SIAM J. SCI. COMPUT.
Vol. 42, No. 2, pp. A771–A796

MULTILEVEL CONVERGENCE ANALYSIS OF MULTIGRID-REDUCTION-IN-TIME*

ANDREAS HESSENTHALER†, BEN S. SOUTHWORTH‡, DAVID NORDSLETTEN§, 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 the 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 (anisotropic diffusion) and hyperbolic (wave equation) 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 estimating MGRIT convergence *a priori* and for the design of a convergent 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 and the benefit of a stronger relaxation scheme. An important observation is that with increasing numbers of levels MGRIT convergence deteriorates for the hyperbolic model problem, while constant convergence factors can be achieved for the diffusion equation. The theory also indicates that L-stable Runge–Kutta schemes are more amendable to multilevel 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, analytic upper bounds, a priori estimates

AMS subject classifications. 15, 35, 65

DOI. 10.1137/19M1238812

1. Introduction.

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

*Submitted to the journal's Methods and Algorithms for Scientific Computing section January 14, 2019; accepted for publication (in revised form) November 27, 2019; published electronically March 19, 2020. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Copyright is owned by SIAM to the extent not limited by these rights.

https://doi.org/10.1137/19M1238812

Funding: This work was supported by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, LLNL-JRNL-763460.

†Institute for Modelling and Simulation of Biomechanical Systems, University of Stuttgart, 70565 Stuttgart, Germany (hessenthaler@mechbau.uni-stuttgart.de, roehrle@simtech.uni-stuttgart.de).

‡Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309 (ben.s.southworth@gmail.com).

§School of Biomedical Engineering and Imaging Sciences, King's College London, St. Thomas Hospital, London, SE1 7EH, UK, and Department of Biomedical Engineering and Cardiac Surgery, University of Michigan, NCRC B20, Ann Arbor, MI 48109 (david.nordsletten@kcl.ac.uk).

¶Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551 (falgout2@llnl.gov).

‖Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131 (jbschroder@unm.edu).

of computational models and accuracy of computed quantities.

Spatial domain decomposition (DD) methods are a widespread 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 [35, 13] increase the amount of parallelism that can be exploited by introducing parallelism in the temporal domain. Many such methods exist, including waveform relaxation [33, 45], space-time multigrid [25], parallel implicit time-integrator [8, 9], revisionist integral deferred correction [1], spectral deferred correction (SDC) [44, 4, 22], Parareal [32], and multigrid-reduction-in-time [10, 5]. These methods have been developed for various application areas and with varying degree of intrusiveness, ease of implementation, level of parallelism, and potential for speedup. For an extensive review, see [13].

In this paper, we focus on multigrid-reduction-in-time (MGRIT), a recently developed iterative, multilevel algorithm, which introduces parallelism in the temporal domain by employing a parallel, iterative coarse-grid correction in time based on multigrid reduction. MGRIT has been explored for various application areas, including the numerical solution of parabolic and hyperbolic partial differential equations (PDEs) [10, 7, 26, 23], investigations of power systems [30, 40], solution of adjoint and optimization problems [17, 18], and neural network training [39]. Two-level convergence theory for MGRIT was developed in [3] for time integration on a uniform time grid, under the assumption of linear, simultaneously diagonalizable time-stepping operators (see section 2.2). The derived *a priori* bounds were shown to be quite accurate when compared with convergence observed in practice. Southworth [41] 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 nonuniform time grids are also provided in [41, 47]. In a special two-level case, MGRIT and Parareal are equivalent, and within this setting, convergence theory was developed for the linear and nonlinear case [16, 14]. However, no work has been done on convergence theory for the general multilevel setting, which is often far superior in practice. The selection of an appropriate cycling strategy and relaxation scheme is fundamental to achieving scalable multilevel performance and, ultimately, for achieving parallel speedup. A theoretical framework for multilevel convergence of MGRIT can help guide these decisions in a rigorous, a priori manner.

This paper introduces a framework for multilevel convergence analysis of MGRIT, laying the groundwork for a better understanding of MGRIT in theory and in practice. While the convergence framework is developed for linear PDEs, we note that MGRIT employs full approximation storage (FAS) multigrid (similar to methods like multilevel SDC, e.g., [4, 22]) and is thus applicable to the general nonlinear case. It was shown in [3], however, that the investigation of linear PDEs can illustrate the strengths and weaknesses of two-level MGRIT. Similarly, the new multilevel convergence framework in this work highlights how the use of a stronger relaxation scheme and carefully selected time-integration schemes for the diffusion equation can benefit MGRIT convergence significantly. On the other hand, it highlights problematic areas for MGRIT, for example, if A-stable Runge–Kutta schemes are used for the parallel-in-time integration of the second-order wave equation. Thus, we expect the analysis presented here to help guide the development and improvement of MGRIT through ideas such as coarsening in integration order as opposed to step size (p-MGRIT; similar to coarsening in collocation order for multilevel SDC [43]). Furthermore, we provide a

parallel C++ implementation of all derived bounds and approximate convergence factors in the supplementary materials and as open-source software[1] to guide parameter choices for a particular application in an *a priori* manner.

The paper proceeds as follows. In section 2, notation for a general linear time-stepping problem is introduced, and in section 2.1, the MGRIT algorithm and operators are reviewed. The important assumption of simultaneously diagonalizable operators is discussed in section 2.2, and the connection between two-level and multilevel convergence is discussed in section 2.3. In sections 3 and 4, we first generalize previous two-level convergence theory of MGRIT to the case with an arbitrary number of relaxation steps, referred to as rFCF-relaxation. We then extend the convergence framework to the multilevel setting, presenting analytic formulae to approximate the worst-case convergence factor of MGRIT algorithms, for multiple MGRIT cycling strategies and types of relaxation. 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. Section 5 demonstrates the sharpness of the derived theoretical bounds for model parabolic (including new analysis of the anisotropic diffusion equation) and hyperbolic PDEs, highlighting the benefits of the theory derived in this work.

2. Multigrid-reduction-in-time (MGRIT). For linear problems, sequential time stepping based on a single-step integration operator Φ can be written as

$$\mathbf{u}_n = \Phi_n \mathbf{u}_{n-1} + \mathbf{g}_n \qquad \text{for } n = 1, \ldots, N_t - 1, \tag{2.1}$$

with state vector $\mathbf{u}_n \in \mathbb{R}^{N_x}$ at time $t_n \in (0, T]$, initial condition $\mathbf{u}_0$ at time $t_0 = 0$, and forcing function $\mathbf{g}_n$. 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 one-step integrator, $\Phi_n = \Phi$, for all n.[3]

In matrix form, (2.1) can be written as

$$A\mathbf{u} = \begin{bmatrix} I & & & \\ -\Phi & I & & \\ & -\Phi & I & \\ & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \end{bmatrix} = \mathbf{g}, \tag{2.2}$$

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

MGRIT [10, 5] solves (2.2) iteratively and, like sequential time stepping, is an $O(N_t)$ method, that is, the total number of (block) operations to solution is linear or near-linear with the number of time steps (assuming MGRIT is applicable/convergent). MGRIT introduces a multilevel hierarchy of n_ℓ time grids of varying step size to achieve parallelism in the temporal domain, employing a coarse-grid correction 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 $\ell = 0, \ldots, n_\ell - 2$),[4] such

[1] Github repository: https://github.com/XBraid/XBraid-convergence-est.

[2] Note that in contrast to [3], we include the initial time point.

[3] Multistep time-integration schemes can be addressed in a similar way; see [6].

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

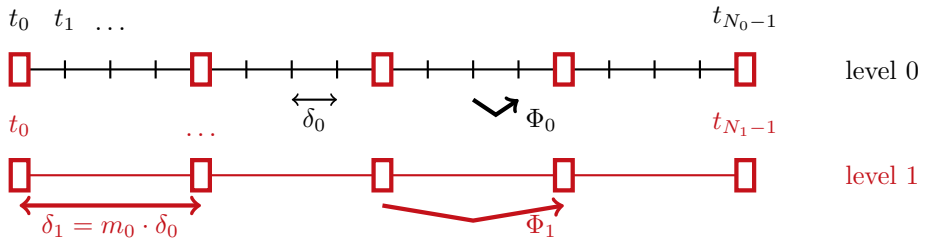

FIG. 1. *Two-grid hierarchy: Time points t_n, fine-/coarse-grid step sizes δ_0 and δ_1, and coarsening factor $m_0 = 4$. On level 0, F-points are denoted as vertical lines and C-points are denoted as squares.*

that the number of time points on each grid level is given by

$$N_\ell = \frac{N_{\ell-1} - 1}{m_{\ell-1}} + 1 \qquad \text{for } \ell = 1, \ldots, n_\ell - 1, \tag{2.3}$$

with corresponding time step size $\delta_{t,\ell}$. On each grid level ℓ, time points are partitioned into F-points (black) and C-points (red), and the C-points on level ℓ compose all points on the next coarser grid level, $\ell + 1$.

2.1. MGRIT operators. MGRIT approximates the exact coarse-grid time-stepping operator[5] on level ℓ by introducing

$$\Phi_\ell \approx \Phi_{\ell-1}^{m_{\ell-1}} \qquad \text{for } \ell = 1, \ldots, n_\ell - 1,$$

and we write

$$A_\ell = \begin{bmatrix} I & & & & \\ -\Phi_\ell & I & & & \\ & -\Phi_\ell & I & & \\ & & \ddots & \ddots & \end{bmatrix} \in \mathbb{R}^{N_x N_\ell \times N_x N_\ell} \qquad \text{for } \ell = 1, \ldots, n_\ell - 1. \tag{2.4}$$

MGRIT constructs coarse grids from a Schur complement decomposition of (2.4), relative to the F/C-splitting from Figure 1 [3]. The Schur complement arises from certain so-called ideal multigrid restriction and interpolation operators. Define ideal restriction, $R_\ell \in \mathbb{R}^{N_x N_{\ell+1} \times N_x N_\ell}$ (level ℓ to $\ell + 1$, for $\ell = 0, \ldots, n_\ell - 2$), as

$$R_\ell = \begin{bmatrix} I & \Phi_\ell^{m_\ell - 1} & \Phi_\ell^{m_\ell - 2} & \cdots & \Phi_\ell & I & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & \Phi_\ell^{m_\ell - 1} & \Phi_\ell^{m_\ell - 2} & \cdots & \Phi_\ell & I \end{bmatrix}, \tag{2.5}$$

and ideal interpolation, $P_\ell \in \mathbb{R}^{N_x N_\ell \times N_x N_{\ell+1}}$ (level $\ell + 1$ to ℓ, for $\ell = 0, \ldots, n_\ell - 2$),

[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.

along with an auxiliary operator $S_\ell \in \mathbb{R}^{N_x N_\ell \times N_x(N_\ell - N_{\ell+1})}$ for $\ell = 1, \ldots, n_\ell - 2$, as

$$P_\ell = \begin{bmatrix} I & & & \\ \Phi_\ell & & & \\ \vdots & & & \\ \Phi_\ell^{m_\ell - 1} & & & \\ & I & & \\ & \Phi_\ell & & \\ & \vdots & & \\ & \Phi_\ell^{m_\ell - 1} & & \\ & & \ddots & \\ & & & I \end{bmatrix}, \qquad S_\ell = \begin{bmatrix} 0 & & & & & & \\ I & & & & & & \\ & I & & & & & \\ & & \ddots & & & & \\ & & & I & & & \\ & & & 0 & & & \\ & & & 0 & I & & \\ & & & & & \ddots & \\ & & & & & & I \\ & & & & & & 0 \end{bmatrix}. \tag{2.6}$$

Note that the interpolation operator is not defined as the transpose of the restriction operator, since in general $\Phi_\ell \neq \Phi_\ell^T$ (for example, see [23, eqn. (28)]), and thus $P_\ell \neq R_\ell^T$. The number of block columns in S_ℓ corresponds to the total number of F-points on level ℓ. If the operator S_ℓ is applied to the right of A_ℓ, the result $A_\ell S_\ell$ is composed of all block rows in A_ℓ, and all block columns in A_ℓ that correspond to F-points on level ℓ (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_ℓ that correspond to F-points.

Using the above definitions, it is straightforward to work out that the multigrid coarse-grid operator, $R_\ell A_\ell P_\ell$, is then given by the Schur complement [3] of A_ℓ (2.4),

$$R_\ell A_\ell P_\ell = R_{I_\ell} A_\ell P_\ell = \begin{bmatrix} I & & & \\ -\Phi_\ell^{m_\ell} & I & & \\ & -\Phi_\ell^{m_\ell} & I & \\ & & \ddots & \ddots \end{bmatrix}, \tag{2.7}$$

where restriction by injection is given by the operator,

$$R_{I_\ell} = \begin{bmatrix} I & & & & & & & & & \\ & 0 & 0 & \cdots & I & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 0 & 0 & \cdots & 0 \quad I \end{bmatrix} \qquad \text{for } \ell = 1, \ldots, n_\ell - 2. \tag{2.8}$$

Here, $R_{I_\ell} \in \mathbb{R}^{N_x N_{\ell+1} \times N_x N_\ell}$ has a block structure similar to that of R_ℓ, but with all blocks Φ_ℓ^d (for $d = 1, \ldots, m_\ell - 1$) set to zero. Thus, R_{I_ℓ} restricts the C-points from level ℓ to level $\ell + 1$, omitting the respective F-points. The number of block rows corresponds to the total number of C-points on level ℓ, i.e. $N_{\ell+1}$. Also note that the inverse of A_ℓ is given analytically by [3]

$$A_\ell^{-1} = \begin{bmatrix} I & & & & \\ \Phi_\ell & I & & & \\ \Phi_\ell^2 & \Phi_\ell & \ddots & & \\ \vdots & \vdots & & I & \\ \Phi_\ell^{N_\ell - 1} & \Phi_\ell^{N_\ell - 2} & & \Phi_\ell & I \end{bmatrix} \in \mathbb{R}^{N_x N_\ell \times N_x N_\ell} \qquad \text{for } \ell = 1, \ldots, n_\ell - 1. \tag{2.9}$$

In a typical multigrid fashion, MGRIT uses a complementary relaxation process to reduce errors that are not adequately reduced on coarser grids. Because MGRIT is a reduction-based method, coarse-grid correction (should) eliminate error effectively at C-points, so this is coupled with an F-relaxation scheme to eliminate error at F-points. F-relaxation can be seen as a block-Jacobi-like method, where in this case each block consists of a set of contiguous F-points in the time domain (that is, F-relaxation updates all F-points based on sequential time integration from the closest (previous) C-point). Algebraically, this is equivalent to an application of the idempotent operator

$$F_\ell = P_\ell R_{I_\ell} = I - S_\ell (S_\ell^T A_\ell S_\ell)^{-1} S_\ell^T. \tag{2.10}$$

When F-relaxation alone is insufficient, a stronger relaxation scheme can be used. Define $T_\ell = R_{I_\ell}^T$. Then C-relaxation updates a C-point based on taking one time step from the previous F-point (equivalent to block Jacobi applied to the C-point block rows of A_ℓ). Algebraically, this corresponds to an application of

$$C_\ell = I - T_\ell (T_\ell^T A_\ell T_\ell)^{-1} T_\ell^T A_\ell, \tag{2.11}$$

where the zero columns correspond to C-points. Working through the algebra, FCF-relaxation (subsequent F-, C-, and F-relaxation steps) can be written as

$$F_\ell C_\ell F_\ell = P_\ell (I - R_\ell A_\ell P_\ell) R_{I_\ell}. \tag{2.12}$$

2.2. Simultaneous diagonalization of $\{\Phi_\ell\}$. Let Φ_ℓ denote the time-stepping operator on level ℓ. Primary results in this paper rest on the assumption that $\{\Phi_\ell\}$ are 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, $\Phi_i \Phi_j = \Phi_j \Phi_i$ for all i, j, and that Φ_ℓ is diagonalizable for all ℓ. The concept of simultaneous diagonalization (albeit, in the Fourier basis) was introduced in [11] and was modified and used as the basis for the improved two-grid convergence bounds developed in [3].

In terms of when such an assumption is valid, let $\mathcal{L}$ be a time-independent operator (such as a spatial discretization) that is propagated through time by operators $\{\Phi_\ell\}$. Note that all rational functions of $\mathcal{L}$ commute and that if $\mathcal{L}$ is diagonalizable, so is any rational function of $\mathcal{L}$. Indeed, nearly all standard time-integration routines, including all single-step Runge–Kutta-type methods, consist of some rational function of $\mathcal{L}$, and all such schemes are simultaneously diagonalizable with the eigenvectors of $\mathcal{L}$. To that end, let $\mathcal{L} = UDU^{-1}$, where $D_{kk} = \xi_k$ is a diagonal matrix containing the eigenvalues of $\mathcal{L}$ and columns of U are the corresponding eigenvectors. Denote the Butcher tableau of a general s-stage Runge–Kutta method as

$$\begin{array}{c|c} \mathfrak{c} & \mathfrak{A} \\ \hline & \mathfrak{b}^T \end{array}.$$

With some algebra, one can show that Φ_ℓ corresponding to a given Butcher tableau is exactly given by the Runge–Kutta stability function applied to $\mathcal{L}$ in block form,

$$\begin{aligned} \Phi_\ell &= I + \delta_{t,\ell} (\mathfrak{b}^T \otimes I) \left(I - \delta_{t,\ell} \mathfrak{A} \otimes \mathcal{L}\right)^{-1} (\mathbf{1}_s \otimes \mathcal{L}) \\ &= U \Big(I + \delta_{t,\ell} \left(\mathfrak{b}^T \otimes I\right) \left(I - \delta_{t,\ell} \mathfrak{A} \otimes D\right)^{-1} (\mathbf{1}_s \otimes D) \Big) U^{-1} \\ &= U \Lambda_\ell U^{-1}, \end{aligned} \tag{2.13}$$

where $(\Lambda_\ell)_{kk} = \lambda_{\ell,k}$, for $k = 1, \ldots, N_x$, are the eigenvalues of Φ_ℓ, given by

$$\lambda_{\ell,k} = 1 + \delta_{t,\ell} \xi_k \mathfrak{b}^T (I - \delta_{t,\ell} \xi_k \mathfrak{A})^{-1} \mathbf{1}. \tag{2.14}$$

Note that (2.14) is exactly the stability function for a Runge–Kutta scheme applied to $\delta_{t,\ell}\xi_k$ for time step $\delta_{t,\ell}$ and spatial eigenvalue $\{\xi_k\}$ [27, sect. 2.1, §4]. This highlights the fact that solving the spatial eigenvalue problem also provides the eigenvalues of all Φ_ℓ for arbitrary Runge–Kutta schemes and time step sizes.

Now suppose $\mathcal{A}$ is some matrix operator, where each entry is a rational function of time-stepping operators in $\{\Phi_\ell\}$. In particular, this applies to error and residual propagation operators of MGRIT that are derived in section 3. Let $\widetilde{U}$ denote a block-diagonal matrix with diagonal blocks given by U. Then, as in [41],

$$\|\mathcal{A}(\Phi_0,\dots,\Phi_{n_\ell-1})\|_{(\widetilde{U}\widetilde{U}^*)^{-1}} = \sup_k \|\mathcal{A}(\lambda_{0,k},\dots,\lambda_{n_\ell-1,k})\|_2. \tag{2.15}$$

Thus, the $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm of $\mathcal{A}(\Phi_0,\dots,\Phi_{n_\ell-1})$ can be computed by maximizing the ℓ^2-norm of $\mathcal{A}$ over eigenvalues of $\{\Phi_\ell\}$. In the case that $\{\Phi_\ell\}$ are normal matrices, U is unitary and the $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm reduces to the standard Euclidean 2-norm. More generally, we have bounds on the ℓ^2-norm of $\mathcal{A}(\Phi_0,\dots,\Phi_{n_\ell-1})$,

$$\begin{aligned}\frac{1}{\kappa(U)}\left(\sup_k \|\mathcal{A}(\lambda_{0,k},\dots,\lambda_{n_\ell-1,k})^i\|_2\right)\\ \le \|\mathcal{A}(\Phi_0,\dots,\Phi_{n_\ell-1})^i\|_2 \le \kappa(U)\left(\sup_k \|\mathcal{A}(\lambda_{0,k},\dots,\lambda_{n_\ell-1,k})^i\|_2\right)\end{aligned} \tag{2.16}$$

for $i = 1, 2, \dots$ applications of $\mathcal{A}$, where $\kappa(U)$ denotes the matrix condition number of U.[6]

Here, we are interested in $\mathcal{A}$ corresponding to the error- and residual-propagation operators of n_l-level MGRIT, denoted as $\mathcal{E}^{n_l}$ and $\mathcal{R}^{n_l}$, respectively. Convergence of MGRIT requires

$$\|\left(\mathcal{E}^{n_l}\right)^i\|, \|\left(\mathcal{R}^{n_l}\right)^i\| \to 0, \tag{2.17}$$

as iteration i increases. To that end, bounding $\sup_k \|\left(\mathcal{E}^{n_l}(\lambda_{0,k},\dots,\lambda_{n_\ell-1,k})\right)^i\| < 1$ for all k provides necessary and sufficient conditions for $\|\left(\mathcal{E}^{n_l}(\Phi_0,\dots,\Phi_{n_\ell-1})\right)^i\| \to 0$ with i (eventually), and similarly for $\mathcal{R}^{n_l}(\Phi_0,\dots,\Phi_{n_\ell-1})$.

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, thus better approximating a two-grid method. The non-Galerkin coarse grid used in MGRIT makes this relationship more complicated, and it is not definitive 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.

Bounds on two-grid convergence obtained in [3, 41] are tight to $\mathcal{O}(1/N_1)$. However, these bounds only apply to either (i) error/residual on all points and all iterations *except* the first, or (ii) error/residual for all iterations, but *only on C-points*. In the multilevel setting, it is necessary to consider convergence over *all* points for *one* iteration. To understand why this is, consider a three-level MGRIT V-cycle, with levels

[6] A similar modified norm also occurs in the case of integrating in time with a mass matrix [3].

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 [3, 41] 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 desired exact residual correction than no correction at all (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.

Extensions to the theory developed in [41] can be derived to place tight bounds on error/residual propagation for all points and one iteration [42]. It turns out that indeed convergence factors can be larger and the region of convergence with respect to $\delta_{t,\ell}\xi_k$ smaller compared with bounds on C-point error or later iterations [42]. Here, we do not analyze the two-level setting further and, rather, use this as motivation to consider the multilevel setting in detail. The remainder of this paper derives analytical multilevel error and residual propagation operators and proceeds to develop upper bounds on convergence in the ℓ^2-norm.

3. Multilevel residual and error propagation. As noted in [41], residual and error propagation are formally similar, that is,

$$\mathcal{R}^{n_\ell} = A_0\mathcal{E}^{n_\ell}A_0^{-1} = A_0(I - M^{-1}A_0)A_0^{-1} = I - A_0M^{-1}, \tag{3.1}$$

where M^{-1} denotes the MGRIT preconditioner for A_0^{-1}. Noting that there is a closed form for A^{-1} (see (2.9)), 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, which are then used to develop analytic *a priori* bounds on MGRIT convergence in section 4, as well as to construct the error propagation operator and compute its norm directly in numerical tests in section 5.

In the remainder of this work, we use the following convention for sums and products: For $b < a$, $\sum_{i=a}^{b} f_i = 0$ and $\prod_{i=a}^{b} f_i = 1$. We further write $\mathcal{E}^{n_\ell=2}$ to refer to the two-grid error propagation operator, and similarly for other numbers of levels n_ℓ.

3.1. Two-level MGRIT with *r*FCF-relaxation. Here, we generalize the two-level error propagator, as given in [3], to two-level MGRIT with rFCF-relaxation. rFCF-relaxation refers to F-relaxation followed by r CF-relaxation steps. A similar result can be found in [15], 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

$$\begin{aligned} 0 &= I - A_0^{-1}A_0 = (I - P_0(R_0A_0P_0)^{-1}R_0A_0)(F_0C_0)^rF_0 \\ &= (I - P_0(R_0A_0P_0)^{-1}R_0A_0)P_0(I - R_0A_0P_0)^rR_{I_0}, \end{aligned} \tag{3.2}$$

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*

$$\mathcal{E}_{rFCF}^{n_\ell=2} = (I - P_0A_1^{-1}R_0A_0)P_0(I - R_0A_0P_0)^rR_{I_0}. \tag{3.3}$$

Proof. This follows by substituting the coarse-grid operator $A_1 \approx R_0 A_0 P_0$ into (3.2). ∎

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 ℓ,

$$\begin{aligned} 0 &= I - A_\ell^{-1} A_\ell = (I - P_\ell (R_\ell A_\ell P_\ell)^{-1} R_\ell A_\ell)(I - S_\ell (S_\ell^T A_\ell S_\ell)^{-1} S_\ell^T A_\ell) \\ &= I - (P_\ell (R_\ell A_\ell P_\ell)^{-1} R_\ell + S_\ell (S_\ell^T A_\ell S_\ell)^{-1} S_\ell^T) A_\ell, \end{aligned} \tag{3.4}$$

and the additional relation

$$A_\ell^{-1} = P_\ell (R_\ell A_\ell P_\ell)^{-1} R_\ell + S_\ell (S_\ell^T A_\ell S_\ell)^{-1} S_\ell^T. \tag{3.5}$$

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

$$\begin{aligned} \mathcal{E}_F^{n_\ell} = P_0 R_{I_0} &- \left(\prod_{k=0}^{n_\ell - 2} P_k \right) A_{n_\ell - 1}^{-1} \left(\prod_{k=n_\ell - 2}^{0} R_k \right) A_0 P_0 R_{I_0} \\ &- \sum_{i=0}^{n_\ell - 3} \left(\prod_{k=0}^{i} P_k \right) S_{i+1} (S_{i+1}^T A_{i+1} S_{i+1})^{-1} S_{i+1}^T \left(\prod_{k=i}^{0} R_k \right) A_0 P_0 R_{I_0} \end{aligned} \tag{3.6}$$

for $n_\ell \geq 2$ levels.

Proof. For $n_\ell = 2$, we have

$$\mathcal{E}_F^{n_\ell = 2} = P_0 R_{I_0} - P_0 A_1^{-1} R_0 A_0 P_0 R_{I_0} = (I - P_0 A_1^{-1} R_0 A_0) P_0 R_{I_0},$$

which is equivalent to (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, (3.5), yields,

$$\begin{aligned} \mathcal{E}_F^{n_\ell = n} &= P_0 R_{I_0} - \left(\prod_{k=0}^{n-2} P_k \right) \Big[P_{n-1} (R_{n-1} A_{n-1} P_{n-1})^{-1} R_{n-1} \\ &\qquad + S_{n-1} (S_{n-1}^T A_{n-1} S_{n-1})^{-1} S_{n-1}^T \Big] \left(\prod_{k=n-2}^{0} R_k \right) A_0 P_0 R_{I_0} \\ &\qquad - \sum_{i=0}^{n-3} \left(\prod_{k=0}^{i} P_k \right) S_{i+1} (S_{i+1}^T A_{i+1} S_{i+1})^{-1} S_{i+1}^T \left(\prod_{k=i}^{0} R_k \right) A_0 P_0 R_{I_0} \\ &= P_0 R_{I_0} - \left(\prod_{k=0}^{n-1} P_k \right) (R_{n-1} A_{n-1} P_{n-1})^{-1} \left(\prod_{k=n-1}^{0} R_k \right) A_0 P_0 R_{I_0} \\ &\qquad - \sum_{i=0}^{n-2} \left(\prod_{k=0}^{i} P_k \right) S_{i+1} (S_{i+1}^T A_{i+1} S_{i+1})^{-1} S_{i+1}^T \left(\prod_{k=i}^{0} R_k \right) A_0 P_0 R_{I_0}. \end{aligned}$$

Approximating the exact coarse grid operator on level $n+1$ by $A_n \approx R_{n-1} A_{n-1} P_{n-1}$ completes the proof. ∎

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 ℓ,

$$\begin{aligned} 0 &= I - A_\ell^{-1}A_\ell = (I - P_\ell(R_\ell A_\ell P_\ell)^{-1}R_\ell A_\ell)F_\ell C_\ell F_\ell \\ &= I - P_\ell(R_\ell A_\ell P_\ell)^{-1}R_\ell A_\ell - S_\ell(S_\ell^T A_\ell S_\ell)^{-1}S_\ell^T A_\ell - T_\ell(T_\ell^T A_\ell T_\ell)^{-1}T_\ell^T A_\ell \\ &\quad + S_\ell(S_\ell^T A_\ell S_\ell)^{-1}S_\ell^T A_\ell T_\ell(T_\ell^T A_\ell T_\ell)^{-1}T_\ell^T A_\ell \\ &\quad + P_\ell(R_\ell A_\ell P_\ell)^{-1}R_\ell A_\ell T_\ell(T_\ell^T A_\ell T_\ell)^{-1}T_\ell^T A_\ell, \end{aligned} \tag{3.7}$$

and the additional relation

$$\begin{aligned} A_\ell^{-1} &= T_\ell(T_\ell^T A_\ell T_\ell)^{-1}T_\ell^T \\ &\quad + \left[S_\ell(S_\ell^T A_\ell S_\ell)^{-1}S_\ell^T + P_\ell(R_\ell A_\ell P_\ell)^{-1}R_\ell\right]\left[I - A_\ell T_\ell(T_\ell^T A_\ell T_\ell)^{-1}T_\ell^T\right]. \end{aligned} \tag{3.8}$$

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

$$\begin{aligned} \mathcal{E}_{FCF}^{n_\ell} &= P_0(I - (T_0^T A_0 T_0)^{-1}R_{I_0}A_0P_0)R_{I_0} \\ &\quad - \left(\prod_{k=0}^{n_\ell-2} P_k\right) A_{n_\ell-1}^{-1}\left(\prod_{k=n_\ell-2}^{0} R_k\left[I - A_kT_k(T_k^T A_kT_k)^{-1}T_k^T\right]\right) A_0P_0R_{I_0} \\ &\quad - \sum_{i=1}^{n_\ell-2}\left(\prod_{k=0}^{i-1} P_k\right)\Big[S_i(S_i^T A_iS_i)^{-1}S_i^T\left[I - A_iT_i(T_i^T A_iT_i)^{-1}T_i^T\right] \\ &\quad + T_i(T_i^T A_iT_i)^{-1}T_i^T\Big]\left(\prod_{k=i-1}^{0} R_k\left[I - A_kT_k(T_k^T A_kT_k)^{-1}T_k^T\right]\right) A_0P_0R_{I_0}, \end{aligned} \tag{3.9}$$

with $n_\ell \geq 2$ levels.

Proof. The proof is analogous to the proof of Lemma 3.2. ∎

3.4. Multilevel F-cycles with rFCF-relaxation. Similar to the notation used for V-cycle error propagation, let $\mathcal{F}^{n_\ell}$ denote error propagation of MGRIT F-cycles with n_ℓ levels, with a subscript denoting relaxation scheme. Following [19, p. 53], error propagation of MGRIT for a multilevel F-cycle with rFCF-relaxation can be defined recursively,

$$\mathcal{F}_{rFCF}^{n_\ell} = M_{rFCF,0}^F \qquad \text{for } n_\ell \geq 2, \tag{3.10}$$

with

$$\begin{aligned} M_{rFCF,\ell-1}^F &= P_{\ell-1}\left(I - \left(I - M_\ell^V M_\ell^F\right)A_\ell^{-1}R_{\ell-1}A_{\ell-1}P_{\ell-1}\right)(I - R_{\ell-1}A_{\ell-1}P_{\ell-1})^r R_{I_{\ell-1}}, \\ M_{rFCF,\ell-1}^V &= P_{\ell-1}\left(I - \left(I - M_\ell^V\right)A_\ell^{-1}R_{\ell-1}A_{\ell-1}P_{\ell-1}\right)(I - R_{\ell-1}A_{\ell-1}P_{\ell-1})^r R_{I_{\ell-1}} \end{aligned}$$

for $l = 1, \ldots, n_\ell - 2$, and

$$\begin{aligned} M_{rFCF,n_\ell-2}^F &= M_{rFCF,n_\ell-2}^V \\ &= P_{n_\ell-2}\left(I - A_{n_\ell-1}^{-1}R_{n_\ell-2}A_{n_\ell-2}P_{n_\ell-2}\right)(I - R_{n_\ell-2}A_{n_\ell-2}P_{n_\ell-2})^r R_{I_{n_\ell-2}}. \end{aligned}$$

It is easy to verify that for $n_\ell = 2$, the recursive formulae result in $\mathcal{F}_{rFCF}^{n_\ell=2} = \mathcal{E}_{rFCF}^{n_\ell=2}$. For $n_\ell = 3$ and $r = 0$, we can write

$$\mathcal{F}_F^{n_\ell=3} = \mathcal{E}_F^{n_\ell=2} + P_0P_1\left(I - A_2^{-1}R_1A_1P_1\right)^2 R_{I_1}A_1^{-1}R_0A_0P_0R_{I_0}. \tag{3.11}$$

However, it is not straightforward to convert the recursive definition in (3.10) into a summation similar to (3.6) or (3.3), for arbitrary n_ℓ. Nevertheless, this formula is still useful for numerically computing bounds of $\mathcal{F}_{rFCF}^{n_\ell}$ and is thus included for completeness.

4. Bounds for MGRIT residual and error propagation. Following the work in [3], we assume that operators Φ_ℓ, $\ell = 0, \dots, n_\ell - 1$, can be diagonalized by the same set of eigenvectors (see (2.13)), with eigenvalues denoted $\{\lambda_{\ell,k}\}$ for $1 \le k \le N_x$. We also assume that Φ_ℓ are strongly stable time-stepping operators, that is, $\|\Phi_\ell\| < 1$, which implies $|\lambda_{\ell,k}| < 1$ for all $\ell = 0, \dots, n_\ell - 1$ and $k = 1, \dots, N_x$.[7] To simplify notation in the following derivations, we use Φ_ℓ to denote the diagonalized time-stepping operator moving forward. Results then follow in a $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm, which (as discussed in section 2.2) is equivalent to the ℓ^2-norm if Φ_ℓ is normal (and, thus, U is unitary).

For ease of presentation and because many of the derivations are fairly involved, but repetitive, a number of steps are moved to the supplementary materials. We refer the interested reader to section SM2.

4.1. Residual and error on levels 0 and 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. The norms of the residual and error propagation operators are equal in the $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm (recall that $\widetilde{U}$ is a block-diagonal matrix of eigenvectors, U).[8] If $\{\Phi_\ell\}$ are normal operators, they are diagonalizable by unitary transformation, in which case $\widetilde{U}\widetilde{U}^* = I$, and error and residual propagation are equal in the ℓ^2-norm.

Similarly to section 3, let $\mathcal{E}_{rFCF}^{n_\ell}$ be the n_l-level error propagator, acting on all points on level 0. We further refer to $\mathcal{E}_{rFCF}^{n_\ell,\Delta}$ as the error propagator that acts on all points on level 1, i.e., on the error at the C-points on level 0. In the two-grid setting, we also refer to $\mathcal{E}_{rFCF}^{n_\ell,\Delta}$ as the coarse-grid error propagator.

To quantify how *fast* MGRIT converges in the *worst case*, we can bound the convergence factor of the fine-grid residual [3] $\mathbf{r}_{i+1}$ at iteration $i+1$, $i \in \mathbb{N}_0$, by the norm of the error propagator on level 1 (in the unitary case),

$$\|\mathbf{r}_{i+1}\|_2 / \|\mathbf{r}_i\|_2 = \|A_1 \mathbf{e}_{i+1}^\Delta\|_2 / \|A_1 \mathbf{e}_i^\Delta\|_2 \le \|A_1 \mathcal{E}_{rFCF}^{n_\ell,\Delta} A_1^{-1}\|_2 = \|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2, \tag{4.1}$$

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

$$\begin{aligned} \mathbf{e}_{i+1}^\Delta = \mathcal{E}_{rFCF}^{n_\ell,\Delta} \mathbf{e}_i^\Delta = \mathcal{E}_{rFCF}^{n_\ell,\Delta} R_{I_0} \mathbf{e}_i \\ \Leftrightarrow \quad P_0 \mathbf{e}_{i+1}^\Delta = P_0 \mathcal{E}_{rFCF}^{n_\ell,\Delta} \mathbf{e}_i^\Delta = P_0 \mathcal{E}_{rFCF}^{n_\ell,\Delta} R_{I_0} \mathbf{e}_i, \end{aligned} \tag{4.2}$$

we can identify $\mathcal{E}_{rFCF}^{n_\ell,\Delta} = R_{I_0} \mathcal{E}_{rFCF}^{n_\ell} P_0$, which is a generalization of the approach in [3], 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).

[7] Note that it is possible to have a stable time-integration scheme with $\|\Phi_\ell\| > 1$ if $\|\Phi_\ell^i\| < 1$ for some i [31, 29, sect. 9.5, eqn. (9.22)], but we do not consider such schemes.

[8] Although [41] specifically addresses two-grid bounds, equality of error and residual propagation in the $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm follows if Φ_ℓ is simultaneously diagonalizable for all levels ℓ.

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

$$\begin{aligned}\mathbf{e}_{i+1} &= \mathcal{E}_{rFCF}^{n_\ell}\mathbf{e}_i = \cdots = \left(\mathcal{E}_{rFCF}^{n_\ell}\right)^{i+1}\mathbf{e}_0 \\ &= \left(P_0\mathcal{E}_{rFCF}^{n_\ell,\Delta}R_{I_0}\right)^{i+1}\mathbf{e}_0 = P_0\left(\mathcal{E}_{rFCF}^{n_\ell,\Delta}\right)^{i+1}R_{I_0}\mathbf{e}_0,\end{aligned} \tag{4.3}$$

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.

LEMMA 4.1. *Error propagation on level* 0 *for an MGRIT V-cycle method can be bounded by error propagation on level* 1*,*

$$\|\mathcal{E}_{rFCF}^{n_\ell}\|_2 \leq \sqrt{m_0}\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2, \tag{4.4}$$

with temporal coarsening factor m_0 *on level* 0*.*

Proof. This follows from

$$\begin{aligned}\|\mathcal{E}_{rFCF}^{n_\ell}\|_2 &= \|P_0\mathcal{E}_{rFCF}^{n_\ell,\Delta}R_{I_0}\|_2 \leq \|P_0\|_2\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2\|R_{I_0}\|_2 \\ &\leq \sqrt{\|P_0\|_1\|P_0\|_\infty}\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2\sqrt{\|R_{I_0}\|_1\|R_{I_0}\|_\infty} \leq \sqrt{m_0}\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2,\end{aligned}$$

with submultiplicativity and the inequality $\|D\|_2 \leq \sqrt{\|D\|_1\|D\|_\infty}$ (see [24]). ∎

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 0 converges as well. This is the basis for the theory developed in the rest of this work, where convergence is attained by bounding $\mathcal{E}_{rFCF}^{n_\ell,\Delta}$ 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 [41] 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 (*a priori*) from above via

$$c_f = \max_i \|\mathbf{r}_{i+1}\|_2/\|\mathbf{r}_i\|_2 \leq \|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2, \tag{4.5}$$

which corresponds to the maximum singular value of $\mathcal{E}_{rFCF}^{n_\ell,\Delta}$. In practice, the dimension of $\mathcal{E}_{rFCF}^{n_\ell}$ grows with the problem size in space and time, N_x and N_0. Similarly, $\mathcal{E}_{rFCF}^{n_\ell,\Delta}$ 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 *a priori* 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 ℓ^2-norm of the error propagator on level 1 using the well-known inequality [24]

$$\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_2^2 \leq \|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_1\|\mathcal{E}_{rFCF}^{n_\ell,\Delta}\|_\infty. \tag{4.6}$$

In [3], this was used to develop an upper-bound on two-grid convergence, which was proven to be sharp in [41]. This section extends this approach to three and four grid levels based on analytic formulae. Although the sharpness of (4.6) 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 *r*FCF-relaxation. The coarse-grid error propagator follows from (3.3) with $n_\ell = 2$ (see section SM2.1),

$$\mathcal{E}_{rFCF}^{n_\ell=2,\Delta} = \begin{bmatrix} 0 & & & & & \\ \vdots & & & & & \\ 0 & & & & & \\ (\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & & & & & \\ \Phi_1(\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & & (\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & & & \\ \vdots & & & & & \\ \Phi_1^{N_1-r-2}(\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & \Phi_1^{N_1-r-3}(\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & \cdots & (\Phi_0^{m_0} - \Phi_1)\Phi_0^{rm_0} & 0 & \cdots & 0 \end{bmatrix}, \tag{4.7}$$

where the first $r+1$ rows and last $r+1$ columns are zero.

THEOREM 4.3. *Let $\{\lambda_{\ell,k}\}$ be the eigenvalues of $\{\Phi_\ell\}$. Then the* worst-case *convergence factor of the fine-grid residual of two-level MGRIT with rFCF-relaxation is bounded by*

$$c_f \leq \max_{1\leq k\leq N_x} |\lambda_{0,k}^{m_0} - \lambda_{1,k}||\lambda_{0,k}|^{rm_0} \frac{1-|\lambda_{1,k}|^{N_1-1-r}}{1-|\lambda_{1,k}|}. \tag{4.8}$$

Proof. This follows from equation (4.7) and inequality (4.6),

$$\begin{aligned} \|\mathcal{E}_{rFCF}^{n_\ell=2,\Delta}\|_2 &\leq \sqrt{\|\mathcal{E}_{rFCF}^{n_\ell=2,\Delta}\|_1 \|\mathcal{E}_{rFCF}^{n_\ell=2,\Delta}\|_\infty} = \|\mathcal{E}_{rFCF}^{n_\ell=2}\|_1 \\ &= \max_{1\leq k\leq N_x} |\lambda_{0,k}^{m_0} - \lambda_{1,k}||\lambda_{0,k}|^{rm_0} \frac{1-|\lambda_{1,k}|^{N_1-1-r}}{1-|\lambda_{1,k}|}, \end{aligned}$$

where the relationship $\sum_{i=0}^{N-1} a^i = (1-a^N)/(1-a)$ was used. ∎

Remark 4.4. The cases of F- and FCF-relaxation (i.e., $r=0$ and $r=1$) yield the result in [3],

$$\begin{aligned} \|\mathcal{E}_{F}^{n_\ell=2,\Delta}\|_2 &\leq \max_{1\leq k\leq N_x} |\lambda_{0,k}^{m_0} - \lambda_{1,k}| \frac{1-|\lambda_{1,k}|^{N_1-1}}{1-|\lambda_{1,k}|}, \\ \|\mathcal{E}_{FCF}^{n_\ell=2,\Delta}\|_2 &\leq \max_{1\leq k\leq N_x} |\lambda_{0,k}^{m_0} - \lambda_{1,k}||\lambda_{0,k}|^{m_0} \frac{1-|\lambda_{1,k}|^{N_1-2}}{1-|\lambda_{1,k}|}. \end{aligned}$$

In [41], 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 (4.7) 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 reorder the rows and columns of the coarse-grid error propagator, yielding a block-diagonal form with lower triangular nilpotent blocks.

LEMMA 4.5. *Let $\{\Phi_\ell\}$ be simultaneously diagonalizable by the same unitary transformation, with eigenvalues $\{\lambda_{\ell,k}\}$, such that $|\lambda_{\ell,k}| < 1$. Then the ℓ^2-norm of the coarse-grid error propagator of two-level MGRIT with rFCF-relaxation can be computed as*

$$\|\mathcal{E}_{rFCF}^{n_\ell=2,\Delta}\|_2 = \sup_{1\leq k\leq N_x} \|\tilde{\mathcal{E}}_{rFCF}^{n_\ell=2,\Delta}(k)\|_2, \tag{4.9}$$

with the coarse-grid error propagator $\tilde{\mathcal{E}}_{rFCF}^{n_\ell=2,\Delta}(k)$ *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 [3, Rem. 3.1]. ∎

Remark 4.6. Lemma 4.5 implies that computing a bound of the form of (4.5) can be parallelized over the number of spatial modes. Thus, the time complexity of evaluating (4.5) is $O(N_x N_1^3/p)$ with $1 \leq p \leq N_x$ parallel processors.

Remark 4.7. Lemma 4.5 formalizes and generalizes the discussion for two-level MGRIT with F- and FCF-relaxation in [11, sect. 4.2].

Remark 4.8. The result in Lemma 4.5 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 (3.6) for a three-level V-cycle with F-relaxation on level 1 (see equation (SM2.5)) and comparison with the two-level error propagator for F-relaxation in [3] 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.[9] 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* $\{\lambda_{\ell,k}\}$ *be the eigenvalues of* $\{\Phi_\ell\}$. *Then the* worst-case *convergence factor of three-level MGRIT with F-relaxation is bounded by*

$$c_f \leq \sqrt{\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_1 \|\mathcal{E}_F^{n_\ell=3,\Delta}\|_\infty}, \tag{4.10}$$

and $\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_1$ *and* $\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_\infty$ *are given analytically as*

$$\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_1 = \max_{1\leq k\leq N_x} \begin{cases} |\lambda_{2,k} - \lambda_{0,k}^{m_0}\lambda_{1,k}^{m_1-1}| \left(|\lambda_{2,k}|^{N_2-2} + \frac{1-|\lambda_{2,k}|^{N_2-2}}{1-|\lambda_{2,k}|}\frac{1-|\lambda_{1,k}|^{m_1}}{1-|\lambda_{1,k}|}\right) \\ \quad +|\lambda_{1,k} - \lambda_{0,k}^{m_0}|\frac{1-|\lambda_{1,k}|^{m_1-1}}{1-|\lambda_{1,k}|} \\ \hline |\lambda_{1,k}|^{j-1}|\lambda_{1,k} - \lambda_{0,k}^{m_0}| \left[|\lambda_{2,k}|^{N_2-2} + \frac{1-|\lambda_{2,k}|^{N_2-2}}{1-|\lambda_{2,k}|}\frac{1-|\lambda_{1,k}|^{m_1}}{1-|\lambda_{1,k}|}\right] \\ \quad +|\lambda_{1,k} - \lambda_{0,k}^{m_0}|\frac{1-|\lambda_{1,k}|^{m_1-2}}{1-|\lambda_{1,k}|} \qquad \text{for } j = 1,\dots,m_1-1 \end{cases} \tag{4.11}$$

and

$$\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_\infty = \max_{1\leq k\leq N_x} \begin{cases} |\lambda_{2,k} - \lambda_{0,k}^{m_0}\lambda_{1,k}^{m_1-1}|\frac{1-|\lambda_{2,k}|^{N_2-1}}{1-|\lambda_{2,k}|} \\ \quad +|\lambda_{1,k} - \lambda_{0,k}^{m_0}|\frac{1-|\lambda_{2,k}|^{N_2-1}}{1-|\lambda_{2,k}|}\frac{1-|\lambda_{1,k}|^{m_1-1}}{1-|\lambda_{1,k}|} \\ \hline |\lambda_{1,k} - \lambda_{0,k}^{m_0}|\frac{1-|\lambda_{1,k}|^{j}}{1-|\lambda_{1,k}|} + |\lambda_{1,k}|^j\frac{1-|\lambda_{2,k}|^{N_2-2}}{1-|\lambda_{2,k}|}|\lambda_{2,k} - \lambda_{0,k}^{m_0}\lambda_{1,k}^{m_1-1}| \\ \quad +|\lambda_{1,k}|^j\frac{1-|\lambda_{2,k}|^{N_2-2}}{1-|\lambda_{2,k}|}|\lambda_{1,k} - \lambda_{0,k}^{m_0}|\frac{1-|\lambda_{1,k}|^{m_1-1}}{1-|\lambda_{1,k}|} \qquad \text{for } j = 1,\dots,m_1-1. \end{cases} \tag{4.12}$$

Proof. The proof is analogous to that of Theorem 4.3. ∎

The benefit of Theorem 4.9 is that evaluating the $2m_1$ analytic formulae is significantly cheaper than constructing $\mathcal{E}_F^{n_\ell=3,\Delta}$ numerically and directly computing $\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_1$ and $\|\mathcal{E}_F^{n_\ell=3,\Delta}\|_\infty$.

[9] Note that additional relaxation steps did not break symmetry of $\mathcal{E}_{rFCF}^{n_l=2,\Delta}$ in (4.7).

4.2.3. Analytic formulae for other cases. Analogous to sections 4.2.1 and 4.2.2, analytic formulae for a four-level V-cycle with F-relaxation and a three-level V-cycle with FCF-relaxation are derived in sections SM2.2.1 and SM2.2.2.

4.3. Approximate convergence factor of multilevel V-cycle algorithm. Section 4.2 presented analytic formulae for the inequality bound (4.6) as the maximum of a certain function over eigenvalues of $\{\Phi_\ell\}$. 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. Here, we propose an analytic approximate convergence factor for multilevel V-cycles with F- and FCF-relaxation, as a function of eigenvalues of the time-stepping operators $\{\lambda_{\ell,k}\}$, number of time points $\{N_\ell\}$, and temporal coarsening factors for each level, $\{m_\ell\}$. This yields approximate *a priori* convergence factors with linear memory and time complexity.[10]

The proposed approximate convergence factors are based on approximating the inequality bound (4.6) and therefore are expected to be a conservative upper bound in a large number of cases. More specifically, in the case of multilevel V-cycles with F-relaxation, the approximate convergence factor is derived by identifying the recursive structure in the analytic formulae for two, three, and four levels (see (4.8), (4.10), and (SM2.10)) and estimating how this recursion continues for $n_l > 4$ levels (and similarly for FCF-relaxation with (4.8) and (SM2.11)).

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

Approximation 1. Let $\{\lambda_{\ell,k}\}$ be the eigenvalues of $\{\Phi_\ell\}$. Then an approximate worst-case convergence factor of multilevel MGRIT V-cycles with F-relaxation is given by

$$\tilde{c}_{f,F} := \max_{1\le k\le N_x} \sqrt{s_0^{\mathrm{row}}(k,n_\ell)\; s_{N_1-1}^{\mathrm{col}}(k,n_\ell)} \approx \sqrt{\|\mathcal{E}_F^{n_\ell,\Delta}\|_1 \|\mathcal{E}_F^{n_\ell,\Delta}\|_\infty}, \tag{4.13}$$

with approximate maximum absolute column and row sums

$$\begin{aligned} s_0^{\mathrm{col}}(k,n_\ell) \approx \sum_{\ell=1}^{n_\ell-1} \left|\lambda_{\ell,k} - \lambda_{0,k}^{m_0}\left(\prod_{p=1}^{l-1}\lambda_{p,k}^{\tilde{m}_p-1}\right)\right| \left(\prod_{q=1}^{l}\frac{1-|\lambda_{q,k}|^{\tilde{m}_q-1}}{1-|\lambda_{q,k}|}\right) \\ + (n_\ell>2)\cdot|\lambda_{n_\ell-1,k}|^{\tilde{m}_{n_\ell-1}-1}\left|\lambda_{n_\ell-1,k} - \lambda_{0,k}^{m_0}\left(\prod_{p=1}^{n_\ell-2}\lambda_{p,k}^{\tilde{m}_p-1}\right)\right|, \end{aligned} \tag{4.14}$$

$$s_{N_1-1}^{\mathrm{row}}(k,n_\ell) \approx \sum_{\ell=1}^{n_\ell-1} \left|\lambda_{\ell,k} - \lambda_{0,k}^{m_0}\left(\prod_{p=1}^{l-1}\lambda_{p,k}^{\tilde{m}_p-1}\right)\right| \left(\prod_{q=l}^{n_\ell-1}\frac{1-|\lambda_{q,k}|^{\tilde{m}_q}}{1-|\lambda_{q,k}|}\right) \tag{4.15}$$

for $\tilde{m}_\ell = [m_0,\dots,m_{n_\ell-2},N_{n_\ell-1}-1]^T$. In many cases, $\|\mathcal{E}_F^{n_\ell,\Delta}\|_2 \le \tilde{c}_{f,F}$, because $\tilde{c}_{f,F}$ directly approximates an upper bound on $\|\mathcal{E}_F^{n_\ell,\Delta}\|_2$ (4.6).

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

Approximation 2. Let $\{\lambda_{\ell,k}\}$ be the eigenvalues of $\{\Phi_\ell\}$. Then an approximate worst-case convergence factor of multilevel MGRIT V-cycles with FCF-relaxation is

[10] The generalization of Lemma 4.5 implies that time complexity is in fact $O(N_x/p)$ with $1 \le p \le N_x$ parallel processors.

given by

$$\tilde{c}_{f,FCF} := \max_{1\le k\le N_x} \sqrt{s_0^{\text{row}}(k,n_\ell)\; s_{N_1-1}^{\text{col}}(k,n_\ell)} \approx \sqrt{\|\mathcal{E}_{FCF}^{n_\ell,\Delta}\|_1 \|\mathcal{E}_{FCF}^{n_\ell,\Delta}\|_\infty}, \tag{4.16}$$

with approximate maximum absolute column and row sum

$$\begin{aligned}
s_0^{\text{col}}(k,n_\ell) \approx{}& (n_\ell > 2)\cdot |\lambda_{0,k}|^{m_0} |\lambda_{1,k} - \lambda_{0,k}^{m_0}| \frac{1-|\lambda_{1,k}|^{m_1}}{1-|\lambda_{1,k}|} \\
&+ \frac{1}{n_\ell - 1} |\lambda_{0,k}|^{m_0} \left[\sum_{p=2}^{n_\ell-2} \left(\prod_{j=1}^{p-1} |\lambda_{j,k}| \right) \left| \lambda_{p,k} - \lambda_{0,k}^{m_0} \left(\prod_{j=1}^{p-1} \lambda_{j,k}^{m_j-1} \right) \right| \left(\prod_{j=1}^{p} \frac{1-|\lambda_{j,k}|^{m_j}}{1-|\lambda_{j,k}|} \right) \right] \\
&+ \frac{1}{n_\ell - 1} \frac{1-|\lambda_{n_\ell-1,k}|^{N_{n_\ell-1}-1}}{1-|\lambda_{n_\ell-1,k}|} |\lambda_{0,k}| \left(\prod_{j=0}^{n_\ell-2} |\lambda_{j,k}|^{m_j-1} \right) \left(\prod_{j=1}^{n_\ell-2} \frac{1-|\lambda_{j,k}|^{m_j}}{1-|\lambda_{j,k}|} \right) |\lambda_{1,k} - \lambda_{0,k}^{m_0}| \\
&+ \frac{1}{n_\ell - 1} \frac{1-|\lambda_{n_\ell-1,k}|^{N_{n_\ell-1}-1}}{1-|\lambda_{n_\ell-1,k}|} |\lambda_{0,k}|^{m_0} \left(\prod_{j=1}^{n_\ell-2} |\lambda_{j,k}| \right) \left(\sum_{p=2}^{n_\ell-1} \left| \lambda_{p,k} - \lambda_{0,k}^{m_0} \left(\prod_{j=1}^{p-1} \lambda_{j,k}^{m_j-1} \right) \right| \right) \left(\prod_{j=1}^{n_\ell-2} \frac{1-|\lambda_{j,k}|^{m_j}}{1-|\lambda_{j,k}|} \right), \\
s_{N_1-1}^{\text{row}}(k,n_\ell) \approx{}& |\lambda_{0,k}|^{m_0} \frac{1-|\lambda_{n_\ell-1,k}|^{N_{n_\ell-1}-1}}{1-|\lambda_{n_\ell-1,k}|} \left[\sum_{p=1}^{n_\ell-1} \left(\prod_{j=1}^{p-1} |\lambda_{j,k}| \right) \left| \lambda_{p,k} - \lambda_{0,k}^{m_0} \left(\prod_{j=1}^{p-1} \lambda_{j,k}^{m_j-1} \right) \right| \left(\prod_{j=p}^{n_\ell-2} \frac{1-|\lambda_{j,k}|^{m_j}}{1-|\lambda_{j,k}|} \right) \right].
\end{aligned}$$

In many cases, $\|\mathcal{E}_{FCF}^{n_\ell,\Delta}\|_2 \le \tilde{c}_{f,FCF}$, because $\tilde{c}_{f,FCF}$ directly approximates an upper bound on $\|\mathcal{E}_{FCF}^{n_\ell,\Delta}\|_2$ (4.6).

5. Numerical results. All numerical, analytic, and approximate bounds on convergence from section 4 are implemented in MPI/C++,[11] using the open-source library Armadillo [37, 38]. In this section, we evaluate these bounds for various model problems. Analytic formulae, e.g., for the inequality bound (4.6), are employed whenever available: for example, for a two-, three-, and four-level V-cycle with F-relaxation, we evaluate the analytic formulae derived in section 4.2, while for more than four levels, we construct the error propagator numerically and directly compute its 1-/∞-norm bounds.

This section assesses 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 [20, 21] of orders 1–4 (Butcher tableaux provided in section SM3). In [3], 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 δ_x). Derived bounds and approximate convergence factors are compared with the maximum observed convergence factor in numerical simulations, in terms of the ℓ^2-norm of the residual (see (4.1)),

$$\max_i \|\mathbf{r}_{i+1}\|_2 / \|\mathbf{r}_i\|_2. \tag{5.1}$$

[11] Github repository: https://github.com/XBraid/XBraid-convergence-est. For more details, see section SM1 in the supplementary materials.

All test cases are implemented in MPI/C++, using the open-source libraries Armadillo [37, 38] and XBraid [46]. The absolute stopping tolerance for MGRIT is selected as $\|\mathbf{r}_i\|_2 < 10^{-11}$, 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 $\mathbf{x} \in \Omega = (0, 2\pi) \times (0, 2\pi)$,

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

with homogeneous boundary and discontinuous initial condition (see Figure SM1),

$$u(\mathbf{x}, \cdot) = 0 \qquad \text{for } \mathbf{x} \in \partial\Omega,$$
$$u(\cdot, 0) = 1 - \max\Big\{\operatorname{sign}\Big(\big(4 - (x_1 - \pi + 1)^2 - 4(x_2 - \pi)^2\big)^2 + 1.2(1 + \pi - x_1)^3 - 10\Big), 1\Big\} \qquad \text{for } \mathbf{x} \in \Omega \cup \partial\Omega,$$

for a scalar solution $u(\mathbf{x}, t)$ and boundary $\partial\Omega$. Here, $K = \operatorname{diag}(k_1, k_2) = \text{const}$ 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 Φ_ℓ 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,

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

ranges between $\mathrm{CFL}_0 \approx 0.376$ on level 0 and $\mathrm{CFL}_5 \approx 12.036$ on level 5. Results for F-relaxation are shown in Figure 2 and FCF-relaxation in Figure 3 (note the difference in y-axis limits; results for SDIRK3 can be seen in Figures SM2 and SM3 in the supplementary materials).

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 [5]. 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 Approximations 1 and 2 estimate observed convergence as well as 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 Runge–Kutta schemes.

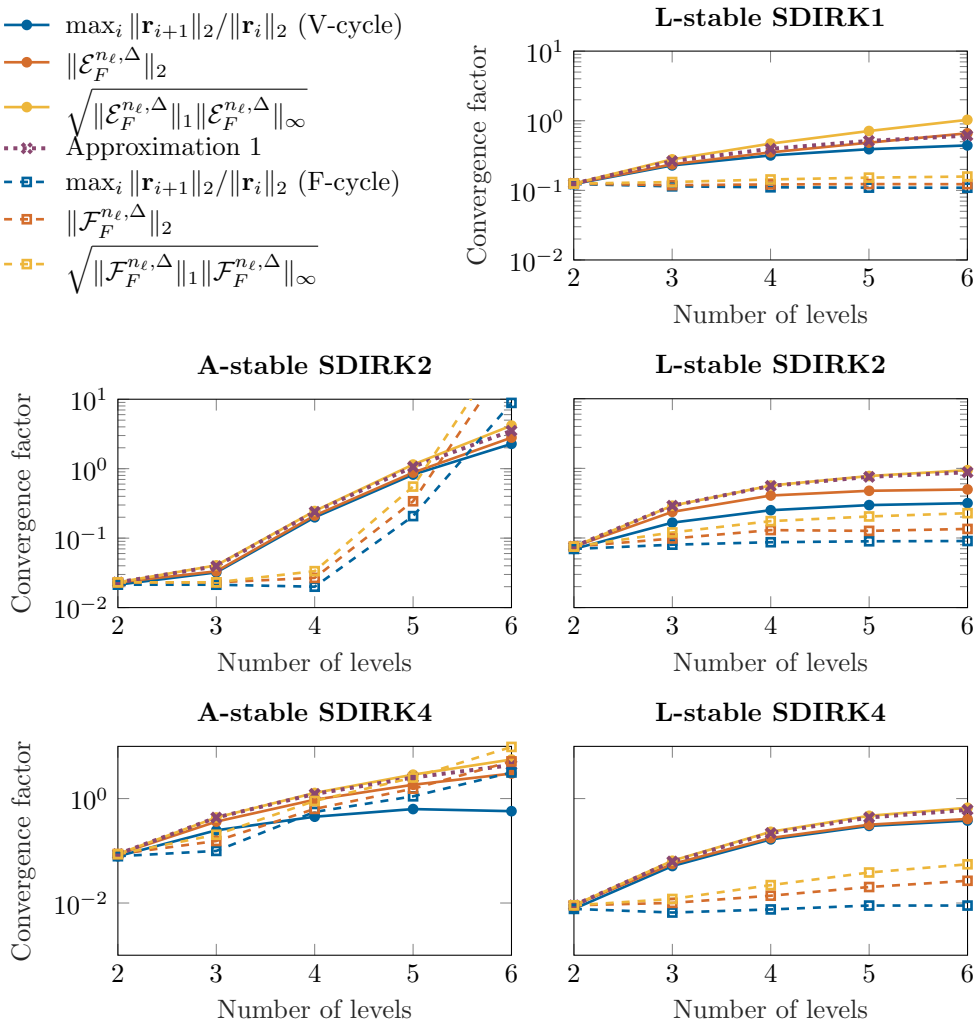

FIG. 2. *Isotropic diffusion: Comparison of V- and F-cycle MGRIT with F-relaxation. Convergence of A-stable schemes deteriorates much quicker with a growing number of time grid levels and V-cycle MGRIT than for L-stable schemes and V-cycle MGRIT. The convergence factor for L-stable schemes and F-cycle MGRIT is almost constant.*

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. Here, conductivity parameters are given by $k_1 = 0.5$ and $k_2 = 0.001$, and the CFL number on each level, $\mathrm{CFL}_\ell = 2\pi(k_1 + k_2)/[\delta_x^2(N_\ell - 1)]$, ranges between $\mathrm{CFL}_0 \approx 0.009$ on level 0 and $\mathrm{CFL}_5 \approx 0.302$ on level 5. Results are presented in Figure SM4 in the supplementary

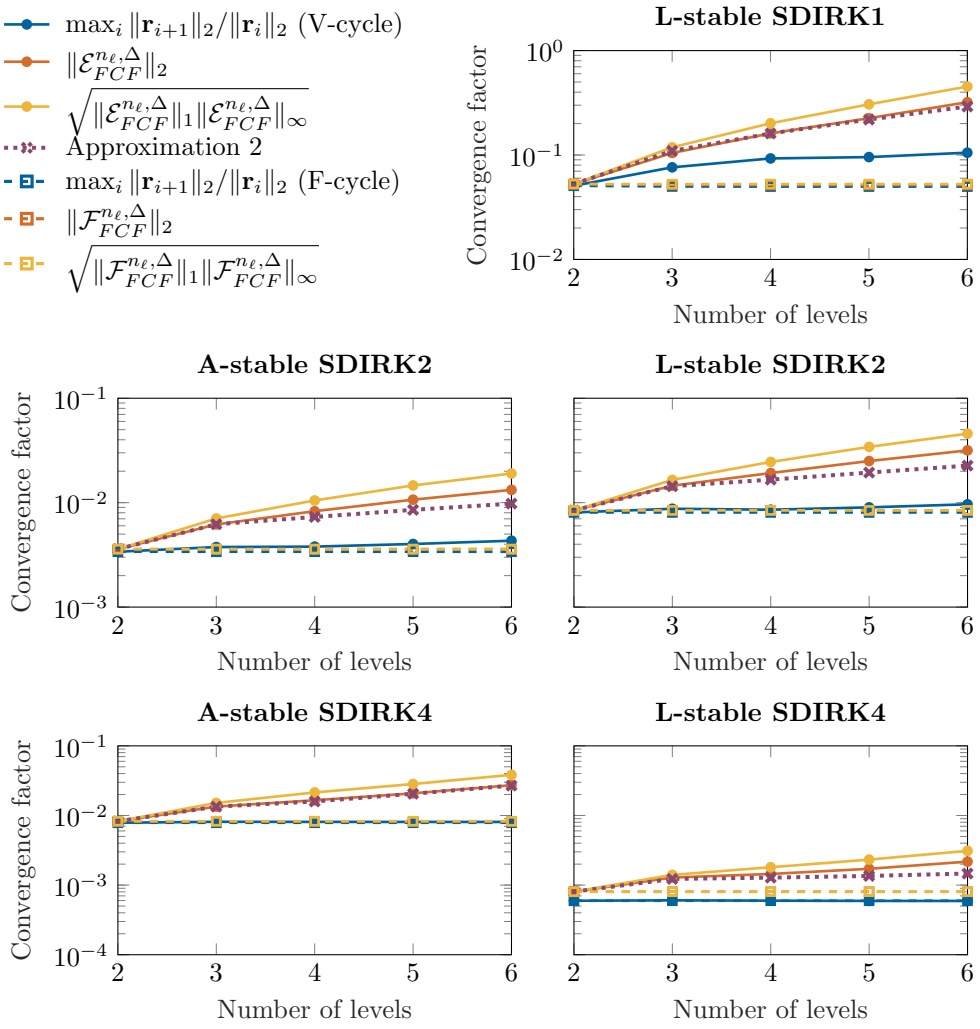

FIG. 3. *Isotropic diffusion: Comparison of V- and F-cycle MGRIT with FCF-relaxation. Convergence of A-stable and L-stable schemes deteriorates only slightly for an MGRIT V-cycle algorithm. On the other hand, the convergence factor for an F-cycle MGRIT algorithm is constant for all considered Runge–Kutta schemes and cases.*

materials.

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 means that, for this problem, an F-cycle solves the coarse-grid problem sufficiently accurately that residual and error reduction is more or less equivalent to a two-level method. Conversely, convergence in the case of V-cycles deteriorates due to inexact solves of the coarse-grid problem on each level. However, the fact that solving the coarse-grid problem more accurately (such as with F-cycles) improves convergence indicates 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 that this is in contrast to using algebraic multigrid to solve anisotropic diffusion discretizations in the spatial setting, where stronger cycles such as F- and W-cycles often do not improve convergence [34].

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 lose sharpness as the number of time grid levels increases, similarly to section 5.1.1, but still provide reasonable estimates on convergence. Indeed, for V-cycles with F-relaxation, Approximation 1 is quite sharp for all tested numbers of levels.

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

$$\partial_{tt} u = c^2 \nabla \cdot \nabla u \qquad \text{for } \mathbf{x} \in \Omega, t \in (0, 2\pi], \tag{5.2}$$

with scalar solution $u(\mathbf{x}, t)$ and wave speed $c = \sqrt{10}$. We transform (5.2) into a system of PDEs that are first order in time,

$$\partial_t u = v, \qquad \partial_t v = c^2 \nabla \cdot \nabla u, \qquad \text{for } \mathbf{x} \in \Omega, t \in (0, 2\pi], \tag{5.3}$$

with initial condition (see Figure SM1) and boundary conditions

$$u(\cdot, 0) = \sin(x)\sin(y), \quad v(\cdot, 0) = 0, \qquad \text{for } \mathbf{x} \in \Omega \cup \partial\Omega, \tag{5.4}$$

$$u(\mathbf{x}, \cdot) = v(\mathbf{x}, \cdot) = 0 \qquad \text{for } \mathbf{x} \in \partial\Omega. \tag{5.5}$$

This problem corresponds to a two-dimensional membrane with imposed nonzero 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]. Similarly to section 5.1, we use second-order centered finite differences to discretize the spatial operator in (5.3). The time-stepping operators Φ_ℓ are then simultaneously diagonalizable and the Courant number on each level is given by $\mathcal{C}_\ell = 2c\pi/[\delta_x(N_\ell - 1)]$, ranging between $\mathcal{C}_0 \approx 0.034$ on level 0 and $\mathcal{C}_5 \approx 1.087$ on level 5.[12]

An MGRIT V-cycle algorithm with FCF-relaxation shows quickly increasing convergence factors with a growing number of time grid levels (see Figure 4 and supplementary Figure SM6). 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 (see Figure 5 and supplementary Figure SM8) 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 trends. For example, convergence factors are initially constant in most cases, then increase almost linearly with the number of

[12] Note that the Courant number is smaller than in section 5.1.1. This is motivated by selecting a configuration that captures multiple periods of the oscillatory temporal behavior.

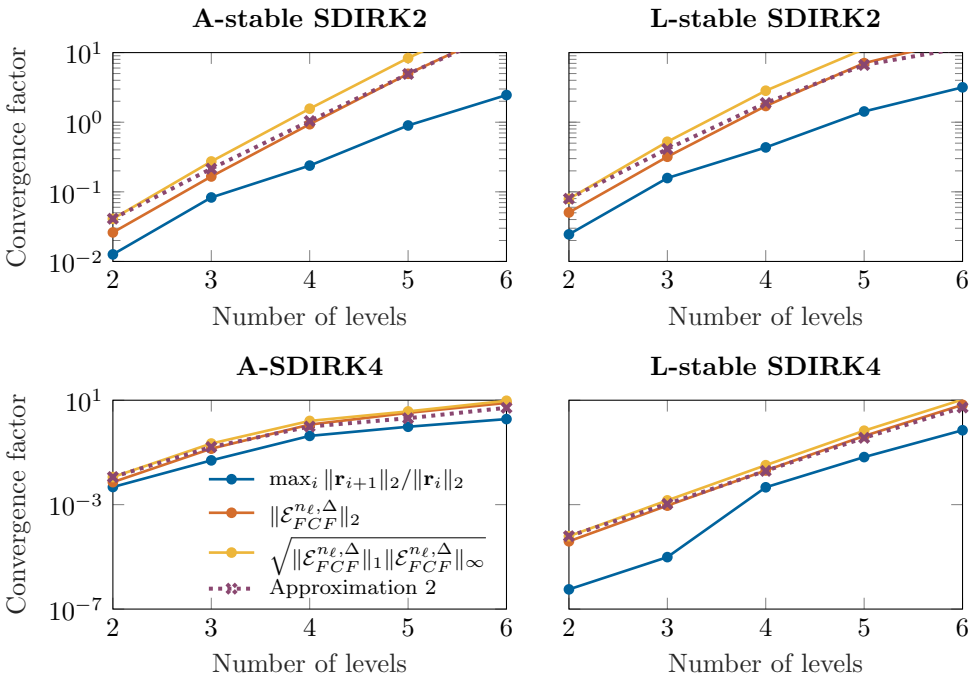

FIG. 4. *Wave equation: The convergence factor of MGRIT with V-cycles and FCF-relaxation increases substantially with a growing number of time grid levels and eventually exceeds* 1. *This means that MGRIT V-cycles will likely yield a divergent algorithm in practice, which is in line with observations for hyperbolic PDEs in the literature* [8, 9, 23].

levels, such as the case of L-stable SDIRK3 in supplementary Figure SM8. This highlights the fact that designing robust and convergent parallel-in-time algorithms for hyperbolic problems is generally perceived as difficult, and it 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 [12], such *a priori* estimates can provide valuable guidance.

Note that for the problem considered here, the benefit of FCF-relaxation over F-relaxation observed for the diffusion equation does not seem to apply for the wave equation (compare supplementary Figures SM5 and SM6, or supplementary Figures SM7 and SM8). 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 FCF-relaxation is not sufficient to design a scalable multilevel solver for the wave equation is a major difference from 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 using 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.

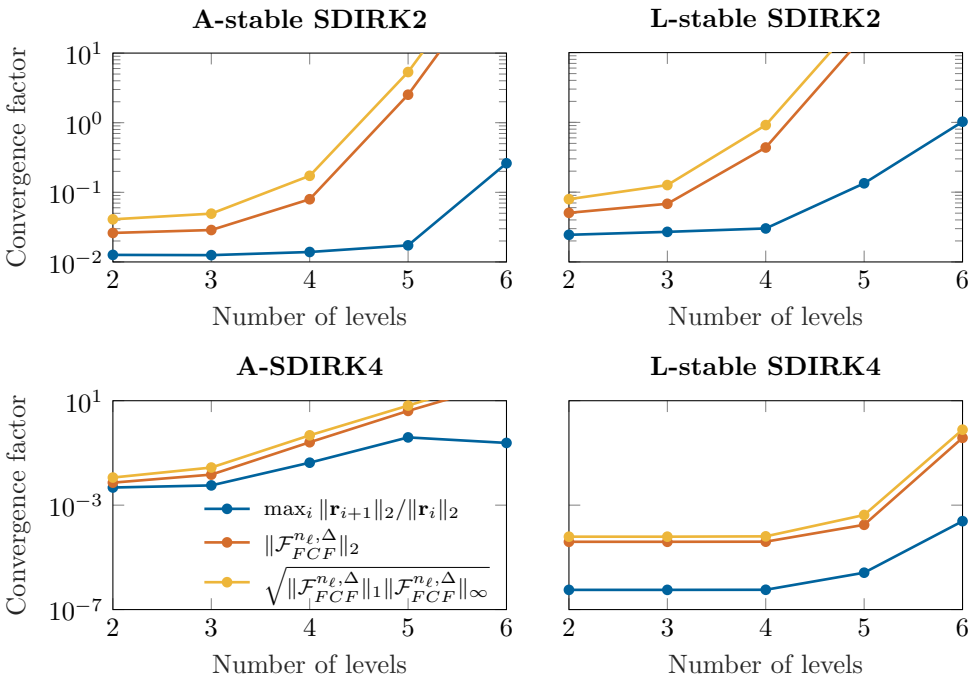

FIG. 5. *Wave equation: The convergence of MGRIT with F-cycles and FCF-relaxation deteriorates with a larger number of time grid levels compared to MGRIT with V-cycles; see Figure 4. Generally, convergent algorithms are given for a larger range of time grid levels, and observed convergence is better 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.*

6. Discussion. In section 5, we compared the *a priori* estimate for various developed bounds and approximate convergence factors with observed *worst-case* convergence in numerical experiments. These investigations were performed for a fixed temporal domain (in particular, a fixed fine-grid size) and a fixed choice of the spatial discretization scheme (e.g., the large spatial step size underresolves the discontinuous initial condition in section 5.1). An interesting future research question is how the convergence framework can help guide the selection of an optimal space-time discretization for the parallel-in-time integration with MGRIT with regards to best convergence and best speedup (e.g., by combining the convergence framework with performance models [12]). Here, important aspects are the dependency of the sharpness of the bounds on material parameters, spatial discretization and resolution, fine-grid size of the temporal domain, and others.

The presented upper bounds vary in their respective time complexity, and it is demonstrated that, e.g., using inequality (4.6) 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 (Approximations 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 yield *a priori* estimates that are at least as good as more expensive bounds. All these observations, however, rest on the assumption that the eigenvalues of the time-stepping operator Φ_ℓ can be computed, which might result in prohibitive computational cost for large-scale problems.

For Runge–Kutta schemes, the number of such operations can be reduced by only computing the eigenvalues of the spatial operator $\mathcal{L}$ and evaluating the stability function, as opposed to computing the eigenvalues for the family of Φ_ℓ for all considered Runge–Kutta schemes. While this might not be possible in general, the derivation of Fourier symbols [2] can provide another viable path to reduce the time complexity of computing such bounds. Furthermore, with prior knowledge of $\mathcal{L}$ (e.g., $\mathcal{L}$ is symmetric positive definite or skew symmetric), the need for solving an eigenvalue problem can be avoided. Despite the difficulty of performing multilevel convergence analysis for large-scale problems, there is yet a lot that can be learned from investigating smaller-scale problems.

MGRIT natively supports the parallel-in-time integration of nonlinear problems by using FAS multigrid (similar to other methods, e.g., [22]). The multilevel convergence framework developed in this work is limited to the linear case; however, it is able to illustrate the strengths and weaknesses of the algorithm in this setting (similar to the two-level theory developed in [3]). Future work will investigate how theoretical results for linear (or linearized) problems can guide the application of MGRIT for nonlinear problems. On the other hand, (two-level) convergence theory for Parareal [16] was extended to the nonlinear case [14] and, naturally, a similar extension of MGRIT convergence theory would be desirable.

Numerical studies demonstrate the benefit of using FCF-relaxation for parabolic model problems (see Figure 3), but FCF-relaxation does not significantly affect convergence of MGRIT for the hyperbolic model problem. However, theoretical results do confirm the advantage of using F-cycles over V-cycles for the hyperbolic model problem, a result that also applies to the isotropic diffusion equation if integrated by L-stable Runge–Kutta schemes. Results here confirm the observation that A-stable schemes are generally less suited for parallel-in-time integration than L-stable schemes [3]. For the diffusion model problem, 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 for A-stable schemes. However, results here indicate that stronger cycling and relaxation, such as F-cycles and FCF-relaxation, can alleviate this weakness.

7. Conclusion. In this work, we develop a framework for multilevel convergence of MGRIT for linear PDEs. 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 [3] and based on similar assumptions (for example, simultaneously diagonalizable and stable time-stepping operators). This work also presents a generalization of the two-level bounds derived in [3] to the case of arbitrary numbers of relaxation steps.

In complementary numerical studies, the theoretical results are assessed for two different model problems: the anisotropic diffusion equation and the second-order wave equation. It is 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.

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 and where *a priori* bounds can avoid (in parts) extensive numerical testing. Thus, future development and improvement of MGRIT (different coarse-grid operators [28], investigation of relation between errors and phase-shifts [36], and others) can be guided by the proposed multilevel convergence framework, where important recommendations can be made, such as the use of higher-order L-stable SDIRK methods and F-cycles with F-relaxation for hyperbolic problems.

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

REFERENCES

[1] A. J. Christlieb, C. B. Macdonald, and B. W. Ong, *Parallel high-order integrators*, SIAM J. Sci. Comput., 32 (2010), pp. 818–835, https://doi.org/10.1137/09075740X.

[2] H. De Sterck, S. Friedhoff, A. J. M. Howse, and S. P. MacLachlan, *Convergence Analysis for Parallel-in-Time Solution of Hyperbolic Systems*, preprint, https://arxiv.org/abs/1903.08928, 2019.

[3] V. A. Dobrev, T. Kolev, N. A. Petersson, and J. B. Schroder, *Two-level convergence theory for multigrid reduction in time (MGRIT)*, SIAM J. Sci. Comput., 39 (2017), pp. S501–S527, https://doi.org/10.1137/16M1074096.

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

[5] R. D. Falgout, S. Friedhoff, T. V. Kolev, S. P. MacLachlan, and J. B. Schroder, *Parallel time integration with multigrid*, SIAM J. Sci. Comput., 36 (2014), pp. C635–C661, https://doi.org/10.1137/130944230.

[6] R. D. Falgout, M. Lecouvez, and C. S. Woodward, *A parallel-in-time algorithm for variable step multistep methods*, J. Comput. Sci., 37 (2019), 101029.

[7] 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 J. Sci. Comput., 39 (2017), pp. S298–S322, https://doi.org/10.1137/16M1082330.

[8] C. Farhat and M. Chandesris, *Time-decomposed parallel time-integrators: Theory and feasibility studies for fluid, structure, and fluid-structure applications*, Internat. J. Numer. Methods Engrg., 58 (2003), pp. 1397–1434.

[9] C. Farhat, J. Cortial, C. Dastillung, and H. Bavestrello, *Time-parallel implicit integrators for the near-real-time prediction of linear structural dynamic responses*, Internat. J. Numer. Methods Engrg., 67 (2006), pp. 697–724.

[10] 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 Laboratory (LLNL), Livermore, CA, 2012.

[11] S. Friedhoff and S. MacLachlan, *A generalized predictive analysis tool for multigrid methods*, Numer. Linear Algebra Appl., 22 (2015), pp. 618–647.

[12] H. Gahvari, 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 International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), IEEE, 2016, pp. 22–31.

[13] M. J. Gander, 50 *years of time parallel time integration*, in Multiple Shooting and Time Domain Decomposition Methods, Springer, 2015, pp. 69–113.

[14] M. J. Gander and E. Hairer, *Nonlinear convergence analysis for the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering XVII, Springer, 2008, pp. 45–56.

[15] M. J. Gander, F. Kwok, and H. Zhang, *Multigrid interpretations of the parareal algorithm leading to an overlapping variant and MGRIT*, Comput. Vis. Sci., 19 (2018), pp. 59–74.

[16] M. J. Gander and S. Vandewalle, *On the superlinear and linear convergence of the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering XVI, Springer, 2007, pp. 291–298.

[17] S. Günther, N. R. Gauger, and J. B. Schroder, *A non-intrusive parallel-in-time adjoint solver with the XBraid library*, Comput. Vis. Sci., 19 (2018), pp. 85–95.

[18] S. Günther, N. R. Gauger, and J. B. Schroder, *A non-intrusive parallel-in-time approach for simultaneous optimization with unsteady PDEs*, Optim. Methods Softw., (2018), pp. 1–16.
[19] W. Hackbusch and U. Trottenberg, *Multigrid Methods: Proceedings of the Conference Held at Köln-Porz*, Lecture Notes in Math. 960, Springer, 1981.
[20] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I: *Non-stiff Problems*, Springer Ser. Comput. Math. 8, Springer, 1993.
[21] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations* II: *Stiff and Differential-Algebraic Problems*, Springer Ser. Comput. Math. 14, Springer, 1996.
[22] F. P. Hamon, M. Schreiber, and M. L. Minion, *Multi-level spectral deferred corrections scheme for the shallow water equations on the rotating sphere*, J. Comput. Phys., 376 (2019), pp. 435–454.
[23] A. Hessenthaler, D. Nordsletten, O. Röhrle, J. B. Schroder, and R. D. Falgout, *Convergence of the multigrid reduction in time algorithm for the linear elasticity equations*, Numer. Linear Algebra Appl., 25 (2018), e2155.
[24] N. J. Higham, *Estimating the matrix p-norm*, Numer. Math., 62 (1992), pp. 539–555.
[25] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 848–864, https://doi.org/10.1137/0916050.
[26] A. J. Howse, H. De Sterck, R. D. Falgout, S. MacLachlan, and J. Schroder, *Parallel-in-time multigrid with adaptive spatial coarsening for the linear advection and inviscid Burgers equations*, SIAM J. Sci. Comput., 41 (2019), pp. A538–A565, https://doi.org/10.1137/17M1144982.
[27] J. F. B. M. Kraaijevanger, *Contractivity of Runge-Kutta methods*, BIT, 31 (1991), pp. 482–528.
[28] O. A. Krzysik, H. De Sterck, S. P. MacLachlan, and S. Friedhoff, *On Selecting Coarse-Grid Operators for Parareal and MGRIT Applied to Linear Advection*, preprint, https://arxiv.org/abs/1902.07757, 2019.
[29] P. D. Lax and R. D. Richtmyer, *Survey of the stability of linear finite difference equations*, Comm. Pure Appl. Math., 9 (1956), pp. 267–293.
[30] 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), IEEE, 2016, pp. 1–5.
[31] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, 2007, https://doi.org/10.1137/1.9780898717839.
[32] J.-L. Lions, Y. Maday, and G. Turinici, *A "parareal" in time discretization of PDE's*, C. R. Acad. Sci. Paris Sér. I Math., 332 (2001), pp. 661–668.
[33] C. Lubich and A. Ostermann, *Multi-grid dynamic iteration for parabolic equations*, BIT, 27 (1987), pp. 216–234.
[34] T. A. Manteuffel, L. N. Olson, J. B. Schroder, and B. S. Southworth, *A root-node–based algebraic multigrid method*, SIAM J. Sci. Comput., 39 (2017), pp. S723–S756, https://doi.org/10.1137/16M1082706.
[35] J. Nievergelt, *Parallel methods for integrating ordinary differential equations*, Comm. ACM, 7 (1964), pp. 731–733.
[36] D. Ruprecht, *Wave propagation characteristics of Parareal*, Comput. Vis. Sci., 19 (2018), pp. 1–17.
[37] C. Sanderson and R. Curtin, *Armadillo: A template-based C++ library for linear algebra*, J. Open Source Software, 1 (2016), 26.
[38] C. Sanderson and R. Curtin, *A User-Friendly Hybrid Sparse Matrix Class in C++*, preprint, https://arxiv.org/abs/1805.03380, 2018.
[39] J. B. Schroder, *Parallelizing over Artificial Neural Network Training Runs with Multigrid*, preprint, https://arxiv.org/abs/1708.02276, 2017.
[40] J. B. Schroder, R. D. Falgout, C. S. Woodward, P. Top, and M. Lecouvez, *Parallel-in-time solution of power systems with scheduled events*, in 2018 Power and Energy Society General Meeting (PESGM), IEEE, 2018, pp. 1–5.
[41] B. S. Southworth, *Necessary conditions and tight two-level convergence bounds for Parareal and multigrid reduction in time*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 564–608, https://doi.org/10.1137/18M1226208.
[42] B. S. Southworth, W. Mitchell, and A. Hessenthaler, *Tight Two-level Convergence of Linear Parareal and MGRIT: Extensions and Implications in Practice*, in preparation.
[43] R. Speck, D. Ruprecht, M. Emmett, M. Minion, M. Bolten, and R. Krause, *A multi-level spectral deferred correction method*, BIT, 55 (2015), pp. 843–867.

[44] R. Speck, D. Ruprecht, R. Krause, M. Emmett, M. L. Minion, M. Winkel, and P. Gibbon, *A massively space-time parallel N-body solver*, in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC'12, IEEE Computer Society Press, 2012, pp. 92:1–92:11.

[45] S. Vandewalle and E. Van de Velde, *Space-time concurrent multigrid waveform relaxation*, Ann. Numer. Math., 1 (1994), pp. 347–360.

[46] *XBraid: Parallel Time Integration with Multigrid*, http://llnl.gov/casc/xbraid.

[47] X. Yue, S. Shu, X. Xu, W. Bu, and K. Pan, *Parallel-in-Time with Fully Finite Element Multigrid for* 2*-D Space-Fractional Diffusion Equations*, preprint, https://arxiv.org/abs/1805.06688, 2018.