot [K \nabla u] \quad \text{for} \quad \mathbf{x} \in \Omega, \ t \in (0, 2\pi],
\]

with homogeneous boundary and discontinuous initial condition (see Figure 3),

\[
\begin{align*}
  u(\mathbf{x}, \cdot) &= 0 \quad \text{for} \ \mathbf{x} \in \partial \Omega, \\
  u(\cdot, 0) &= 1 - \max \left\{ \text{sign}(\{4 - (x_1 - \pi + 1)^2 - 4(x_2 - \pi)^2\}^2 + 1.2(1 + \pi - x_1)^3 - 10), 1 \right\} \\
  &\quad \text{for} \ \mathbf{x} \in \Omega \cup \partial \Omega,
\end{align*}
\]

for a scalar solution \( u(\mathbf{x}, t) \) and boundary \( \partial \Omega \). Here, \( K = \text{diag}(k_1, k_2) = \text{const} \) is the grid-aligned conductivity tensor. If \( k_1 = k_2 \), the problem is isotropic, while if \( k_1 \ll k_2 \) or \( k_2 \ll k_1 \), the problem is anisotropic. The spatial problem is discretized using second-order centered finite differences, in which case the time-stepping operators \( \Phi_\ell \) are unitarily diagonalizable.

5.1.1. Isotropic diffusion. First, we consider the isotropic case with \( k_1 = k_2 = 10 \). The CFL number on each level,

\[
\text{CFL}_\ell = 2\pi/(N_\ell - 1) \left( k_1/\delta_x^2 + k_2/\delta_y^2 \right) = 4\pi k_1/\delta_x^2(N_\ell - 1),
\]

where \( \delta_x = \delta_y = \delta \) and \( N_\ell \) is the number of levels. The CFL numbers on each level are

\[
\text{CFL}_\ell = \frac{2\pi}{(N_\ell - 1)\left(\frac{k_1}{\delta_x^2} + \frac{k_2}{\delta_y^2}\right)} = \frac{4\pi k_1}{\delta_x^2(N_\ell - 1)}.
\]
ranges between CFL_0 ≈ 0.376 on level 0 and CFL_5 ≈ 12.036 on level 5. Results for F-relaxation are shown in Figure 4 and FCF relaxation in Figure 5 (note the difference in y-axis limits).

In the case of F-relaxation, there is a considerable difference in convergence behavior between the A-stable and L-stable Runge-Kutta schemes. For A-stable schemes, convergence of MGRIT deteriorates with a growing number of time grid levels, which corresponds to a growing CFL number on the coarse grid, and eventually diverges. On the other hand, L-stable schemes show a less dramatic increase in the convergence factor. In fact, the estimated and observed convergence factors plateau for V-cycle algorithms with L-stable time integration. For F-cycle algorithms with F-relaxation and L-stable schemes, observed convergence is flat for all considered time grid hierarchies and only a slight increase can be observed in the upper bound values and approximate convergence factor.

In the case of FCF-relaxation, all observed convergence factors for SDIRK orders 2-4 are constant with respect to number of levels, and only a slight increase in convergence factor occurs for SDIRK1. FCF-relaxation was shown to be a critical ingredient for a scalable multilevel solver in [8]. An important observation for F-cycle convergence is that all upper bounds predict constant convergence factors, suggesting that an MGRIT algorithm with F-cycles and FCF-relaxation yields a robust and scalable multilevel solver for the isotropic diffusion equation.

In general, all upper bounds and approximate convergence factors provide good qualitative a priori estimates of the observed convergence. These estimates become less sharp for larger numbers of time grid levels, but the estimates do appear to be robust across changes in time integration order. Furthermore, note that Approximation 1 and Approximation 2 estimate observed convergence as well or better than more expensive upper bounds, demonstrating their applicability and efficacy. Overall, results in this section demonstrate that theoretical results presented in this work provide a valuable tool for designing robust and scalable multilevel solvers. It further provides guidance to avoid less optimal parameter choices for MGRIT, such as F-relaxation with A-stable RK schemes.

5.1.2. Anisotropic diffusion. In this section, we investigate the anisotropic diffusion case for the L-stable SDIRK1 scheme, backward Euler, to assess how sensitive the estimates are with respect to conductivity parameters. Conductivity parameters are given as k_1 = 0.5 and k_2 = 0.001, and the CFL number on each level, CFL_ℓ = 2π(k_1 + k_2)/[σ^2(N_ℓ - 1)], ranges between CFL_0 ≈ 0.009 on level 0 and CFL_5 ≈ 0.302 on level 5. Results are presented in Figure 6.
For V-cycle algorithms with F- and FCF-relaxation, the estimated and observed convergence factors grow with the number of grid levels, similar to the isotropic case. Again, FCF-relaxation yields a quicker plateauing of the observed convergence factor. On the other hand, for F-cycle algorithms with F- and FCF-relaxation, observed and estimated convergence are effectively constant. This is indicative that convergence in the case of V-cycles is limited by solving the coarse-grid problem well, but that the non-Galerkin coarse-grid operator (that is, taking larger time steps on the coarse grid using the same integration scheme) is indeed an effective preconditioner. Note, this behavior differs compared with using multigrid to solve anisotropic diffusion discretizations in the spatial setting, where stronger cycles such as F- and W-cycles often do not improve convergence [31].

The approximate bounds on convergence of F-cycles are fairly sharp for F- and FCF-relaxation and all numbers of levels tested. In the case of V-cycles, the bounds and approximate convergence factors loose sharpness as the number of time grid levels increases, similar to Section 5.1.1, but still provide reasonable estimates on convergence. Indeed, for V-cycles with F-relaxation, Approximation 1 is quite sharp.
max \| r_{i+1} \|_2 / \| r_i \|_2 \quad (V-cycle)

\| e_{F_{FCF}}^{n_{i+\Delta}} \|_2

\sqrt{\| e_{F_{FCF}}^{n_i,\Delta} \|_1 \| e_{F_{FCF}}^{n_i,\Delta} \|_\infty}

Approximation 2

max \| r_{i+1} \|_2 / \| r_i \|_2 \quad (F-cycle)

\| e_{F_{FCF}}^{n_i,\Delta} \|_2

\sqrt{\| e_{F_{FCF}}^{n_i,\Delta} \|_1 \| e_{F_{FCF}}^{n_i,\Delta} \|_\infty}

\[ \frac{\| e_{F_{FCF}}^{n_{i+\Delta}} \|_2}{\| e_{F_{FCF}}^{n_i,\Delta} \|_2} \]

\[ \sqrt{\| e_{F_{FCF}}^{n_i,\Delta} \|_1 \| e_{F_{FCF}}^{n_i,\Delta} \|_\infty} \]

Fig. 5: Isotropic diffusion: Comparison of V- and F-cycle MGRIT with FCF-relaxation. Convergence of A-stable and L-stable schemes deteriorates only slightly for an MGRIT V-cycle algorithm. On the other hand, the convergence factor for an F-cycle MGRIT algorithm is constant for all considered RK schemes and cases.

for all tested number of levels.

5.2. Wave equation. Consider the wave equation in two spatial dimensions over domain \( \Omega = (0, 2\pi) \times (0, 2\pi) \),

\[ \partial_{tt} u = c^2 \nabla \cdot \nabla u \quad \text{for} \ x \in \Omega, t \in (0, 2\pi), \]

with scalar solution \( u(x, t) \) and wave speed \( c = \sqrt{10} \). We transform Equation (5.1) into a system of PDEs that are first-order in time,

\[ \partial_t u = v, \quad \partial_t v = c^2 \nabla \cdot \nabla u, \quad \text{for} \ x \in \Omega, t \in (0, 2\pi), \]

with initial condition (see Figure 3) and boundary conditions,

\[ u(\cdot, 0) = \sin(x) \sin(y), \quad v(\cdot, 0) = 0, \quad \text{for} \ x \in \Omega \cup \partial\Omega, \]

\[ u(x, \cdot) = v(x, \cdot) = 0, \quad \text{for} \ x \in \partial\Omega. \]

This problem corresponds to a 2D membrane with imposed non-zero initial displacement \( u \) and zero initial velocity \( v \). The membrane enters an oscillatory motion
pattern due to initial stresses in the material. Thus, it is a simplified representative of a hyperbolic model that shares characteristic behavior with PDEs in solid dynamics research, such as linear elasticity [23]. Similar to Section 5.1, we use second-order centered finite differences to discretize the spatial operator in Equation (5.2). The time stepping operators $\Phi_\ell$ are then simultaneously diagonalizable and the Courant number on each level is given by $C_\ell = 2c\pi/[\delta x(N_\ell - 1)]$, ranging between $C_0 \approx 0.034$ on level 0 and $C_5 \approx 1.087$ on level 5.

An MGRIT V-cycle algorithm with FCF-relaxation shows quickly increasing convergence factors with a growing number of time grid levels (Figure 7). The worst-case convergence factors quickly exceed 1, and thus diverge, which is correctly predicted by all upper bounds and Approximation 2. Similarly, using an F-cycle results in a less dramatic, but still significant increase in observed and predicted convergence factors with respect to the number of levels. For some schemes, particularly L-stable ones, an F-cycle is able to retain convergence up to the six levels in time considered here, but the bounds and approximations developed here do not predict these results.

In general, the upper bounds on convergence applied to the wave equation are significantly less sharp compared to the diffusion equation (see Section 5.1), but they are still able to accurately represent correct trends. For example, convergence factors are initially constant in most cases, then increase almost linearly with the number of levels, such as the case of L-stable SDIRK 3 in Figure 8. This highlights the fact that designing robust and convergent parallel-in-time algorithms for hyperbolic problems is generally perceived as difficult, and emphasizes the benefit of the presented upper bounds for F-cycle algorithms. For example, the convergence factor can be estimated a priori to select a time grid hierarchy that is likely to yield a significant speedup. In combination with performance modeling [15], such a priori estimates can provide valuable guidance.

It is noted that in the investigated cases, a similarly strong benefit of FCF-relaxation over F-relaxation as for the diffusion equation cannot be observed for the wave equation, see Figure 9 and Figure 10 in the Appendix. However, in some cases FCF-relaxation increases the maximum number of time grid levels for which convergence can be achieved. Thus, in practice one would prefer F-relaxation over FCF-relaxation to reduce the computational cost of a given algorithm. The fact that
Fig. 7: Wave equation: The convergence factor of MGRIT with V-cycles and FCF-relaxation increases substantially with a growing number of time grid levels and eventually exceeds 1. This means that in the worst case, MGRIT V-cycles yields a divergent algorithm, which is in line with observations for hyperbolic PDEs in the literature [11, 12, 23].

FCF-relaxation is not sufficient to design a scalable multilevel solver for the wave equation is a major difference to observations for the diffusion equation.

We further note, that the observed convergence factors and upper bound values are smaller with higher time integration order, especially when L-stable SDIRK schemes are employed. For example, the theory suggests to use five-level MGRIT with F-cycles and L-stable SDIRK4 with an estimated upper bound on the convergence factor of $O(10^{-3})$, which is a very fast algorithm for hyperbolic PDEs.

6. Conclusion. In this work, we develop a framework for multilevel convergence of MGRIT. This framework provides a priori bounds and approximations for the convergence factor of various types of MGRIT configurations, including different cycling strategies (V- and F-cycles) and relaxation schemes (rFCF-relaxation). The new theoretical results are a generalization of the two-grid theory derived in [5] and based on similar assumptions (for example, simultaneously diagonalizable and stable time-stepping operators). This work also presents a generalization of the two-grid bounds derived in [42] to be exact for every iteration, as opposed to all but one.

In complementary numerical studies, the theoretical results in Section 3 are assessed for two different model problems, the diffusion equation (including new analysis for the anisotropic case) and the second-order wave equation. It was found that the a priori upper bounds are relatively sharp upper bounds on observed convergence for the diffusion equation, and accurately describe qualitative behavior for the wave equation. Generally, these bounds are sharper for smaller numbers of time grids. The bounds also appeared sharper for lower integration orders (e.g., see Figure 4). This, however, is most likely due to the relatively small convergence factors for higher-order schemes and, thus, the small number of MGRIT iterations required to satisfy the
Fig. 8: Wave equation: The convergence of MGRIT with F-cycles and FCF-relaxation deteriorates with a larger number of time grid levels compared to MGRIT with V-cycles, see Figure 7. Generally, convergent algorithms are given for a larger range of time grid levels and observed convergence is better than the predictions from the upper bounds. This shows that the choice of F-cycles over V-cycles is one likely ingredient for future improvements of MGRIT for hyperbolic-type PDEs.

The numerical studies have demonstrated the benefit of using FCF-relaxation for parabolic model problems, see Figure 5. On the other hand, FCF-relaxation did not significantly affect convergence of MGRIT for the hyperbolic model problem. The theoretical results further confirm the advantage of using F-cycles over V-cycles for the hyperbolic model problem. The same is true for the isotropic diffusion equation, if integrated by L-stable Runge-Kutta schemes. The observation that A-stable schemes are generally less suited for parallel-in-time integration [5] was confirmed in this work. For the diffusion model problem, however, it was shown that using MGRIT with F-cycles and FCF-relaxation can alleviate this weakness. Further, the theory implies that naive multilevel Parareal (i.e., MGRIT V-cycles with F-relaxation) does not yield a scalable algorithm and has increasing iteration counts with an increasing number of levels.

Overall, the theoretical convergence results lay the groundwork for future in-depth examination and understanding of MGRIT. This is especially true for the solution of convergence criterion. Therefore, the maximum observed convergence factors might not have reached the worst-case.

The presented upper bounds vary in their respective time complexity and it was demonstrated that, e.g., using the inequality (4.3) reduces the time complexity but still gives good and reasonably sharp upper bounds. Further, the proposed approximate convergence factors for V-cycle algorithms with F- and FCF-relaxation (Approximation 1 and 2) provide analytic formulae to estimate observed convergence a priori with effectively constant time complexity (if implemented in parallel). In the investigated cases, the approximate convergence factors yielded a priori estimates that were at least as good as more expensive bounds.

The numerical studies have demonstrated the benefit of using FCF-relaxation for parabolic model problems, see Figure 5. On the other hand, FCF-relaxation did not significantly affect convergence of MGRIT for the hyperbolic model problem. The theoretical results further confirm the advantage of using F-cycles over V-cycles for the hyperbolic model problem. The same is true for the isotropic diffusion equation, if integrated by L-stable Runge-Kutta schemes. The observation that A-stable schemes are generally less suited for parallel-in-time integration [5] was confirmed in this work. For the diffusion model problem, however, it was shown that using MGRIT with F-cycles and FCF-relaxation can alleviate this weakness. Further, the theory implies that naive multilevel Parareal (i.e., MGRIT V-cycles with F-relaxation) does not yield a scalable algorithm and has increasing iteration counts with an increasing number of levels.

Overall, the theoretical convergence results lay the groundwork for future in-depth examination and understanding of MGRIT. This is especially true for the solution of
hyperbolic PDEs; an application area where the design of robust and efficient parallel-in-time algorithms has proven challenging. Thus, future development and improvement of MGRIT can be guided by the proposed multilevel convergence framework, where already important recommendations, such as the usefulness of higher-order L-stable SDIRK methods, along with F-relaxation and F-cycles, for hyperbolic problems, are available. The theory also lays the groundwork to explore MGRIT variants, such as coarsening in integration order (as opposed to step size; p-MGRIT), weighted relaxation schemes, and others. Another interesting aspect is the use of dissipative spatial discretizations for hyperbolic PDEs, which was demonstrated to be beneficial in [38]. Investigating this aspect within the presented multilevel convergence framework could provide invaluable insight into how these observations are founded in the theory and suggest new ways to accelerate MGRIT performance for hyperbolic PDEs.

Future work will explore these ideas. We will further employ the theoretical results to guide parameter choices for MGRIT in various application areas and focus on extending the multilevel convergence framework. For example, it is desirable to also derive analytic formulae for F-cycle convergence and reduce the time complexity of the respective upper bounds similar to the V-cycle bounds, if possible. The relation between two-grid and multilevel convergence is a further interesting research topic. Additionally, the framework could be combined with performance models [15], to enable a priori estimates of the expected wall clock time of a given algorithm and its potential speedup.

Lastly, we provide a parallel C++ implementation of all derived bounds and approximate convergence factors that is made publicly available. The code takes the (complex or real) eigenvalues of \( \{\Phi_\ell\} \) as input along with a definition of the desired MGRIT algorithm (cycling, relaxation, coarsening, etc.) and computes the bound or approximate convergence factor values. We further implemented functionality for the user to supply the eigenvalues of a spatial operator and to compute the respective eigenvalues of \( \{\Phi_\ell\} \) based on the stability function of a given Runge-Kutta time integration scheme and its Butcher tableau.

Acknowledgements. AH would like to acknowledge the Lawrence Livermore National Laboratory, USA, and RDF/JBS for funding and hosting a summer internship during which most of the theoretical results of this manuscript were derived.

REFERENCES

[1] L. Bonaventura and A. Della Rocca, Monotonicity, positivity and strong stability of the tr-bdf2 method and of its ssp extensions, arXiv preprint arXiv:1510.04303, (2015).
[2] J. C. Butcher, Numerical methods for ordinary differential equations, John Wiley & Sons, 2016.
[3] A. J. Christlieb, C. B. Macdonald, and B. W. Ong, Parallel high-order integrators, SIAM Journal on Scientific Computing, 32 (2010), pp. 818–835.
[4] J. Cortial and C. Farhat, A time-parallel implicit method for accelerating the solution of non-linear structural dynamics problems, International Journal for Numerical Methods in Engineering, 77 (2009), pp. 451–470.
[5] V. A. Dobrev, T. V. Kolev, N. A. Petersson, and J. B. Schroder, Two-level convergence theory for multigrid reduction in time (mgrit), SIAM Journal on Scientific Computing, 39 (2017), pp. S501–S527.
[6] M. Duarte, R. Dobbins, and M. Smooke, High order implicit time integration schemes on multiresolution adaptive grids for stiff pdes, arXiv preprint arXiv:1604.00355, (2016).
[7] M. Emmett and M. L. Minion, Toward an Efficient Parallel in Time Method for Partial Differential Equations, Communications in Applied Mathematics and Computational Science, 7 (2012), pp. 105–132.
[8] R. D. Falgout, S. Friedhoff, T. V. Kolev, S. P. MacLachlan, and J. B. Schroder,
Parallel time integration with multigrid, SIAM Journal on Scientific Computing, 36 (2014), pp. C635–C661.

[9] R. D. Falgout, M. Lecouvez, and C. S. Woodward, A parallel-in-time algorithm for variable step multistep methods, 2017.

[10] R. D. Falgout, T. A. Manteuffel, B. O’Neill, and J. B. Schroder, Multigrid reduction in time for nonlinear parabolic problems: A case study, SIAM Journal on Scientific Computing, 39 (2017), pp. S208–S322.

[11] C. Farhat and M. Chandreskhar, Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications, International Journal for Numerical Methods in Engineering, 58 (2003), pp. 1397–1434.

[12] C. Farhat, J. Cortial, C. Dastillung, and H. Bavestrello, Time-parallel implicit integrators for the near-real-time prediction of linear structural dynamic responses, International Journal for Numerical Methods in Engineering, 67 (2006), pp. 697–724.

[13] S. Friedhoff, R. D. Falgout, T. V. Kolev, S. MacLachlan, and J. B. Schroder, A multigrid-in-time algorithm for solving evolution equations in parallel, tech. rep., Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 2012.

[14] S. Friedhoff and S. P. MacLachlan, A generalized predictive analysis tool for multigrid methods, Numerical Linear Algebra with Applications, 22 (2015), pp. 618–647.

[15] H. Garavani, V. A. Dobrev, R. D. Falgout, T. V. Kolev, J. B. Schroder, M. Schulz, and U. M. Yang, A performance model for allocating the parallelism in a multigrid-in-time solver, in Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), International Workshop on, IEEE, 2016, pp. 22–31.

[16] M. J. Gander, 50 years of Time Parallel Time Integration, in Multiple Shooting and Time Domain Decomposition, Springer, 2015.

[17] M. J. Gander, F. Kwok, and H. Zhang, Multigrid interpretations of the parareal algorithm leading to an overlapping variant and mgrit, Computing and Visualization in Science, 19 (2018), pp. 59–74.

[18] S. Günther, N. R. Gauger, and J. B. Schroder, A non-intrusive parallel-in-time adjoint solver with the xbraid library, Computing and Visualization in Science, 19 (2018), pp. 85–95.

[19] ________, A non-intrusive parallel-in-time approach for simultaneous optimization with unsteady pdes, Optimization Methods and Software, (2018), pp. 1–16.

[20] W. Hackbusch and U. Trottenberg, Multigrid methods: proceedings of the conference held at Köln-Porz, November 23-27, 1981, vol. 960, Springer, 2006.

[21] E. Hairer, S. P. Norsett, and G. Wanner, Solving ordinary differential equations i: Nonstiff problems, 1993.

[22] E. Hairer and G. Wanner, Solving ordinary differential equations ii: Stiff and differential-algebraic problems, Springer Series in Computational Mathematics, 14 (1996).

[23] A. Hessenthaler, D. Nordsetten, O. Röhrle, J. B. Schroder, and R. D. Falgout, Convergence of the multigrid reduction in time algorithm for the linear elasticity equations, Numerical Linear Algebra with Applications, 25 (2018), p. e2155.

[24] N. J. Higham, Estimating the matrixp-norm, Numerische Mathematik, 62 (1992), pp. 539–555.

[25] G. Horton, F. S. Vandewalle, and P. Worley, An Algorithm with Polylog Parallel Complexity for Solving Parabolic Partial Differential Equations, SIAM Journal on Scientific Computing, 16 (1995), pp. 848–864.

[26] G. Horton, F. S. Vandewalle, and P. Worley, An Algorithm with Polylog Parallel Complexity for Solving Parabolic Partial Differential Equations, SIAM Journal on Scientific Computing, 16 (1995), pp. 531–541.

[27] A. Howse, H. De Sterck, S. MacLachlan, R. Falgout, and J. Schroder, Multigrid reduction in time with adaptive spatial coarsening for the linear advection equation, tech. rep., Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2017.

[28] M. Lecouvez, R. D. Falgout, C. S. Woodward, and P. Top, A parallel multigrid reduction in time method for power systems, in Power and Energy Society General Meeting (PESGM), 2016, IEEE, 2016, pp. 1–5.

[29] J.-L. Lions, Y. Maday, and G. Turinici, A "parareal" in time discretization of PDE’s, Comptes Rendus de l’Académie des Sciences - Series I - Mathematics, 332 (2001), pp. 661–668.

[30] C. Luch and A. Ostermann, Multi-grid dynamic iteration for parabolic equations, BIT Numerical Mathematics, 27 (1987), pp. 216–234.

[31] T. A. Manteuffel, L. N. Olson, J. B. Schroder, and B. S. Southworth, A root-node based algebraic multigrid method, SIAM Journal on Scientific Computing, (accepted) (2017).

[32] M. Miranda, P. Tilli, and 2000, Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results, SIAM Journal on Matrix Analysis and Applications, 21 (2000),
Appendix A. Proofs of Section 2.

Proof of Theorem 2.3. Here we consider a slightly modified coarsening of points: let the first \( k \) points be F-points, followed by a C-point, followed by \( k \) F-points, and so on, finishing with a C-point (as opposed to starting with a C-point). This is simply a theoretical tool that permits a cleaner analysis, but is not fundamental to the result. Then,

\[
R_{\text{ideal}} = \begin{bmatrix} -A_{cf}(A_{ff})^{-1} & I \\ \Phi_0 & \Phi_0 & \ldots & \Phi_0 \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_0 & \Phi_0 & \ldots & \Phi_0 \\ I \\ \end{bmatrix}
\]

Let \( \tilde{P} \) be the orthogonal (column) block permutation matrix such that

\[
R_{\text{ideal}} \tilde{P} = \begin{bmatrix} 
\Phi_0 & \ldots & \Phi_0 & I \\
\vdots & & \ddots & \vdots \\
\Phi_0 & \ldots & \Phi_0 & I \\
\end{bmatrix} := \begin{bmatrix} 
W \\
\vdots \\
W \\
\end{bmatrix}
\]

where \( W \) is the block row matrix \( W = (\Phi_0, \ldots, \Phi_0, I) \). Then, the \( \ell^1 \)-norm of residual
propagation of MGRIT with F-relaxation is

\[ \| \mathcal{E}_F \|_{A^*A} = \| \mathcal{R}_F \| = \| (I - A \Delta B_{\Delta}^{-1}) R_{\text{ideal}} \| = \| (I - A \Delta B_{\Delta}^{-1}) R_{\text{ideal}} \tilde{P} \|, \]

where

\[
(I - A \Delta B_{\Delta}^{-1}) R_{\text{ideal}} \tilde{P} = \text{diag} \left[ \Phi_1 - \Phi_0^k \right] \begin{bmatrix} 0 & I & 0 & 0 \\ \Phi_1 & I & 0 & 0 \\ \Phi_1^2 & \Phi_1 & I & 0 \\ \vdots & \vdots & \vdots & \ddots \\ \Phi_1^{N_k-2} & \Phi_1^{N_k-3} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \vdots \\ \mathbf{W} \end{bmatrix}
\]

Excuse the slight abuse of notation and denote \( \mathcal{R}_F := (I - A \Delta B_{\Delta}^{-1}) R_{\text{ideal}} \tilde{P} \); that is, ignore the upper zero blocks in \( \mathcal{R}_F \). Define a tentative pseudoinverse, \( \mathcal{R}_F^\dagger \), as

\[
\mathcal{R}_F^\dagger = \begin{bmatrix} 0 & \bar{W}^{-1}(\Phi_1 - \Phi_0^k)^{-1} \\ -W^{-1} \Phi_1 (\Phi_1 - \Phi_0^k)^{-1} & \bar{W}^{-1}(\Phi_1 - \Phi_0^k)^{-1} \\ \vdots & \vdots \\ -W^{-1} \Phi_1 (\Phi_1 - \Phi_0^k)^{-1} & \bar{W}^{-1}(\Phi_1 - \Phi_0^k)^{-1} \\ 0 & 0 \end{bmatrix},
\]

for some \( \bar{W}^{-1} \), and observe that

\[
\mathcal{R}_F^\dagger \mathcal{R}_F = \begin{bmatrix} \bar{W}^{-1} \mathbf{W} \\ \vdots \\ \bar{W}^{-1} \mathbf{W} \\ 0 \end{bmatrix}.
\]

Three of the four properties of a pseudoinverse require that,

\[
\mathcal{R}_F^\dagger \mathcal{R}_F \mathcal{R}_F^\dagger = \mathcal{R}_F^\dagger, \quad \mathcal{R}_F \mathcal{R}_F^\dagger \mathcal{R}_F = \mathcal{R}_F \quad \text{and} \quad \left( \mathcal{R}_F^\dagger \mathcal{R}_F \right)^* = \mathcal{R}_F^\dagger \mathcal{R}_F.
\]

These three properties follow by picking \( \bar{W}^{-1} \) such that \( \left( \bar{W}^{-1} \mathbf{W} \right)^* = \bar{W}^{-1} \mathbf{W} \), and \( \mathbf{W} \bar{W}^{-1} \mathbf{W} = \mathbf{W}, \quad \bar{W}^{-1} \mathbf{W} \bar{W}^{-1} = \bar{W}^{-1} \). Notice these are exactly the first three properties of a pseudoinverse of \( \mathbf{W} \). To that end, define \( \bar{W}^{-1} \) as the pseudoinverse of a full row rank operator,

\[
\bar{W}^{-1} = \mathbf{W}^* (\mathbf{W} \mathbf{W}^*)^{-1}.
\]

Note that here, \( \mathbf{W} \bar{W}^{-1} = I \), and the fourth property of a pseudoinverse for \( \mathcal{R}_F \),

\[
\left( \mathcal{R}_F \mathcal{R}_F^\dagger \right)^* = \mathcal{R}_F \mathcal{R}_F^\dagger,
\]

follows immediately.
Recall the maximum singular value of $\mathcal{R}_F$ is given by the minimum nonzero singular value of $\mathcal{R}_F^\dagger$, which is equivalent to the minimum nonzero singular value of $(\mathcal{R}_F^\dagger)^*\mathcal{R}_F^\dagger$. Following from [42], the minimum nonzero eigenvalue of $(\mathcal{R}_F^\dagger)^*\mathcal{R}_F^\dagger$ is bounded from above by the minimum eigenvalue of a block Toeplitz matrix, with real-valued matrix generating function

$$F(x) = (e^{ix} \tilde{W}^{-1} \Phi_1(\Phi_1 - \Phi_0^k) - \tilde{W}^{-1}(\Phi_1 - \Phi_0^k)) \left( e^{ix} \tilde{W}^{-1} \Phi_1(\Phi_1 - \Phi_0^k) - \tilde{W}^{-1}(\Phi_1 - \Phi_0^k) \right)^*.$$ 

Let $\lambda_k(A)$ and $\sigma_k(A)$ denote the $k$th eigenvalue and singular value of some operator, $A$. Then,

$$\min_{x \in [0,2\pi]} \lambda_k(F(x)) = \min_{x \in [0,2\pi]} \sigma_k \left( e^{ix} \tilde{W}^{-1} \Phi_1(\Phi_1 - \Phi_0^k) - \tilde{W}^{-1}(\Phi_1 - \Phi_0^k) \right)^2$$

$$= \min_{x \in [0,2\pi], \nu \neq 0} \frac{\| (e^{ix} \tilde{W}^{-1} \Phi_1(\Phi_1 - \Phi_0^k) - \tilde{W}^{-1}(\Phi_1 - \Phi_0^k)) \nu \|^2}{\|\nu\|^2}$$

$$= \min_{x \in [0,2\pi], \nu \neq 0} \frac{\| \tilde{W}^{-1}(e^{ix} \Phi_1 - I)\nu \|^2}{\| (\Phi_1 - \Phi_0^k)\nu \|^2}$$

$$\sigma_{\min} \left( \mathcal{R}_F^\dagger \right) \leq \min_{x \in [0,2\pi], \nu \neq 0} \frac{\| (WW^*)^{-1/2}(e^{ix} \Phi_1 - I)\nu \|^2}{\| (\Phi_1 - \Phi_0^k)\nu \|^2} + O(1/N_1).$$

Then,

$$\|\mathcal{E}_F\|_{A^*A} = \|\mathcal{R}_F\| \geq \frac{1}{\sqrt{\min_{x \in [0,2\pi], \nu \neq 0} \| (WW^*)^{-1/2}(e^{ix} \Phi_1 - I)\nu \|^2} + O(1/N_1)}$$

$$\geq \frac{1}{\min_{x \in [0,2\pi], \nu \neq 0} \| (WW^*)^{-1/2}(e^{ix} \Phi_1 - I)\nu \|^2} + O(1/\sqrt{N_1})$$

$$= \max_{\nu \neq 0} \min_{x \in [0,2\pi]} \| (WW^*)^{-1/2}(e^{ix} \Phi_1 - I)\nu \|^2 + O(1/\sqrt{N_1}),$$

and the result for $\mathcal{E}_F$ follows.\(^{12}\)

The case of $\mathcal{E}_{FCF}$ follows an identical derivation with the modified operator $\tilde{W} = (\Phi_0^k, ..., \Phi_0^k)$, which follows from the right multiplication by $A_{cf}A_{jf}^{-1}A_{fc}$ in $\mathcal{R}_{FCF} = (I - A_{cf}B_{A_{fc}}^{-1})A_{cf}A_{jf}^*A_{fc}R_{\text{ideal}}$, which is effectively a right-diagonal scaling by $\Phi_0^k$. The cases of error propagation in the $\ell^2$-norm follow a similar derivation based on $P_{\text{ideal}}$.\(\square\)

**Proof of Corollary 2.4.** The derivations follow a similar path as those in Theorem 2.3. However, when considering Toeplitz operators defined over the complex scalars

\(^{12}\)More detailed steps for this proof involving the block Toeplitz matrix and generating function can be found in similar derivations in [42, 40, 41, 32].
which is equivalent with (3.3) for \( r \) the steps follow easily when coupled with the pseduoinverse derived in Theorem 2.3. Derivations for each of these steps can be found in the diagonalizable case in [42], and function \( F \) of a second-order zero of \( F(x) - \min_{y \in [0,2\pi]} F(y) \), when the Toeplitz generating function \( F(x) \) is defined over complex scalars as opposed to operators [40]. Analogous derivations for each of these steps can be found in the diagonalizable case in [42], and the steps follow easily when coupled with the pseduoinverse derived in Theorem 2.3.

Then, noting that

\[
W_F = \sqrt{\sum_{\ell=0}^{m_0-1} (|\lambda_{0,i}|^2)^\ell} = \sqrt{\frac{1-|\lambda_{0,i}|^{2m_0}}{1-|\lambda_{0,i}|^2}},
\]

\[
W_{FCF} = \sqrt{\sum_{\ell=m_0}^{2m_0-1} (|\lambda_{0,i}|^2)^\ell} = |\lambda_{0,i}|^{m_0} \sqrt{\frac{1-|\lambda_{0,i}|^{2m_0}}{1-|\lambda_{0,i}|^2}},
\]

and substituting \( \lambda_{0,i} \) for \( \Phi_0 \) and \( \lambda_{1,i} \) for \( \Phi_1 \) in Theorem 2.3, the result follows.

**Appendix B. Proofs of Section 3.**

**Proof of Lemma 3.2.** For \( n_\ell = 2 \), we have,

\[
\xi_{F_\ell}^{n_\ell=2} = P_0R_{I_0} - P_0A_1^{-1}R_0A_0P_0R_{I_0} = (I - P_0A_1^{-1}R_0A_0)P_0R_{I_0},
\]

which is equivalent with (3.3) for \( r = 0 \). Now, assume it is true for \( n_\ell = n \) levels. Substituting an exact two-level method on the coarse grid, that is (3.4), yields,

\[
\xi_{F_\ell}^{n_\ell=n} = P_0R_{I_0} - \left( \prod_{k=0}^{n-2} P_k \right) \left[ P_{n-1}(R_{n-1}A_{n-1}P_{n-1})^{-1}R_{n-1} + S_{n-1}(S_{n-1}^T A_{n-1}S_{n-1})^{-1}S_{n-1}^T \right] \left( \prod_{k=n-2}^0 R_k \right) A_0P_0R_{I_0}
\]

\[
- \sum_{i=0}^{n-3} \left( \prod_{k=0}^i P_k \right) S_{i+1}(S_{i+1}^T A_{i+1}S_{i+1})^{-1}S_{i+1}^T \left( \prod_{k=i}^0 R_k \right) A_0P_0R_{I_0}
\]

\[
= P_0R_{I_0} - \left( \prod_{k=0}^{n-1} P_k \right) (R_{n-1}A_{n-1}P_{n-1})^{-1} \left( \prod_{k=n-1}^0 R_k \right) A_0P_0R_{I_0}
\]

\[
- \sum_{i=0}^{n-2} \left( \prod_{k=0}^i P_k \right) S_{i+1}(S_{i+1}^T A_{i+1}S_{i+1})^{-1}S_{i+1}^T \left( \prod_{k=i}^0 R_k \right) A_0P_0R_{I_0}.
\]

Approximating the exact coarse grid operator on level \( n + 1 \) by \( A_n \approx R_{n-1}A_{n-1}P_{n-1} \) concludes the proof.

**Appendix C. Derivations: Bounds for MGRIT residual and error propagation.**

Here, we present derivations and proofs that were omitted in Section 4.

**C.1. Residual and error on level 0 and level 1.**
Proof of Lemma 4.1. This follows from,

\[ \|E_{rFCF}^{n_\ell}\|_2 = \|P_0E_{rFCF}^{n_\ell}\|_2 \|E_{rFCF}^{n_\ell}\|_2 \leq \sqrt{\|P_0\|_1 \|P_0\|_\infty} \|E_{rFCF}^{n_\ell}\|_2 \|E_{rFCF}^{n_\ell}\|_\infty \leq \sqrt{m_0} \|E_{rFCF}^{n_\ell}\|_2, \]

where we have used submultiplicativity and the inequality \(\|D\|_2 \leq \sqrt{\|D\|_1 \|D\|_\infty}\) (see [24]). \(\square\)

C.2. Two-level MGRIT with \(rFCF\)-relaxation. The coarse-grid error propagator follows from Equation (3.3) for \(n_\ell = 2\),

\[ E_{rFCF}^{n_\ell=2,\Delta} = R_{I_0}E_{rFCF}^{n_\ell=2}P_0 = (I - A_1^{-1}R_0A_0P_0)(I - R_0A_0P_0)^r \]

with,

(C.1) \[ R_0A_0P_0 = \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Phi_0^{m_0} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \cdots & 1 \end{bmatrix}, \]

(C.2) \[ I - A_1^{-1}R_0A_0P_0 = -\begin{bmatrix} 0 & \Phi_1 - \Phi_0^{m_0} & 0 & \cdots & 0 \\ \Phi_1 (\Phi_1 - \Phi_0^{m_0}) & \Phi_1 - \Phi_0^{m_0} & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots \\ \Phi_1^{N_1-2} (\Phi_1 - \Phi_0^{m_0}) & \cdots & 0 \end{bmatrix}, \]

and

\[ (I - R_0A_0P_0)^r = \begin{bmatrix} 0 & \Phi_0^{m_0} & 0 & \cdots & 0 \\ \Phi_0^{m_0} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & \cdots & 0 & \Phi_0^{m_0} \end{bmatrix} (I - R_0A_0P_0)^{r-1} \]

\[ = \begin{bmatrix} 0 & \Phi_0^{m_0} & 0 & \cdots & 0 \\ \Phi_0^{m_0} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & \cdots & 0 & \Phi_0^{m_0} \end{bmatrix} (I - R_0A_0P_0)^{r-2} \]

\[ = \cdots = \begin{bmatrix} 0 & \cdots & 0 & \Phi_0^{m_0} \\ \vdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \]

Proof of Theorem 4.3. This follows from Equation (4.4) and inequality (4.3),

\[ \|E_{rFCF}^{n_\ell=2,\Delta}\|_2 \leq \sqrt{\|E_{rFCF}^{n_\ell=2,\Delta}\|_1 \|E_{rFCF}^{n_\ell=2,\Delta}\|_\infty} = \|E_{rFCF}^{n_\ell=2}\|_1 \]

34
In the last step, the relationship $\sum_{i=0}^{N-1} a^i = (1 - a^N)/(1 - a)$ was used.

**C.3. Three-level V-cycles with F-relaxation.** Evaluating the error propagator in Equation (3.5) for a three-level V-cycle with F-relaxation on the coarse-grid yields,

$$E_{F}^{\delta=3,\Delta} = R_0 E_{F}^{\delta=3} P_0 = I - [P_1 A_2^{-1} R_1 + S_1 (S_1^T A_1 S_1)^{-1} S_1^T] R_0 A_0 P_0,$$

where,

$$S_1 (S_1^T A_1 S_1)^{-1} S_1^T = \begin{bmatrix} I & & & \\ \Phi_1 & I & & \\ \Phi_1^2 & \Phi_1 & I & \\ \vdots & \ddots & \ddots & \ddots \\ \Phi_{m-2} & \Phi_{m-3} & \cdots & \Phi_1 & I \\ 0 & & & & \\ \Phi_1 & I & & \\ \Phi_1^2 & \Phi_1 & I & \\ \vdots & \ddots & \ddots & \ddots \\ \Phi_{m-2} & \Phi_{m-3} & \cdots & \Phi_1 & I \\ 0 & & & & \end{bmatrix}$$

and

$$P_1 A_2^{-1} R_1 = \begin{bmatrix} I & & & & & & & & & & & & & & & \\ \Phi_1 & I & & & & & & & & & & & & & & \\ \Phi_1^2 & \Phi_1 & I & & & & & & & & & & & & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \Phi_{m-1} & \Phi_{m-2} & \Phi_{m-3} & \cdots & \Phi_1 & I & & & & & & & & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \Phi_{m-1} & \Phi_{m-2} & \Phi_{m-3} & \cdots & \Phi_1 & I & & & & & & & & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \Phi_{m-1} & \Phi_{m-2} & \Phi_{m-3} & \cdots & \Phi_1 & I & & & & & & & & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$
With (C.1), (C.3) and (C.4) follows the nilpotent operator, 

\begin{equation}
\mathcal{E}_{p}^{\Delta} = \begin{bmatrix}
0 & 0 & 0 \\
\Phi_m - \Phi_0 & 0 & 0 \\
\Phi_{m-1} - \Phi_0 & \Phi_{m-1} - \Phi_1 & 0 \\
\vdots & \vdots & \vdots \\
\Phi_{m-n} - \Phi_0 & \Phi_{m-n-1} - \Phi_1 & \Phi_{m-n} - \Phi_1 \\
\end{bmatrix}
\end{equation}

**Proof of Theorem 4.9.** The proof is analogous to Theorem 4.3.

### C.3.1. Four-level V-cycles with F-relaxation

Evaluating the error propagator in Equation (3.5) for a four-level V-cycle with F-relaxation on the coarse-grid yields,

\[
\mathcal{E}_{P}^{n+1} = R_{0} \mathcal{E}_{P}^{n-1} \mathcal{E}_{0}
\]

\[
= I - \left[ P_{1}P_{2}A_{3}^{-1}R_{2}R_{1} + S_{1}(S_{1}^{T}A_{1}S_{1})^{-1}S_{1}^{T} + P_{1}S_{2}(S_{2}^{T}A_{2}S_{2})^{-1}S_{2}^{T}R_{1} \right] R_{0}A_{0}P_{0},
\]

where,
\[
\xi_i \quad (C.6) \quad P_1 S_2 (S_2^T A_2 S_2)^{-1} S_2^T R_1 = \begin{bmatrix}
0 \\
\vdots \\
0 \\
\Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \\
\Phi_2^{m_2-2} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-2} \Phi_1^{m_1-1} \\
\Phi_1^{m_1-1} \Phi_2 \Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \Phi_2 \Phi_1^{m_1-1} \\
\Phi_2 \Phi_1^{m_1-1} & \cdots & \Phi_2 \Phi_1^{m_1-1} \\
\vdots & \cdots & \vdots \\
\Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\vdots & \cdots & \vdots \\
\Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\vdots & \cdots & \vdots \\
\Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_1^{m_1-1} \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\Phi_2^{m_2-3} \Phi_1^{m_1-1} & \cdots & \Phi_2^{m_2-3} \Phi_1^{m_1-1} \\
\vdots & \cdots & \vdots \\
0 \\
0 \\
\end{bmatrix}
\]
and,

$$(C.7) \quad \Phi_{m} - \Phi_{N} \cdot \Phi_{m} + S$$

The sum of (C.3), (C.6) and (C.7) yields,

$$P_{1}P_{2}A_{3}^{-1}R_{2}R_{1} + S_{1}(S_{1}^{T}A_{1}S_{1})^{-1}S_{1}^{T} + P_{1}S_{2}(S_{2}^{T}A_{2}S_{2})^{-1}S_{2}^{T}R_{1}$$

$$(C.8) \quad =$$

where we notice that the sparsity patterns are non-overlapping. Then, the error propagator is given as,
(C.9) \[ \mathcal{L}^{n_F=1, \Delta} = \]
Proof of Theorem 4.10. The proof is analogous to Theorem 4.9.

C.3.2. Three-level V-cycles with FCF-relaxation.

Proof of Theorem 4.12. The proof is analogous to Theorem 4.9.

Appendix D. Butcher tableaux of SDIRK schemes.

\[
\begin{array}{c|ccccc}
1 & 1 & 1 - \gamma & 1 - \gamma & 0 & 0 \\
\gamma & 2\gamma - 1 & 1 - \gamma & 1/2 & 1/2 & q \\
s & s - q & q & 0 & 0 & 0 \\
r & 1 - q - r & q & r & 1 - q - r & q \\
\end{array}
\]

Table 1: Butcher tableau for L-stable SDIRK scheme of orders 1 - 3 with \( \gamma = 1/\sqrt{2} \), \( q = 0.4358665215 \ldots \), \( r = 1.2084966491 \ldots \) and \( s = 0.7179332607 \ldots \); See [5].

\[
\begin{array}{cccccccc}
1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\
3/4 & 1/2 & 1/4 & 0 & 0 & 0 & 0 & 0 \\
11/20 & 17/50 & -1/25 & 1/4 & 0 & 0 & 0 & 0 \\
1/2 & 371/1360 & -137/2720 & 15/544 & 1/4 & 0 & 0 & 0 \\
1 & 25/24 & -49/48 & 125/16 & -85/12 & 1/4 & 0 & 0 \\
\end{array}
\]

Table 2: Butcher tableau for L-stable SDIRK scheme of orders 4; See [6], Appendix C.

\[
\begin{array}{cccccccc}
1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\
3/4 & 1/2 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 \\
1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 \\
\end{array}
\]

Table 3: Butcher tableau for A-stable SDIRK scheme of orders 2 - 4 with \( \gamma = (3 + \sqrt{3})/6 \), \( q = \cos(\pi/18)/\sqrt{3} + 1/2 \) and \( r = 1/(6(2q - 1)^2) \); See [1].

Appendix E. Additional numerical results.
Fig. 9: Wave equation: Observed convergence and predicted upper bounds on convergence of MGRIT with V-cycles and F-relaxation shows very similar trends as for MGRIT with V-cycles and FCF-relaxation, see Figure 7. This shows that switching from F-relaxation to FCF-relaxation alone is not sufficient to yield a robust MGRIT algorithm for the wave equation (and likely, other hyperbolic PDEs).

Fig. 10: Wave equation: Observed convergence and predicted upper bounds on convergence of MGRIT with F-cycles and F-relaxation shows very similar trends as for MGRIT with F-cycles and FCF-relaxation, see Figure 8. This shows that switching from F-relaxation to FCF-relaxation alone is not sufficient to yield a robust MGRIT algorithm for the wave equation (and likely, other hyperbolic PDEs).