TWO-LEVEL CONVERGENCE THEORY FOR MULTIGRID REDUCTION IN TIME (MGRIT) *

V.A. DOBREV , TZ. KOLEV , N.A. PETERSSON , AND J.B. SCHRODER [†]

Abstract. In this paper we develop a two-grid convergence theory for the parallel-in-time scheme known as multigrid reduction in time (MGRIT), as it is implemented in the open-source XBraid package [25]. MGRIT is a scalable and multi-level approach to parallel-in-time simulations that non-intrusively uses existing time-stepping schemes, and that in a specific two-level setting is equivalent to the widely-known parareal algorithm. The goal of this paper is two-fold. First, we present a two-level MGRIT convergence analysis for linear problems where the spatial discretization matrix can be diagonalized, and use this to analyze our two basic model problems, the heat equation and the advection equation. One important assumption is that the coarse and fine time-grid propagators can be diagaonalized by the same set of eigenvectors, which is often the case when the same spatial discretization operator is used on the coarse and fine time grids. In many cases, the MGRIT algorithm is guaranteed to converge and we demonstrate numerically that the theoretically predicted convergence rates are sharp in practice for our model problems. The second goal of the paper is to explore how the convergence of MGRIT compares to the stability of the chosen time-stepping scheme. In particular, we demonstrate that a stable time-stepping scheme does not necessarily imply convergence of MGRIT, although MGRIT with FCF-relaxation always converges for the diffusion dominated problems considered here.

Key words. multigrid, multigrid-in-time, parallel-in-time, convergence theory, high performance computing

AMS subject classifications. 65F10, 65M22, 65M55

1. Introduction. Research into parallel-in-time algorithms is currently being driven by the rapidly changing nature of computer architectures. Future speedups will come through an increasing numbers of cores, and not through faster clock-speeds. Previously, increasing clock-speeds compensated for decreasing time step sizes when spatial resolution was increased, leading to stable or even reduced overall runtimes. However, clock-speeds have become stagnant, or even decreased, leading to the sequential time integration bottleneck. For instance, this bottleneck occurs when the strong scaling limit in space is reached and no speedup can be achieved by distributing the fixed spatial problem on more cores. The bottleneck can also occur for codes that weakly scale with space-only parallelism. Here, successive spatial refinements with a fixed spatial problem size per core result in a (roughly) constant wall clock time per time step. However, for stability and/or accuracy reasons the number of time steps N_t must also increase, thus making the runtime proportional to N_t. The bottleneck can only be avoided by exploiting parallelism in the time dimension.

Efforts to develop parallel-in-time methods date back at least as far as 1964 [20], with the review paper [10] providing an excellent introduction to the field. One of the most popular algorithms is parareal [15]. Parareal has been extensively analyzed [16, 22, 4] and can be interpreted as a two-level multigrid in time method [12]. However, this implies that there is always a large coarse-level sequential solve, thus limiting parallelism to the finest level. The parallel spectral deferred correction (SDC) schemes

[†]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, `dobrev1,tzanio,petersson1,schroder2@llnl.gov`.

of Minion and Williams [19, 18] are well known and popular, but require the use of an SDC time propagator. There are also various multigrid methods that parallelize both space and time for parabolic [14, 24] as well as hyperbolic [5, 26] problems. To the authors' knowledge, with the exception of parareal and the PFASST library for parallel SDC methods [6], introducing any of the above parallel-in-time methods into an existing application code can be very intrusive and require a major code rewrite. This is because these methods assume space and/or time to be discretized in a specific way. The MGRIT method [7] was developed to provide a less intrusive approach to time-parallelism, which has been implemented in the open-source XBraid package [25]. MGRIT is fairly nonintrusive in that existing sequential time-stepping methods can be parallelized by wrapping the code according to the XBraid software interface. Furthermore, MGRIT is a full multi-level method in the time dimension, as opposed to parareal. [1]

In this paper we analyze the convergence properties of MGRIT in the two-level setting with two different relaxation schemes, known as F- and FCF-relaxation. Our analysis includes parareal as a special case because it is equivalent to MGRIT with F-relaxation [12]. In the following, parareal and MGRIT with F-relaxation refer to the same algorithm. Previous analyses of parareal include [12, 11, 1, 15, 23]. While these studies have added much to the understanding of parareal, our paper distinguishes itself in a few key ways. Most importantly, we incorporate the effects of FCF-relaxation into the analysis and describe how it improves convergence and stability. We develop a general convergence analysis that we evaluate numerically for two important matrix classes (symmetric and anti-symmetric), corresponding to important model problems (heat and advection), illustrating that our bounds are sharp. An additional novelty is that the convergence bound is a function of the temporal coarsening factor, which allows us to show that in some cases only certain coarsening factors yield a convergent method. For instance, we show the novel result of scalable two-level and nearly scalable multi-level [2] behavior for advective problems in certain cases for small coarsening factors. Lastly, we compare the standard stability regions for implicit Runge-Kutta schemes of orders 1 through 3 with the convergence regions of MGRIT with F- and FCF-relaxation.

In [23] a stability condition for parareal is derived and provides some insight into when parareal converges, particularly for the case of stiff ordinary differential equations (ODEs) and very small time steps. Our work is more general, as discussed above. In [11], a convergence bound for parareal is derived for nonlinear problems under fairly weak continuity assumptions. The bound, however, is not used as a predictive tool for convergence. The results explore empirically the behavior of parareal for some model nonlinear problems. The papers [1, 15] explore the order of parareal convergence, e.g., an order p time-stepping method converges like $\mathcal{O}(\delta_t^{p(k+1)})$ after k parareal iterations and time step size δ_t. However, these bounds are not sharp [12]. The work [1] additionally provides theoretical commentary on what conditions are required of the time-stepping method for parareal to converge. However, the authors do not extend these requirements to explore the relationship between parareal stability

---

[1]As an initial theoretical investigation of MGRIT, we only consider coarsening in time, but we note that simultaneous coarsening in space is also possible with XBraid and that this can create a method with classical multigrid O(N) performance for N time steps.

[2]We define scalable in the typical multigrid way to mean an iterative method that exhibits flat iteration counts, i.e., a bounded convergence rate, for a problem discretized over a sequence of increasingly fine space-time meshes. Nearly scalable behavior refers to a method that exhibits a slow growth in the convergence rate such that it is still a practical method.

and combinations of commonly used time-stepping routines, which is a goal here.

Equivalence was established in [12] between parareal and a two-grid multigrid cycle with F-relaxation. This paper additionally derives convergence bounds for parareal that are related to our results for MGRIT with F-relaxation. However some key differences remain. Our convergence bounds include the effects of FCF-relaxation and the temporal coarsening factor, and as such, we can do a more detailed convergence analysis of MGRIT with F- and FCF-relaxation. For instance, some of the bounds in [12] are very pessimistic when compared to results for smaller temporal coarsening factors. A final difference is that we explore in-depth how the stability of the time-stepper relates to the convergence of MGRIT.

The paper [9] uses a semi-algebraic mode analysis (SAMA) to analyze MGRIT convergence by diagonalizing each component of the two-grid error propagator with Fourier modes. The proposed approach applies the same analysis techniques used in [9], but does so by diagonalizing the time-stepper with any suitable basis such as eigenmodes, not Fourier modes. Given this closeness, there are nonetheless a few distinctions that make the proposed analysis novel. As opposed to analyzing the full two-grid MGRIT error propagator as in [9], we use the reduction aspect of MGRIT to derive a simplified error propagator that describes the entire method by only considering error propagation on the coarse grid. This simpler error propagation expression then allows us in some cases to derive convergence *bounds* true for any space-time mesh size, as opposed to the MGRIT convergence *estimates* in [9], which are based on sampling a fixed number of Fourier frequencies on a uniform mesh. Lastly, by focusing our analysis solely on MGRIT, we are able to explore MGRIT in more depth than in [9]. We examine MGRIT for both symmetric and anti-symmetric spatial discretizations, allowing us to consider advection, whereas [9] considered only MGRIT for the heat equation. We also provide theoretical insight into the insensitivity of MGRIT with FCF-relaxation to the temporal coarsening factor for parabolic problems [7]. Lastly, our paper explores MGRIT for a variety of time-stepping schemes, where [9] primarily considers the backward Euler method.

Overall, our analysis has the limitations that we consider only linear problems in the two-grid setting where the coarse and fine time steppers share the same eigenvectors. This is often the case when the same spatial discretization operator is used for both grids. However, the simple examples considered here vividly illustrate the strengths and weaknesses of the algorithm. We note that MGRIT can be trivially extended to the nonlinear setting with the full approximate storage (FAS) multigrid [2] as described in [8].

The rest of the paper is organized as follows. We begin Section 2 with an overview of the MGRIT algorithm. In Section 3.1, we derive the two-grid error propagator. In Section 3.2, we derive our theoretical convergence bound, followed by its application to ODE systems in Section 3.3. In Section 3.4, we explore the stability regions for implicit Runge-Kutta methods of order one through three, followed by consideration of the special cases of purely real and negative, and purely imaginary spatial eigenvalues. In Section 4 we give numerical results, first for parabolic and then hyperbolic problems, as examples of purely real and purely imaginary spatial eigenvalues, respectively. We then give special attention to the case of artificial dissipation which corresponds to complex-valued spatial eigenvalues. Conclusions are given in Section 5.
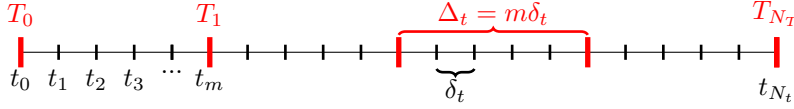
Fig. 2.1: Fine- and coarse-grid temporal meshes with coarsening factor m. F-points (black) are present only on the fine-grid, whereas C-points (red) are on both the fine- and coarse-grid.

2. The MGRIT algorithm.

The MGRIT algorithm is an iterative method for solving time-stepping problems,

$$
(2.1) \qquad
\begin{aligned}
\mathbf{u}_0 &= \mathbf{g}_0, \\
\mathbf{u}_j &= \Phi_j \mathbf{u}_{j-1} + \mathbf{g}_j, \quad j = 1, 2, \ldots, N_t.
\end{aligned}
$$

Here, the solution vector $\mathbf{u}_j \in \mathbb{R}^{N_x}$, where N_x refers to the number of points in space, and we consider the case where $\Phi_j = \Phi$ is a square matrix that does *not* depend on j. In many applications, $\mathbf{u}_j$ corresponds to an approximation of some time-dependent function $\mathbf{q}(t_j) : \mathbb{R} \to \mathbb{R}^{N_x}$, where time is discretized on a grid $t_j = j\delta_t$, $j = 0, 1, \ldots, N_t$, and $\delta_t = T/N_t > 0$ is the time step.

Sequential time stepping solves (2.1) sequentially by forward substitution with optimal complexity, $\mathcal{O}(N_t)$. MGRIT, instead, achieves parallelism in an iterative method that combines the original time-stepping problem with a coarser approximate representation. The method coarsens in time with an integer factor $m > 1$ yielding a coarse time grid of $N_T = N_t/m$ points (T_j) and a coarse time step $\Delta_t = m\delta_t$. This coarse time grid induces a partition of the fine grid into *C-points* (associated with coarse grid points) and *F-points*, as visualized in Figure 2.1.

By applying (2.1) recursively, we get

$$
(2.2) \quad \mathbf{u}_{km} = \Phi \mathbf{u}_{km-1} + \mathbf{g}_{km} = \Phi \left(\Phi \mathbf{u}_{km-2} + \mathbf{g}_{km-1} \right) + \mathbf{g}_{km} = \ldots
$$
$$
= \Phi^m \mathbf{u}_{(k-1)m} + \widetilde{\mathbf{g}}_{km}, \quad k = 1, 2, \ldots, N_T,
$$

where

$$
(2.3) \qquad \widetilde{\mathbf{g}}_{km} = \mathbf{g}_{km} + \Phi \mathbf{g}_{km-1} + \ldots + \Phi^{m-1} \mathbf{g}_{(k-1)m+1},
$$

Thus, all F-point values can be eliminated from the time-stepping problem, resulting in the equivalent coarse grid problem at all C-points

$$
(2.4) \qquad
\begin{aligned}
\mathbf{u}_0 &= \mathbf{g}_0, \\
\mathbf{u}_{km} &= \Phi^m \mathbf{u}_{(k-1)m} + \widetilde{\mathbf{g}}_{km}, \quad k = 1, 2, \ldots, N_T.
\end{aligned}
$$

Unfortunately, this problem is not any easier to solve than (2.1). Instead, MGRIT approximates the exact coarse grid time-stepping operator by introducing

$$
\Phi_\Delta \approx \Phi^m.
$$

A fundamental difference from sequential time-stepping, where the solution at one time point only depends on the solution at the previous time point, is that MGRIT

4

simultaneously computes the solution at all time points and needs to store the solution at all C-points.

Algorithm 1 presents the two-level MGRIT algorithm, which we proceed to explain in detail. To obtain a multi-level method, apply the algorithm recursively in Step 4. Note that the initial solution guess is passed through the variables $\mathbf{u}_j$, $j = 0, 1, \ldots, N_t$, which are overwritten by the solution upon convergence. To emphasize that $\mathbf{u}_j$ corresponds to a C-point for all $j = km$, $k = 0, 1, \ldots, N_T$, we write $\mathbf{u}_{km}^{[c]}$. The remaining indices correspond to F-points and are indicated by $\mathbf{u}_{km+q}^{[f]}$, $q = 1, 2, \ldots, m-1$. The notational convenience of emphasizing F- and C-points in this way is used only in this section and Section 3.1.

Algorithm 1 MGRIT(Φ, Φ_Δ, $\mathbf{u}$, $\mathbf{g}$)

 1: **repeat**
 2:　　Relax the approximate solution using Φ.
 3:　　Compute the residual on the coarse grid with Φ.
 4:　　Solve the coarse grid correction problem using Φ_Δ.
 5:　　Correct the approximate solution at the C-points.
 6:　　Update the solution at the F-points with Φ.
 7: **until** norm of residual is small enough

Relaxing the approximate solution. The first step of the algorithm relaxes (propagates) the approximate solution. There are two fundamental types of relaxation: F- and C-relaxation. The F-relaxation updates the F-point values based on the C-point values, with each updated F-interval of $m-1$ points independent of the other F-intervals,

$$
(2.5) \qquad
\left.
\begin{aligned}
\mathbf{u}_{km+1}^{[f]} &\leftarrow \Phi \mathbf{u}_{km}^{[c]} + \mathbf{g}_{km+1}, \\
\mathbf{u}_{km+2}^{[f]} &\leftarrow \Phi \mathbf{u}_{km+1}^{[f]} + \mathbf{g}_{km+2}, \\
&\vdots \\
\mathbf{u}_{(k+1)m-1}^{[f]} &\leftarrow \Phi \mathbf{u}_{(k+1)m-2}^{[f]} + \mathbf{g}_{(k+1)m-1},
\end{aligned}
\right\}
\qquad \text{for } k = 0, 1, \ldots, N_T - 1.
$$

C-relaxation does the opposite: it updates each C-point value based on the preceding F-point value, with each C-point update independent of the other C-points,

$$
(2.6) \qquad
\begin{aligned}
\mathbf{u}_0^{[c]} &\leftarrow \mathbf{g}_0, \\
\mathbf{u}_{km}^{[c]} &\leftarrow \Phi \mathbf{u}_{km-1}^{[f]} + \mathbf{g}_{km}, \quad \text{for } k = 1, \ldots, N_T.
\end{aligned}
$$

We remark that both F- and C-relaxation can be performed in parallel. The relaxation in step 2 of Algorithm 1 is either an F-relaxation or an FCF-relaxation. An F-relaxation is performed in both cases. In case of FCF-relaxation, the initial F-relaxation is followed by a C-relaxation and a second F-relaxation. The update sequence during FCF-relaxation is illustrated in Figure 2.2

Computing the residual on the coarse grid. The residual on the coarse grid is computed based on the relaxed approximate solution according to

$$
(2.7) \qquad
\begin{aligned}
\mathbf{r}_0^{[c]} &\leftarrow \mathbf{g}_0^{[c]} - \mathbf{u}_0^{[c]}, \\
\mathbf{r}_{km}^{[c]} &\leftarrow \mathbf{g}_{km}^{[c]} - \mathbf{u}_{km}^{[c]} + \Phi \mathbf{u}_{km-1}^{[f]}, \quad k = 1, 2, \ldots, N_T.
\end{aligned}
$$

The standard Euclidean norm of $\mathbf{r}^{[c]}$ is used in the halting criteria.
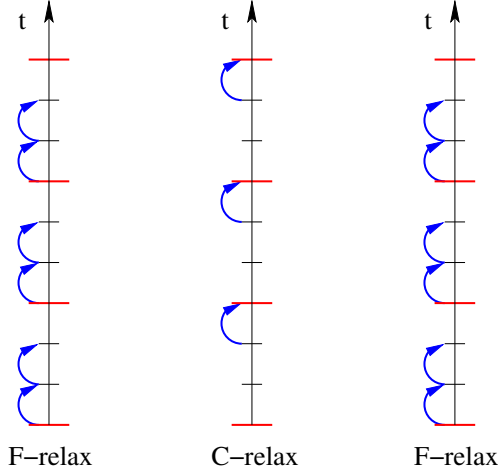
Fig. 2.2: The update sequence during FCF-relaxation. The red and black time points indicate C-points and F-points, respectively. Each blue arrow implies an application of Φ to the approximate solution at the preceding time point.

Solving the coarse grid correction problem using Φ_Δ. MGRIT defines an approximate coarse grid correction meant to approximate the error, which satisfies

$$
\begin{aligned}
\mathbf{c}_0^{[c]} &= \mathbf{r}_0^{[c]}, \\
\mathbf{c}_{km}^{[c]} &= \Phi_\Delta \mathbf{c}_{(k-1)m}^{[c]} + \mathbf{r}_{km}^{[c]}, \quad k = 1, 2, \ldots, N_T.
\end{aligned}
\tag{2.8}
$$

In this paper, we study the two-grid version of MGRIT, and solve this equation with forward substitution, which is a sequential $\mathcal{O}(N_T)$ operation. The recursive version of MGRIT uses a hierarchy of coarser grids to obtain the correction. It will not be described here; see [7].

Correcting the approximate solution on the coarse grid. The approximate coarse grid correction from the previous step is used to update the C-points of the approximate solution according to

$$
\mathbf{u}_{km}^{[c]} \leftarrow \mathbf{u}_{km}^{[c]} + \mathbf{c}_{km}^{[c]}, \quad k = 0, 1, \ldots, N_T.
\tag{2.9}
$$

Each correction can obviously be performed in parallel.

Updating the solution at the F-points with Φ. Given the approximate coarse grid solution values, the approximate solution at the F-points are computed by F-relaxation (2.5). However, this update is redundant because the same operation will be performed by the relaxation in the subsequent iteration. Hence, it is only necessary to update the solution at the F-points if the residual is smaller than the tolerance, in which case the MGRIT iteration is about to be terminated.

3. Convergence analysis of MGRIT.

3.1. Error propagation for one iteration of MGRIT. Let $\mathbf{v}_j$ be an approximate solution of (2.1), where the error is defined by $\mathbf{e}_j = \mathbf{u}_j - \mathbf{v}_j$. The residual on

the fine grid satisfies

$$\mathbf{r}_0 = \mathbf{g}_0 - \mathbf{v}_0,$$
$$\mathbf{r}_j = \mathbf{g}_j - \mathbf{v}_j + \Phi\mathbf{v}_{j-1}, \quad j = 1, 2, \dots, N_t.$$

We want to derive formulas for how the error evolves during one MGRIT iteration and start by considering the case of F-relaxation. By eliminating the intermediate F-point values in (2.5), we see that $\mathbf{v}^{[f]}_{(k+1)m-1}$ can be expressed in terms of $\mathbf{v}^{[c]}_{km}$ and the forcing $\mathbf{g}$,

$$(3.1) \qquad \begin{aligned} \mathbf{v}^{[f]}_{(k+1)m-1} &= \Phi^{m-1}\mathbf{v}^{[c]}_{km} + \widetilde{\mathbf{g}}_{(k+1)m-1}, \\ \widetilde{\mathbf{g}}_{(k+1)m-1} &= \mathbf{g}_{(k+1)m-1} + \Phi\mathbf{g}_{(k+1)m-2} + \dots + \Phi^{m-2}\mathbf{g}_{km+1}, \end{aligned}$$

where $k = 1, 2, \dots, N_T$. The formula for the residual at the C-points (2.7) can therefore be written

$$\mathbf{r}^{[c]}_{km} = \mathbf{g}_{km} - \mathbf{v}^{[c]}_{km} + \Phi\mathbf{v}^{[f]}_{km-1} = \widetilde{\mathbf{g}}_{km} - \mathbf{v}^{[c]}_{km} + \Phi^m\mathbf{v}^{[c]}_{(k-1)m},$$

where $\widetilde{\mathbf{g}}_{km}$ is defined by (2.3). Since the exact solution satisfies $\mathbf{u}_{km} = \Phi^m\mathbf{u}_{(k-1)m} + \widetilde{\mathbf{g}}_{km}$, the residual at the C-points can be expressed in terms of the error at the C-points,

$$(3.2) \qquad \mathbf{r}^{[c]}_0 = \mathbf{e}^{[c]}_0, \quad \mathbf{r}^{[c]}_{km} = \mathbf{e}^{[c]}_{km} - \Phi^m\mathbf{e}^{[c]}_{(k-1)m}, \quad k = 1, 2, \dots, N_T.$$

The coarse grid correction problem (2.8) is solved by forward substitution,

$$(3.3) \qquad \mathbf{c}^{[c]}_{km} = \Phi^k_\Delta \mathbf{r}^{[c]}_0 + \Phi^{k-1}_\Delta \mathbf{r}^{[c]}_m + \dots + \Phi_\Delta \mathbf{r}^{[c]}_{(k-1)m} + \mathbf{r}^{[c]}_{km}, \quad k = 1, 2, \dots, N_T.$$

We can use (3.2) to express the residuals in terms of the errors. We have $\mathbf{c}^{[c]}_0 = \mathbf{e}^{[c]}_0$ and

$$(3.4) \qquad \mathbf{c}^{[c]}_{km} = \Phi^{k-1}_\Delta \left(\Phi_\Delta - \Phi^m\right)\mathbf{e}^{[c]}_0 + \dots + \left(\Phi_\Delta - \Phi^m\right)\mathbf{e}^{[c]}_{(k-1)m} + \mathbf{e}^{[c]}_{km},$$

for $k = 1, 2, \dots, N_T$. The coarse grid correction (2.9) is used to update the approximate solution. Let the approximate solution after the correction be $\widehat{\mathbf{v}}_j = \mathbf{u}_j - \mathbf{f}_j$, where $\mathbf{f}_j$ is the error. After the correction, the error at the C-points satisfies

$$\mathbf{f}^{[c]}_{km} = \mathbf{e}^{[c]}_{km} - \mathbf{c}^{[c]}_{km}, \quad \text{for } k = 1, 2, \dots, N_T.$$

By inserting (3.4),

$$\mathbf{f}^{[c]}_{km} = \Phi^{k-1}_\Delta \left(\Phi^m - \Phi_\Delta\right)\mathbf{e}^{[c]}_0 + \Phi^{k-2}_\Delta \left(\Phi^m - \Phi_\Delta\right)\mathbf{e}^{[c]}_m + \dots + \left(\Phi^m - \Phi_\Delta\right)\mathbf{e}^{[c]}_{(k-1)m},$$

for $k = 1, 2, \dots, N_T$. We summarize the error propagation results in the following lemma.

LEMMA 3.1. *Let $\mathbf{u}_j$ be the solution of (2.1) and let $\mathbf{v}_j = \mathbf{u}_j - \mathbf{e}_j$ be an approximation of the solution, where $\mathbf{e}_j$ is the error and $j = 1, 2, \dots, N_t$. After one iteration of MGRIT with F-relaxation, the approximate solution satisfies $\widetilde{\mathbf{v}}_j = \mathbf{u}_j - \mathbf{f}_j$, where the error at the C-points satisfies*

$$(3.5) \qquad \mathbf{f}^{[c]}_0 = 0,$$

$$(3.6) \qquad \mathbf{f}^{[c]}_{km} = \sum_{q=0}^{k-1} \Phi^{k-1-q}_\Delta \left(\Phi^m - \Phi_\Delta\right)\mathbf{e}^{[c]}_{qm}, \quad k = 1, 2, \dots, N_T.$$

With FCF-relaxation, the approximate solution after F-relaxation is further modified by a C-relaxation and a second F-relaxation, as illustrated by Figure 2.2. As before, let $\mathbf{v}_j$ hold the result of the initial F-relaxation, governed by (3.1), and let $\mathbf{w}_j$ be the approximate solution after the C-relaxation, which is governed by (2.6). It satisfies

$$(3.7) \qquad \mathbf{w}_0^{[c]} = \mathbf{g}_0, \quad \mathbf{w}_{km}^{[c]} = \Phi^m \mathbf{v}_{(k-1)m}^{[c]} + \widetilde{\mathbf{g}}_{km}, \quad \text{for } k = 1, \ldots, N_T.$$

After the second F-relaxation, (3.1) gives

$$\mathbf{w}_{m-1}^{[f]} = \Phi^{m-1} \mathbf{g}_0 + \widetilde{\mathbf{g}}_{m-1},$$
$$\mathbf{w}_{(k+1)m-1}^{[f]} = \Phi^{m-1} \mathbf{w}_{km}^{[c]} + \widetilde{\mathbf{g}}_{(k+1)m-1} = \Phi^{2m-1} \mathbf{v}_{(k-1)m}^{[c]} + \Phi^{m-1} \widetilde{\mathbf{g}}_{km} + \widetilde{\mathbf{g}}_{(k+1)m-1},$$

for $k = 1, 2, \ldots, N_T - 1$. The residual on the coarse grid now becomes

$$\mathbf{r}_{km}^{[c]} = \mathbf{g}_{km} - \mathbf{w}_{km}^{[c]} + \Phi \mathbf{w}_{km-1}^{[f]}.$$

Using the same approach as for F-relaxation, the residual is first expressed in terms of the error, followed by solving the correction equation by forward substitution. The correction is added to the C-point values of the FCF-relaxed solution, giving the approximate solution $\widetilde{\mathbf{v}}_{km}^{[c]} = \mathbf{w}_{km}^{[c]} + \mathbf{c}_{km}^{[c]}$, with error $\mathbf{f}_{km}^{[c]} = \mathbf{u}_{km}^{[c]} - \widetilde{\mathbf{v}}_{km}^{[c]}$. Because $\mathbf{w}_{km}^{[c]}$ satisfies (3.7),

$$\mathbf{f}_0^{[c]} = 0, \quad \mathbf{f}_{km}^{[c]} = \Phi^m \mathbf{e}_{(k-1)m}^{[c]} - \mathbf{c}_{km}^{[c]}, \quad k = 1, 2, \ldots, N_T.$$

The error propagation formula with FCF-relaxation is obtained by substituting the solution of the correction equation into the above expression, as stated in the following lemma.

LEMMA 3.2. *Let $\mathbf{u}_j$ be the solution of (2.1) and let $\mathbf{v}_j = \mathbf{u}_j - \mathbf{e}_j$ be an approximation of the solution, where $\mathbf{e}_j$ is the error. After one iteration of MGRIT with FCF-relaxation, the approximate solution satisfies $\widetilde{\mathbf{v}}_j = \mathbf{u}_j - \mathbf{f}_j$, where the error at the C-points satisfies*

$$(3.8) \qquad \mathbf{f}_0^{[c]} = 0,$$

$$(3.9) \qquad \mathbf{f}_m^{[c]} = 0,$$

$$(3.10) \qquad \mathbf{f}_{km}^{[c]} = \sum_{q=0}^{k-2} \Phi_\Delta^{k-2-q} (\Phi^m - \Phi_\Delta) \Phi^m \mathbf{e}_{qm}^{[c]}, \quad k = 2, 3, \ldots, N_T.$$

Lemmas 3.1 and 3.2 may also be obtained from the Schur complement perspective, common to the sparse linear solver community. See Supplemental Materials Section S1 for more details.

3.2. Convergence estimates for MGRIT. In the following we assume that Φ and Φ_Δ can be diagonalized by the same unitary transformation,

$$(3.11) \qquad \widehat{\Phi} = X^* \Phi X = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{N_x}),$$

$$(3.12) \qquad \widehat{\Phi}_\Delta = X^* \Phi_\Delta X = \text{diag}(\mu_1, \mu_2, \ldots, \mu_{N_x}),$$

8

where

(3.13)
$$X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_x}), \quad X^* X = I.$$

To analyze the convergence properties of MGRIT, we make an eigenvector expansion of the error before and after an MGRIT iteration,

(3.14)
$$\mathbf{e}_j = \sum_{\omega=1}^{N_x} \hat{e}_{j,\omega} \mathbf{x}_\omega, \quad \mathbf{f}_j = \sum_{\omega=1}^{N_x} \hat{f}_{j,\omega} \mathbf{x}_\omega, \quad j = m, 2m, \ldots, mN_T.$$

We introduce the notation

$$\widehat{\mathbf{e}}_\omega = \begin{pmatrix} \hat{e}_{0,\omega} \\ \hat{e}_{m,\omega} \\ \vdots \\ \hat{e}_{mN_T,\omega} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_\omega^* \mathbf{e}_0 \\ \mathbf{x}_\omega^* \mathbf{e}_m \\ \vdots \\ \mathbf{x}_\omega^* \mathbf{e}_{mN_T} \end{pmatrix}, \quad \omega = 1, 2, \ldots, N_x,$$

and similarly for $\widehat{\mathbf{f}}_\omega$. Note that the first index on $\hat{e}$ corresponds to the time point, while the second index corresponds to the eigenmode. Using this notation and that $XX^* = I$, F-relaxation from Lemma 3.1 leads to the error propagation relation

(3.15)
$$\hat{f}_{0,\omega} = 0,$$
$$\hat{f}_{km,\omega} = \sum_{q=0}^{k-1} \mu_\omega^{k-1-q} \left(\lambda_\omega^m - \mu_\omega \right) \hat{e}_{qm,\omega}, \quad k = 1, 2, \ldots, N_T,$$

for $\omega = 1, 2, \ldots, N_x$. This relation can be written in matrix form as

(3.16) $\quad \widehat{\mathbf{f}}_\omega = E_\omega^F \widehat{\mathbf{e}}_\omega, \quad \text{where } E_\omega^F = (\lambda_\omega^m - \mu_\omega) \begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ \mu_\omega & 1 & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ \mu_\omega^{N_T-1} & \cdots & \mu_\omega & 1 & 0 \end{pmatrix}.$

This matrix represents the action of E_Δ^F on one term in the series (3.14), corresponding to the error propagation of eigenmode ω. For FCF-relaxation, the corresponding matrix follows from Lemma 3.2, where $\omega = 1, 2, \ldots, N_x$,

(3.17) $\quad E_\omega^{FCF} = (\lambda_\omega^m - \mu_\omega) \lambda_\omega^m \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ 1 & 0 & 0 & & & \\ \mu_\omega & 1 & 0 & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ \mu_\omega^{N_T-2} & \cdots & \mu_\omega & 1 & 0 & 0 \end{pmatrix}.$

It is straightforward to estimate matrix norms of E_ω^F and E_ω^{FCF} for all omega. Assuming that both Φ and Φ_Δ are stable time-stepping operators, i.e., $|\lambda_\omega| < 1$ and

$|\mu_\omega| < 1$,

$$(3.18) \qquad \|E_\omega^F\|_1 = \|E_\omega^F\|_\infty = |\lambda_\omega^m - \mu_\omega| \sum_{j=0}^{N_T-1} |\mu_\omega|^j = |\lambda_\omega^m - \mu_\omega| \frac{(1 - |\mu_\omega|^{N_T})}{(1 - |\mu_\omega|)},$$

and

$$(3.19) \qquad \|E_\omega^{FCF}\|_1 = \|E_\omega^{FCF}\|_\infty = |\lambda_\omega^m - \mu_\omega| \frac{(1 - |\mu_\omega|^{N_T-1})}{(1 - |\mu_\omega|)} |\lambda_\omega|^m.$$

Both error propagation matrices satisfy

$$(3.20) \qquad \|E_\omega\|_2 \le \sqrt{\|E_\omega\|_1 \|E_\omega\|_\infty} = \|E_\omega\|_1 = \|E_\omega\|_\infty.$$

We can now show the following.

THEOREM 3.3. *Let Φ and Φ_Δ be simultaneously diagonalizable by a unitary transformation X, with eigenvalues λ_ω and μ_ω, respectively. Furthermore, assume that both time-stepping operators are stable, i.e., $|\lambda_\omega| < 1$ and $|\mu_\omega| < 1$. Then the global space-time error vector at the C-points $\bar{\mathbf{e}} = [\mathbf{e}_0^T, \mathbf{e}_m^T, \mathbf{e}_{2m}^T, \ldots, \mathbf{e}_{N_T m}^T]^T$ satisfies*

$$\|E_\Delta^F \bar{\mathbf{e}}\|_2 \le \max_\omega \left\{ |\lambda_\omega^m - \mu_\omega| \frac{(1 - |\mu_\omega|^{N_T})}{(1 - |\mu_\omega|)} \right\} \|\bar{\mathbf{e}}\|_2, \quad \text{for F-relaxation},$$

and

$$\|E_\Delta^{FCF} \bar{\mathbf{e}}\|_2 \le \max_\omega \left\{ |\lambda_\omega^m - \mu_\omega| \frac{(1 - |\mu_\omega|^{N_T-1})}{(1 - |\mu_\omega|)} |\lambda_\omega|^m \right\} \|\bar{\mathbf{e}}\|_2, \quad \text{for FCF-relaxation}.$$

Proof. First, we use the orthonormality of the eigenvectors $\mathbf{x}_\omega$ to obtain the general relationship for any space-time vector $\bar{\mathbf{z}}$

$$(3.21) \qquad \|\bar{\mathbf{z}}\|_2^2 = \sum_{k=0}^{N_T} \|\bar{\mathbf{z}}_k\|_2^2 = \sum_{k=0}^{N_T} \sum_{\omega=1}^{N_x} |\hat{z}_{km,\omega}|^2 = \sum_{\omega=1}^{N_x} \|\hat{\mathbf{z}}_\omega\|_2^2.$$

By using (3.21), followed by (3.16) and (3.21) again, the bound for F-relaxation can be derived as

$$(3.22a) \qquad \|E_\Delta^F \bar{\mathbf{e}}\|_2^2 = \sum_{k=0}^{N_T} \|\mathbf{f}_k\|_2^2 = \sum_{\omega=1}^{N_x} \|\hat{\mathbf{f}}_\omega\|_2^2 = \sum_{\omega=1}^{N_x} \|E_\omega^F \hat{\mathbf{e}}_\omega\|_2^2$$

$$(3.22b) \qquad \le \left(\max_\omega \|E_\omega^F\|_2^2 \right) \sum_{\omega=1}^{N_x} \|\hat{\mathbf{e}}_\omega\|_2^2 = \left(\max_\omega \|E_\omega^F\|_2 \right)^2 \|\bar{\mathbf{e}}\|_2^2.$$

Finally, (3.20) and (3.18) leads to the desired results. The bound for FCF-relaxation is analogously derived. ☐

REMARK 3.1. *Theorem 3.3 implies that we can estimate the error reduction factor independently for each eigenmode, and allows us to simplify the convergence analysis by only examining $\|E_\omega\|_2$ for the worst-case eigenmode.*

REMARK 3.2. *Note that the limit $|\mu_\omega| \to 1$ is a removable singularity in the above estimates because if $|\mu_\omega| = 1 - \varepsilon$ and $0 < \varepsilon \ll 1$, then $1 - |\mu_\omega|^{N_T} = N_t \varepsilon + \mathcal{O}(\varepsilon^2)$ and $1 - |\mu_\omega| = \varepsilon$. Thus, $(1 - |\mu_\omega|^{N_T})/(1 - |\mu_\omega|) \to N_T$ as $\varepsilon \to 0+$. In many applications*

the convergence estimate can be bounded independently of N_T because of the factor $|\lambda_\omega - \mu_\omega^m|$ in the nominator, e.g. see Remark 3.5 below.

REMARK 3.3. *In our convergence estimates the nominator term $1 - |\mu_\omega|^{N_T}$ in Theorem 3.3 is usually replaced by unity. However, it is worth noting that the estimate holds for any value of N_T. This implies that the error at later time values is larger than the error at earlier time values, especially for modes with $|\mu_\omega| \approx 1$. Additionally, MGRIT can be expected to converge faster when N_T is smaller.*

REMARK 3.4. *As we will illustrate later for parabolic and mixed hyperbolic-parabolic model problems, the factor $|\lambda_\omega|^m$ in the estimate for FCF-relaxation leads to significantly faster convergence compared to F-relaxation, because $|\lambda_\omega| \ll 1$ for eigenmodes that are highly oscillatory in space. The FCF-relaxation was shown in [7] to be critical for scalable multi-level iterations.*

3.3. Systems of ODEs. The assumption that Φ and Φ_Δ can be diagonalized by the same transformation holds true, for example, when MGRIT is applied to solve a linear system of ODEs.

$$(3.23) \qquad \frac{d\mathbf{q}}{dt} = G\mathbf{q} + \mathbf{h}(t), \quad 0 \le t \le T,$$
$$\mathbf{q}(0) = \mathbf{g}_0,$$

and the linear operator G can be diagonalized by X,

$$(3.24) \qquad X^* G X = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_{N_x}).$$

We get a problem on the form (2.1) if we discretize (3.23) by a single step method and take $\mathbf{u}_j \approx \mathbf{q}(t_j)$. By assuming a constant time step, $t_j = j\delta_t$, $j = 0, 1, \ldots, N_t$, where $\delta_t N_t = T$, we can handle any single step time-stepping schemes where Φ can be written as a rational function of $\delta_t G$,

$$(3.25) \qquad \Phi = \left(I + \sum_\nu \beta_\nu (\delta_t G)^\nu \right)^{-1} \left(I + \sum_\nu \alpha_\nu (\delta_t G)^\nu \right).$$

The operator Φ_Δ can be diagonalized by the same transformation as Φ, for example, when it is a rediscretization of (3.23) using the time step $\Delta_t = m\delta_t$, and optionally a different time discretization scheme,

$$(3.26) \qquad \Phi_\Delta = \left(I + \sum_\nu \beta'_\nu (\Delta_t G)^\nu \right)^{-1} \left(I + \sum_\nu \alpha'_\nu (\Delta_t G)^\nu \right).$$

The eigenvalues of G are related to those of Φ and Φ_Δ according to
$$(3.27)$$
$$\lambda_\omega(\gamma_\omega) = \frac{1 + \sum_\nu \alpha_\nu (\delta_t \gamma_\omega)^\nu}{1 + \sum_\nu \beta_\nu (\delta_t \gamma_\omega)^\nu}, \quad \mu_\omega(\gamma_\omega) = \frac{1 + \sum_\nu \alpha'_\nu (\Delta_t \gamma_\omega)^\nu}{1 + \sum_\nu \beta'_\nu (\Delta_t \gamma_\omega)^\nu}, \quad \omega = 1, 2, \ldots, N_x.$$

It is worth exploring the upper bound in the estimates of Theorem 3.3 in certain model situations. For example, if we fix the final time and increase the temporal resolution, the time step goes to zero, $\delta_t \to 0$, leading to $|\mu_\omega| \to 1$. More importantly in this case, we also have $|\lambda_\omega| \to 1$. Therefore, in the small time-step limit, the convergence of MGRIT is determined by the ratio $|\lambda_\omega^m - \mu_\omega|/(1 - |\mu_\omega|)$ from Theorem 3.3. In many cases this is less than 1 (and independent of N_T), see e.g. Remark 3.5.

3.3.1. The influence of a mass matrix. So far, we have considered classical discretizations that do not include a mass matrix in the time stepping routine. However, we can generalize the analysis slightly and move to the case of simultaneously diagonalizable Φ and Φ_Δ, without the restriction of a unitary transformation. Let

(3.28a) $$\widehat{\Phi} = X^{-1}\Phi X = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{N_x}),$$

(3.28b) $$\widehat{\Phi}_\Delta = X^{-1}\Phi_\Delta X = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_{N_x}).$$

The difference now is that instead of (3.21), we have for any vector $\mathbf{z}$

(3.29) $$\|\mathbf{z}_i\|_\mathcal{M}^2 = \mathbf{z}_i^* \mathcal{M} \mathbf{z}_i = \|X^{-1}\mathbf{z}_i\|_2^2 = \sum_\omega |\hat{z}_{i,\omega}|^2, \qquad \text{with} \qquad \mathcal{M} = (XX^*)^{-1}.$$

The norm has changed. Introducing the $N_T \times N_T$ block diagonal matrix $\mathfrak{M}$ with diagonal blocks $\mathcal{M}$, we can write the F-relaxation estimate as

(3.30) $$\|E_\Delta^F \overline{\mathbf{e}}\|_{\mathfrak{M}} \leq \left(\max_\omega \|E_\omega^F\|_2 \right) \|\overline{\mathbf{e}}\|_{\mathfrak{M}} = \max_\omega \left\{ |\lambda_\omega^m - \mu_\omega| \frac{1 - |\mu_\omega|^{N_T}}{1 - |\mu_\omega|} \right\} \|\overline{\mathbf{e}}\|_{\mathfrak{M}}.$$

The FCF-relaxation estimate is analogous.

An example of the above case occurs when (3.23) is written in the form

$$M \frac{d\mathbf{q}}{dt} = -S\mathbf{q} + \mathbf{h}(t), \quad 0 \leq t \leq T,$$

corresponding to $G = -M^{-1}S$. Here $M = M^T > 0$ is the mass matrix and $S = S^T > 0$ is the stiffness matrix. Noting that the matrix

$$M^{\frac{1}{2}} G M^{-\frac{1}{2}} = M^{-\frac{1}{2}} S M^{-\frac{1}{2}}$$

is similar to G and symmetric, we can write its eigenvalue decomposition as

$$\left(M^{\frac{1}{2}} G M^{-\frac{1}{2}} \right) U = U\Gamma, \qquad \Gamma = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_{N_x}), \quad UU^T = I.$$

Now it is easy to see that the matrix $V = M^{-\frac{1}{2}} U$, is an eigenvector matrix for G and

$$\mathcal{M} = (VV^*)^{-1} = \left(M^{-\frac{1}{2}} UU^T M^{-\frac{1}{2}} \right)^{-1} = M.$$

As is often the case, the norm of M is close to 1, so we expect convergence here to be very similar to the unitarily diagonalizable case discussed above.

3.3.2. Influence of the spectrum of G. To predict how fast MGRIT will converge when it is applied to solve a system of ODEs of the form (3.23), it is important to consider how the spectrum of the operator G is related to the spectra of Φ and Φ_Δ, i.e., how the eigenvalues γ_ω are related to $(\lambda_\omega, \mu_\omega)$. This relation is determined by the fine and coarse time-stepping methods, through the coarsening factor $m = \Delta_t/\delta_t$ as well as the coefficients α_ν, β_ν, α'_ν, and β'_ν in (3.27). Theorem 3.3 proves that the convergence of MGRIT is governed by the pair of eigenvalues $(\lambda_\omega, \mu_\omega)$ that results in the largest value of $\|E_\omega\|_\infty$. This quantity can be estimated by (3.18) and (3.19) for F- and FCF-relaxation, respectively. Assuming that the same time-stepping method is used for both the fine and the coarse grid, the convergence of MGRIT is therefore determined by the coarsening factor m and the spectrum of the operator G in (3.23).

Classical discretizations of parabolic PDEs, e.g., finite element and finite difference, result in an operator G where the eigenvalues are real and non-positive, $\text{Im}\{\gamma_\omega\} = 0$, $-\gamma_{\max} \leq \text{Re}\{\gamma_\omega\} \leq 0$ and $\gamma_{\max} = \max|\gamma_\omega|$. As the spatial grid is refined, the eigenvalue with largest magnitude grows as $\gamma_{\max} = \mathcal{O}(1/h^2)$, where h represents the smallest spatial grid or element size. For a non-dissipative spatial discretization of a hyperbolic PDE, the eigenvalues are often purely imaginary and come in complex conjugated pairs, $\text{Re}\{\gamma_\omega\} = 0$ and $-\gamma_{\max} \leq \text{Im}\{\gamma_\omega\} \leq \gamma_{\max}$. In this case, the eigenvalue with largest magnitude grows as $\gamma_{\max} = \mathcal{O}(1/h)$ when the spatial grid is refined. For a dissipative spatial discretization of a hyperbolic PDE, the eigenvalues also come in complex conjugated pairs. Here the imaginary part of the spectrum behaves as in the non-dissipative case, but the real part is non-positive, $\text{Re}\{\gamma_\omega\} \leq 0$. Some examples of the spectrum of one-dimensional finite difference discretizations of hyperbolic and parabolic PDEs are given in Supplemental Materials Section S2. Explorations of these basic problem types, hyperbolic and parabolic, will be explored in the upcoming figures by examining the imaginary and real axes, respectively.

When the system of ODEs is stiff, for example when it corresponds to a spatial discretization of a parabolic partial differential equation, the eigenvalues of G are real and non-positive. Furthermore, the most negative eigenvalue γ_ω tends to $-\infty$ as the spatial grid size tends to zero. This means that

$$\lim_{\gamma_\omega \to -\infty} (\lambda_\omega(\gamma_\omega), \mu_\omega(\gamma_\omega)) = (0,0), \quad \text{for all L-stable time-stepping methods.}$$

In this limit, the estimates (3.18) and (3.19) show that the error after one iteration of MGRIT is zero, because the term $|\lambda^m - \mu| = 0$ for both F- and FCF-relaxation.

3.4. Implicit Runge-Kutta methods. Motivated by result regarding L-stable methods in Section 3.3.2, we proceed by evaluating the convergence estimate for three L-stable singly diagonally implicit Runge-Kutta (SDIRK) methods of accuracy order 1-3. The Butcher tableaux for these methods are given in Table 3.1. Note that the SDIRK-1 method is the same as the backward Euler method.

1	1
	1

(a) SDIRK-1

$1-\alpha$	$1-\alpha$	
α	$2\alpha-1$	$1-\alpha$
	$1/2$	$1/2$

(b) SDIRK-2, $\alpha = 1/\sqrt{2}$

a	a		
c	$c-a$	a	
1	b	$1-a-b$	a
	b	$1-a-b$	a

(c) SDIRK-3, $a = 0.435866\ldots$, $b = 1.208496\ldots$, $c = 0.717933\ldots$

Table 3.1: Butcher tableaux for L-stable SDIRK methods of accuracy order 1-3.

In Figure 3.1 we illustrate the convergence estimate for MGRIT for SDIRK methods of orders 1-3. In particular, we show contour level plots of $\|E_\omega\|_2$ as function of the real and imaginary parts of $\delta_t\gamma$. The contours are oriented with the regions of convergence and instability as depicted in Figure 3.1b. Note that the convergence factor is smaller for the third order than the second order SDIRK method, regardless the relaxation type. For all orders, $\|E_\omega\|_2$ is smaller with FCF-relaxation.

The second order accurate Crank-Nicolson method is often used for parabolic problems as it is unconditionally stable and only involves one stage per time step. It

(a) SDIRK-1 with F-relaxation

(b) SDIRK-1 with FCF-relaxation

(c) SDIRK-2 with F-relaxation

(d) SDIRK-2 with FCF-relaxation

(e) SDIRK-3 with F-relaxation

(f) SDIRK-3 with FCF-relaxation
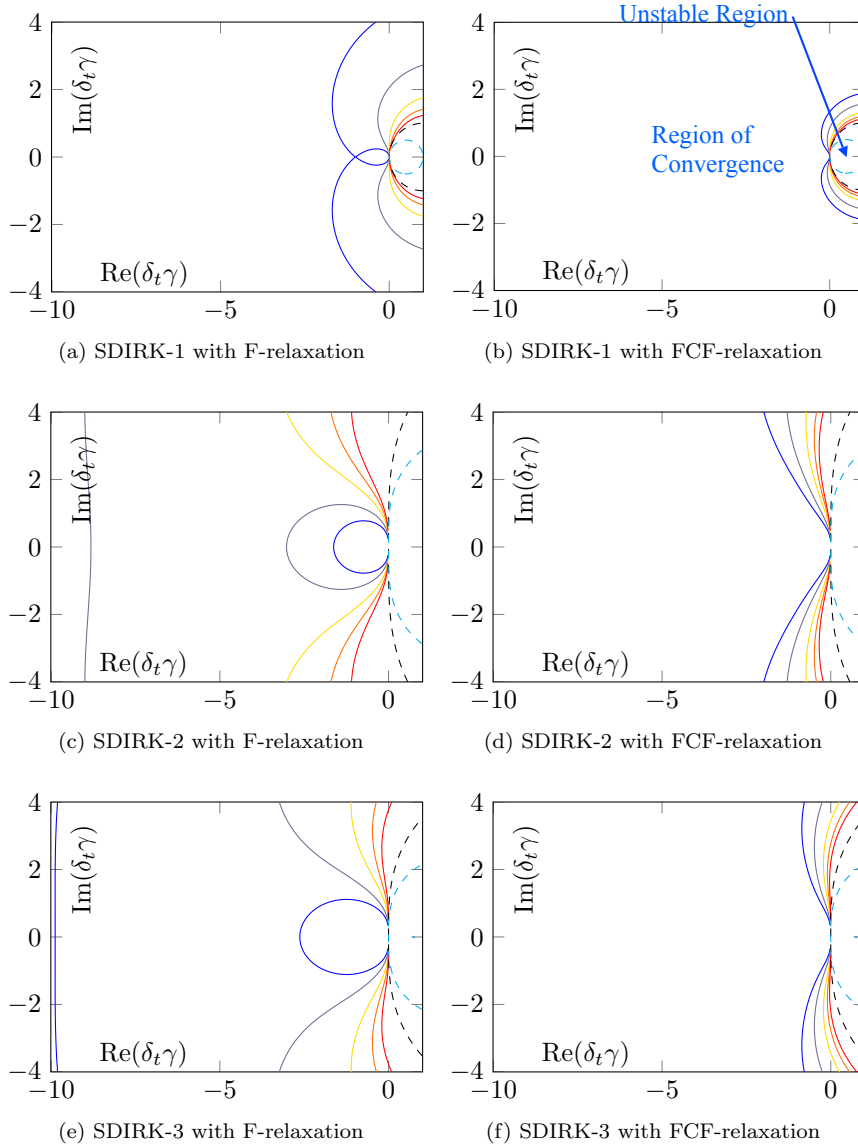
Fig. 3.1: $\|E_\omega\|_2$ as function of the real and imaginary parts of $\delta_t\gamma$ for SDIRK methods with coarsening factor $m = 2$. The solid lines show levels 0.125 (blue), 0.25 (gray), 0.5 (yellow), 0.75 (orange), and 1.0 (red). The dashed black and light blue lines indicate the stability boundaries $|\lambda| = 1$ and $|\mu| = 1$, respectively.

is an A-stable method, but it is not L-stable. However, the corresponding domain of convergence for MGRIT resembles the stability region of an explicit time-integration method, see Figure 3.2. This illustrates that the unconditional stability property of the underlying time-integration method can be lost when it is combined with MGRIT. The fact that Crank-Nicolson is only A-stable, but not L-stable, may be the root problem

for the restricted stability region of MGRIT. However, we leave this investigation for future work.

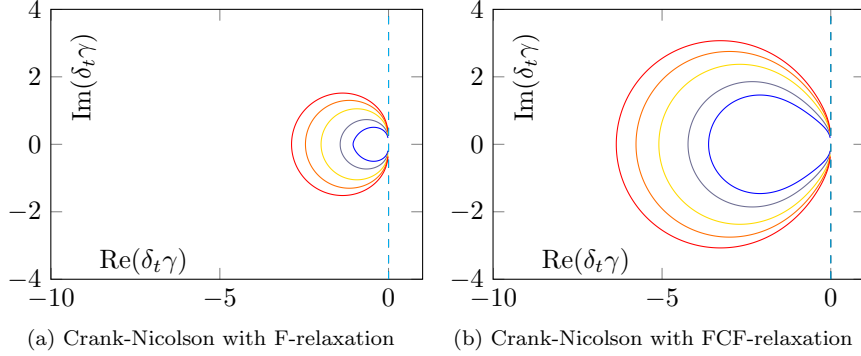(a) Crank-Nicolson with F-relaxation   (b) Crank-Nicolson with FCF-relaxation

Fig. 3.2: $\|E_\omega\|_2$ as function of the real and imaginary parts of $\delta_t\gamma$ for the second order Crank-Nicolson method with coarsening factor $m = 2$. The contour levels are the same as in Figure 3.1. The Crank-Nicolson method is stable for $\mathrm{Re}(\delta_t\gamma) \leq 0$, bounded by the dashed light blue line.

The coarsening factor, $m = \Delta_t/\delta_t$, has a significant influence over the convergence properties of MGRIT. As m becomes larger, say 16 or 32, the contour levels cluster near the boundary of the stability region for Φ, implying that MGRIT converges extremely well throughout almost all of the stability region of the SDIRK-3 method. For more details, see Figure S3.1 in Supplemental Materials. The only remaining region of slow convergence occurs near the imaginary axis, which will be studied in more detail below.

3.4.1. Convergence along the real and imaginary axes. In Figure 3.3 (top) we plot the convergence estimate for MGRIT and SDIRK methods of orders 1-3 when the spatial eigenvalue, γ_ω, is real and negative. The real case includes *any* spatial discretization that yields a symmetric negative definite matrix. We explore $m = 2$ and $m = 32$ and F- versus FCF-relaxation. The global convergence bound is the maximum value attained by a dataset and in all cases, this bound is uniformly below 1. The benefits of FCF-relaxation also increase with the SDIRK order. For SDIRK-3 and FCF-relaxation, the convergence bound is ≈ 0.005. Lastly, the results are insensitive to the coarsening factor m. Overall, the method performs very well for this case.

REMARK 3.5. *While we primarily use plots to illustrate convergence bounds because they convey more information, such as how each mode number responds to the algorithm, one can also derive a general convergence bound in many cases. For instance, let $m = 2$, γ_ω be real with $\gamma_\omega < 0$, and the backward Euler method be used on both levels, then the eigenvalues of Φ and Φ_Δ are a special case of (3.27)*

$$\lambda_\omega = (1 - \delta_t\gamma_\omega)^{-1}, \qquad \mu_\omega = (1 - 2\delta_t\gamma_\omega)^{-1}.$$

15

(a) $m = 2$, real axis

(b) $m = 32$, real axis

(c) $m = 2$, imaginary axis

(d) $m = 32$, imaginary axis
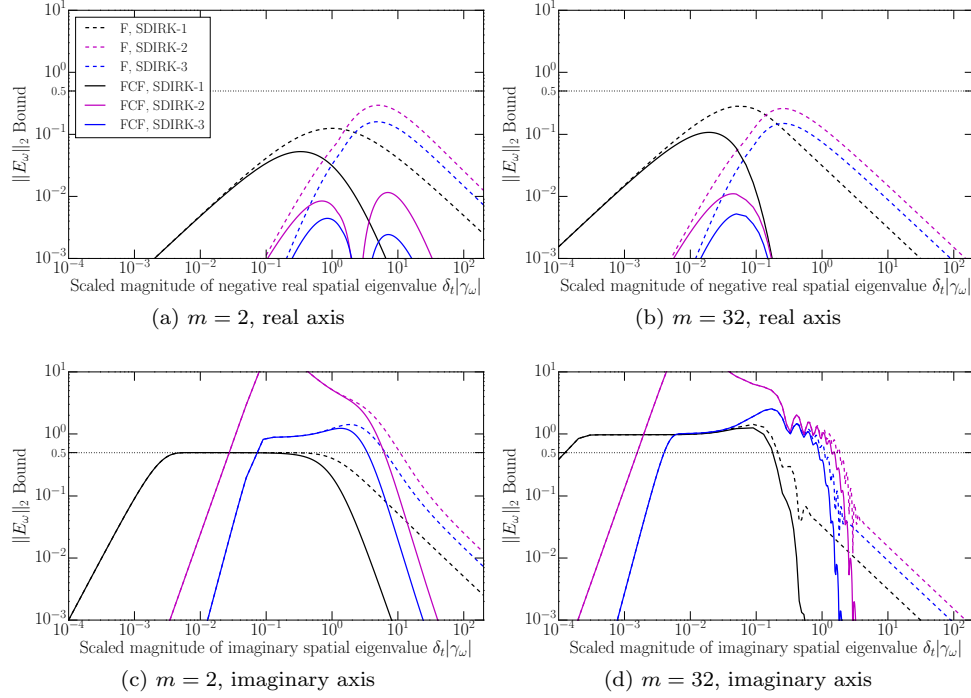
Fig. 3.3: $\|E_\omega\|_2$ as function of the magnitude of the scaled spatial eigenvalue $\delta_t|\gamma_\omega|$, for the SDIRK 1-3 methods, considering $m = 2$ and $m = 32$, for F- and FCF-relaxation.

In this case there is no restriction on δ_t and $\kappa := -\delta_t\gamma_\omega \in [0, \infty)$. Using (3.18) and taking the max over all κ, the estimate for $\|E_\omega^F\|_2$ is

$$\|E_\omega^F\|_2 \leq |(1 + \kappa)^{-2} - (1 + 2\kappa)^{-1}| \frac{(1 - (1 + 2\kappa)^{-N_T})}{(1 - (1 + 2\kappa)^{-1})} \leq |(1 + \kappa)^{-2}(1 + 2\kappa) - 1| \frac{1}{2\kappa}$$

$$= \frac{\kappa}{2(\kappa + 1)^2} \leq \frac{1}{8},$$

Thus, the MGRIT iteration is always convergent, with a rate bounded by 0.125. For FCF-relaxation, we can likewise show

$$\|E_\omega^{FCF}\|_2 \leq \frac{\kappa}{2(\kappa + 1)^4} \leq \frac{27}{512} = 0.0527\ldots$$

Note that these convergence bounds apply to any symmetric negative definite spatial discretization.

In Figure 3.3 (bottom) we plot the convergence estimate for MGRIT for SDIRK methods of orders 1-3, when γ_ω is purely imaginary. In this case, the sign of the imaginary part does not influence $\|E_\omega\|_2$, thus we consider only the magnitude of γ_ω. The convergence estimate is bounded uniformly below 1 only for SDIRK-1 and $m = 2$, where the bound is 0.5. For all other cases, the convergence bound eventually goes above 1 as the scaled spatial eigenvalue $\delta_t\gamma_\omega$ increases. For small enough δ_t, MGRIT can be made to converge for SDIRK-2 and SDIRK-3, but if $\gamma_\omega = \mathcal{O}(1/h)$, this leads to an explicit time-stepping limit $\delta_t/h \leq C$, which is a significant restriction.

16

4. Numerical results. We focus on systems of ODEs that arise from a method of lines discretization of a linear PDE. We will compare the convergence factor of running MGRIT with the bounds from Theorem 3.3. Since the fine-grid residual in MGRIT satisfies

$$\|\mathbf{r}\|_2 = \|AP\overline{\mathbf{e}}\|_2 = \|A_\Delta \overline{\mathbf{e}}\|_2,$$

the reported convergence factor (the ratio of two consecutive fine-grid residuals) is

$$\frac{\|\mathbf{r}_{n+1}\|_2}{\|\mathbf{r}_n\|_2} = \frac{\|A_\Delta \overline{\mathbf{e}}_{n+1}\|_2}{\|A_\Delta \overline{\mathbf{e}}_n\|_2} \leq \|A_\Delta E_\Delta A_\Delta^{-1}\|_2 = \|E_\Delta\|_2,$$

where $\|E_\Delta\|$ refers to both of the cases, F- and FCF-relaxation. The A_Δ and A_Δ^{-1} cancel because Φ and Φ_Δ commute (due to the fact that X is unitary [3]). Thus asymptotically, the estimates from Theorem 3.3 should be (close) upper bounds to the observed convergence factors in MGRIT.

Unless otherwise mentioned, we use a halting tolerance of 10^{-9}. The multi-level results use either stand-alone V-cycles or F-cycles, see [3]. The initial guess for the solver at all time points for $t > 0$ is uniformly random. The experimentally measured asymptotic convergence rate is taken to be the average convergence rate over the last five iterations. We note that the halting tolerance was chosen so that the method iterates long enough to exhibit asymptotic behavior. Overall, our experiments compare various coarsening factors, F-relaxation (i.e., parareal) versus FCF-relaxation and two-level versus multi-level.

4.1. Parabolic problems. We now consider the model heat equation

(4.1a) $$u_t - \nabla \cdot \nabla u = f(\mathbf{x}, t), \qquad (\mathbf{x}, t) \in \Omega \times (0, T),$$

(4.1b) $$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \qquad \mathbf{x} \in \Omega.$$

where Ω is some bounded convex domain in 1, 2 or 3 dimensions, and the boundary conditions are Dirichlet or Neumann in space. Standard discretizations of (4.1) lead to a system of ODEs of the form (3.23) where the linear operator G is symmetric with real and non-positive eigenvalues $\{\gamma_\omega\}, \gamma_\omega \leq 0$.

4.1.1. Heat equation with classic finite-differencing and backward Euler. We start by considering the heat equation (4.1) in 1D. We use standard second-order centered finite-differencing given by (S2.17) with homogeneous Dirichlet boundary conditions, and discretize time by the backward Euler method. [4] The space-time domain is a regular grid of $x \in [0, 1]$ and $t \in [0, 0.625]$, with the compatible initial condition of half a sine-wave, $u(x, 0) = \sin(\pi x)$. We take the ratio $\delta_t / h^2 = 10.0$.

Table 4.1 presents our numerical results. Overall, comparing the observed convergence factors with the theoretical estimates in the last column, we see that they are very close, confirming the accuracy and generality of our analysis. FCF-relaxation offers benefits over F-relaxation such as: insensitivity to the coarsening factor m for the larger problems, multi-level convergence that is essentially as good as two-level, a two-level analysis predictive of multi-level, and overall superior convergence rates.

Regarding F-relaxation as m increases, the two-level results deteriorate, but multi-level results improve. This stems from the fact that for large coarsening factors the

---

[3]This also holds true for the more general case when Φ and Φ_Δ are simultaneously diagonalizable by the same transformation.

[4]Note that backward Euler is the same as SDIRK-1.

difference between multi-level and two-level decreases, i.e., there are fewer levels. Lastly, the convergence rate for the multi-level method combined with F-relaxation always increases (i.e., is non-optimal) when the mesh is refined, but for large coarsening factors this growth is very slow.

Compared to the previously published convergence bound of 0.298 for two-level MGRIT with F-relaxation (c.f. Table 5.1 in [12]), we see that their estimate is only sharp for large coarsening factors. For small coarsening factors, it is overly pessimistic. Here, the new bound from Remark 3.5 is sharper, i.e., our bound is 0.125 for F-relaxation with $m = 2$ and the observed convergence rate is 0.111.

For the corresponding problem in 2D, we find that the one-dimensional results (theoretical and experimental) carry over, including Remark 3.5. See the Supplemental Materials Table S3.1.

	m / Size	$2^7 \times 2^{10}$	$2^8 \times 2^{12}$	$2^9 \times 2^{14}$	$2^{10} \times 2^{16}$	$\|E_\Delta\|_2 \leq$
	2	0.118	0.117	0.113	0.111	0.125
Two-level	4	0.197	0.197	0.196	0.194	0.204
	32	0.267	0.273	0.278	0.278	0.284
F-relaxation	2	0.468	0.527	0.560	0.573	-
V-cycles	4	0.327	0.380	0.398	0.405	-
	32	0.266	0.273	0.278	0.282	-
	2	0.046	0.045	0.044	0.042	0.0527
Two-level	4	0.076	0.075	0.075	0.074	0.0812
	32	0.053	0.095	0.101	0.100	0.108
FCF-relaxation	2	0.106	0.116	0.121	0.123	-
V-cycles	4	0.096	0.100	0.102	0.103	-
	32	0.053	0.095	0.100	0.100	-

Table 4.1: One-dimensional heat equation, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation. The theoretical bound appears in the final column.

4.1.2. Heat equation with finite elements and SDIRK time-steppers. In this section we demonstrate the generality of our convergence theory by considering the setting of 3D finite element problems posed on unstructured grids for various orders. These simulations use the modular finite element library MFEM [17]. We consider the heat equation (4.1) with the Neumann boundary condition,

$$\mathbf{n} \cdot \nabla u = g(\mathbf{x}, t), \quad (\mathbf{x}, t) \in \partial\Omega \times (0, T).$$

We take the computational domain Ω to be defined by the mesh `pipe-nurbs.mesh` from MFEM's `data` directory, see Figure S3.2 in Supplemental Materials. We set the final time to $T = 1$, and use the manufactured solution:

$$u(\mathbf{x}, t) = \sin(\kappa x_1) \sin(\kappa x_2) \sin(\kappa x_3) \sin(\tau t), \quad \mathbf{x} = (x_1, x_2, x_3),$$

with $\kappa = 0.314$ and $\tau = 2\pi + \pi/6$. In space, we apply a standard Galerkin finite element discretization using continuous Q_k elements, $k = 1, 2$, i.e. tri-linear and tri-quadratic polynomials on each element. For the time discretization, we use the SDIRK-2 and SDIRK-3 methods described above with Q_1 and Q_2 spaces, respectively, in order to match the spatial and temporal discretization orders. In this setting, the matrix $G = -M^{-1}S$, where M is the mass matrix and S is the diffusion matrix. It is well known, that the eigenvalues of G are real and non-positive; the smallest and largest (by magnitude) eigenvalues are of orders $\mathcal{O}(1)$ and $\mathcal{O}(h^{-2})$, respectively. The MGRIT convergence estimates for this case are shown in Figure 3.3 (top).

	m / Size	$2^6 \times 2^{10}$	$2^9 \times 2^{11}$	$2^{12} \times 2^{12}$	$\|E_\Delta\|_2 \leq$
	2	0.001	0.007	0.028	0.160
Two-level	4	0.008	0.032	0.090	0.150
	32	0.140	0.140	0.141	0.150
F-relaxation	2	0.201	0.214	0.216	-
V-cycles	4	0.134	0.158	0.156	-
	32	0.140	0.140	0.141	-
	2	0.001	0.002	0.004	0.004
Two-level	4	0.003	0.004	0.004	0.005
	32	0.004	0.004	0.004	0.005
FCF-relaxation	2	0.002	0.003	0.004	-
V-cycles	4	0.003	0.004	0.004	-
	32	0.004	0.004	0.004	-

Table 4.2: Three-dimensional heat equation with Q_2 elements and SDIRK-3, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation. The theoretical bound appears in the final column.

The results from the numerical experiments using Q_2 elements and SDIRK-3 are presented in Table 4.2, where the space-time resolutions are given in the form "number of mesh elements" × "number of time steps". See the Supplemental Materials Table S3.2 for results with Q_1 elements and SDIRK-2. Overall, the accuracy and generality of our analysis is again confirmed. The benefits of FCF-relaxation over F-relaxation are similar to those already discussed in Section 4.1.1, except that the convergence rate is now vastly improved for FCF-relaxation.

4.2. Mixed hyperbolic-parabolic problems. We now consider the model advection-diffusion problem

$$(4.2) \qquad u_t + \mathbf{b}(t, \mathbf{x}) \cdot \nabla u - \eta \nabla \cdot \nabla u = 0, \qquad (\mathbf{x}, t) \in \Omega \times (0, T),$$
$$(4.3) \qquad u(\mathbf{x}, 0) = u_0(\mathbf{x}), \qquad \mathbf{x} \in \Omega.$$

where $\eta \geq 0$, Ω is some bounded convex domain in 1, 2 or 3 dimensions, and the boundary conditions are either periodic or Dirichlet in space. The discretizations

19

considered will yield a linear operator G with either purely imaginary or complex-valued eigenvalues $\{\gamma_\omega\}$ with non-positive real part, complementing the study of the purely real case above.

For the highly advective problems considered here, V-cycles yield quickly rising iteration counts for sequences of refined grids, while F-cycles yield more stable iteration counts and are therefore used. In general, F-cycles are more expensive, but give a better approximation than V-cycles to the two-level method. However since this work considers a two-level analysis, we are unable to explore further the reason behind this advantage for F-cycles.

4.2.1. One-dimensional advection with grid-dependent dissipation. We start by discretizing (4.2) in one space dimension where $\Omega = [0,1]$, $\mathbf{b}(t,\mathbf{x}) = 1$, $\eta = 0$, and let the boundary conditions be periodic. When discretizing u_x, we use centered finite-differencing with artificial (grid-dependent) dissipation on a regular grid. This implies that the spatial eigenvalues $\{\gamma_\omega\}$ are the Fourier symbols from Supplemental Materials Section S2 and are a function of the scaled Fourier frequency $\xi = \omega h \in [-\pi, \pi]$ (see equation (S2.16)). We again consider the SDIRK methods from Table 3.1.

Figure 4.1 plots $\|E_\omega\|_2$ for FCF-relaxation, $m = 2, 32$ and $\delta_t/h = 1, 4$. The x-axis is the scaled Fourier frequency ξ (i.e., spatial eigenvalue). Using the notation from Supplemental Materials Section S2, the plots on the left are for SDIRK-1 with the low-order spatial discretization characterized by $\gamma_\omega^{\{2,\varepsilon\}} = \widehat{D}_1^{\{2,\varepsilon\}}$. The plots on the right are for SDIRK-3 and the higher-order spatial discretization characterized by $\gamma_\omega^{\{4,\varepsilon\}} = \widehat{D}_1^{\{4,\varepsilon\}}$. We consider the three cases $\varepsilon = 0, 0.5, 4$, corresponding to $D_1^{\{2p,\varepsilon\}}$ for $p = 1, 2$ (c.f. (S2.5)), where the artificial dissipation term is consistent with $-\varepsilon h^{2p-1}(-\Delta)^p$.

Repeating the experiment in Figure 4.1 for F-relaxation (see Supplemental Materials Figure S3.3) shows a significant difference only for the case of $\varepsilon = 4$, where the higher viscosity allows FCF-relaxation to accelerate the method. Overall for F- and FCF-relaxation, the method converges with a rate at, or below 0.5, for smaller m and $\delta_t/h \leq 1$. If m or δ_t/h is large, then divergence is likely. FCF-relaxation is beneficial at higher frequencies, and this can lead to a significantly better overall bound, as in the SDIRK-3 and $\varepsilon = 4$ case.

To numerically validate the lower-order case of Figure 4.1 (left-side) for $m = 2$, we run MGRIT and record the asymptotic convergence rate. We take $N_t = N_x$, corresponding to the final time $T = \delta_t N_t = \delta_t/h$. In Table 4.3 we report the case $\delta_t/h = 1$. The initial condition is $u(x,0) = \sin(2\pi x)$ on $[0, 0.5]$ and $u(x,0) = 0$ on $(0.5, 1.0]$. The MGRIT halting tolerance was reduced to $\frac{10^{-14}}{\sqrt{\delta_t h}}$ to better capture the asymptotic behavior.

For $m = 2$, the results agree with the theoretical bound in the final column. The addition of dissipation (larger ε) leads to faster convergence, while larger δ_t/h lead to slower convergence. See Table S3.3 in the supplemental materials section for the case $\delta_t/h = 4$. For the two-level case, FCF-relaxation is only beneficial when there is a sufficient amount of dissipation, i.e. $\varepsilon = 4.0$. The chief benefit of FCF-relaxation is for the multi-level case where FCF is required for good MGRIT performance (convergence generally at or below 0.5). However even with FCF-relaxation, there appears to be a modest log-growth in the iteration count.

The results for $m = 32$ are not presented because they generally diverged. The chief exception was for FCF-relaxation and $\varepsilon = 4.0$, where as illustrated by Figure
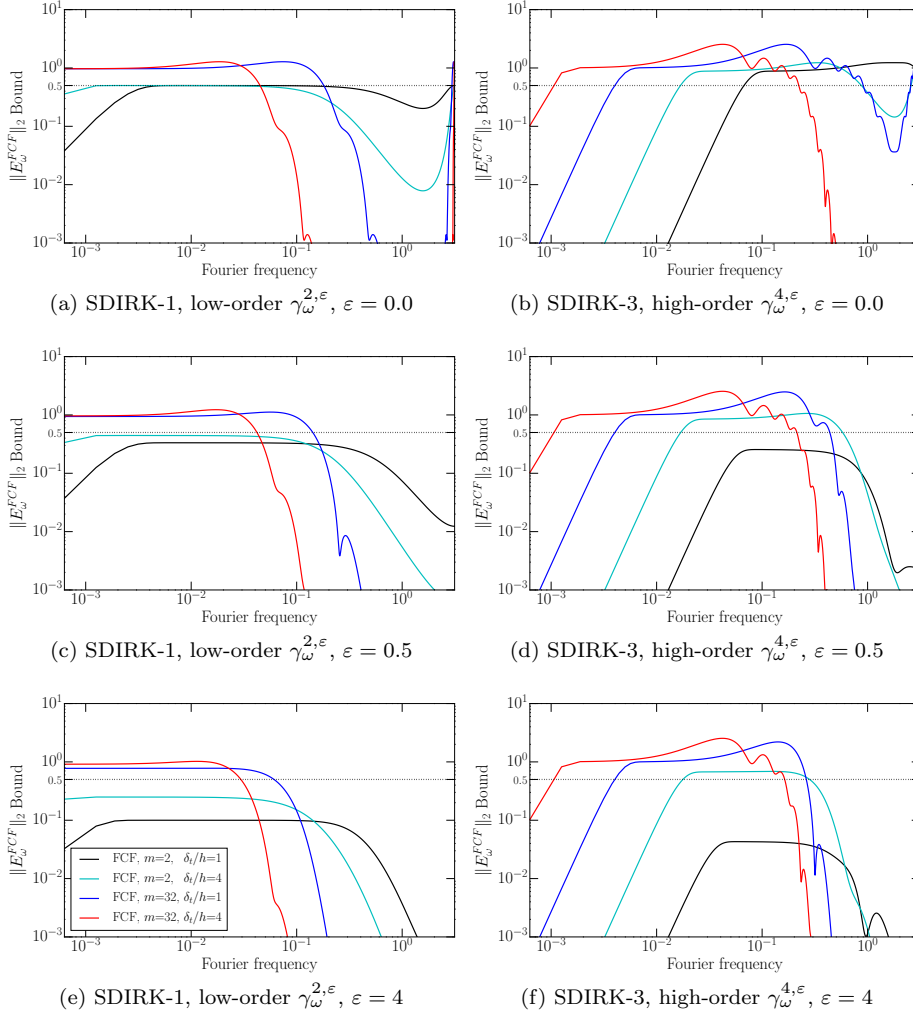
(a) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 0.0$ (b) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 0.0$

(c) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 0.5$ (d) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 0.5$

(e) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 4$ (f) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 4$

Fig. 4.1: $\|E_\omega^{FCF}\|_2$ as function of the scaled Fourier frequency $\xi = \omega h$ for $\xi \in [0, \pi]$, $m = 2, 32$ and $\delta_t/h = 1, 4$. Note that $\|E_\omega^{FCF}\|_2$ is a symmetric function of ξ and reflects around the origin.

4.1, MGRIT converged slowly with rates between 0.6 and 0.9.

4.2.2. One-dimensional advection-diffusion. We now repeat the analysis from Section 4.2.1, but for the case of advection with physical (non grid-dependent) diffusion. To do this, let $\varepsilon = 0$ and $\eta = 1$, which implies spatial eigenvalues of $\{\gamma_\omega\} = \{\widehat{D}_1^{\{2,\varepsilon=0\}} + \widehat{D}_2^{\{2\}}\}$ for the low-order case and $\{\gamma_\omega\} = \{\widehat{D}_1^{\{4,\varepsilon=0\}} + \widehat{D}_2^{\{4\}}\}$ for the high-order case (see Appendix A for notation). Here, the eigenvalues $\{\gamma_\omega\}$ are complex-valued but are dominated by their real part in the limit $h \to 0$.

The convergence bounds are depicted in Figure 4.2. As h decreases, the contribution from the dissipative Fourier symbol $\widehat{D}_2^{\{p\}}$ grows (i.e., the negative real part of $\{\gamma_\omega\}$ grows). The convergence bound improves and approaches the diffusion-only

	ε / Size	$2^8 \times 2^8$	$2^9 \times 2^9$	$2^{10} \times 2^{10}$	$2^{11} \times 2^{11}$	$\|E_\Delta\|_2 \leq$
	0	0.449	0.455	0.456	0.459	0.500
Two-level	0.5	0.288	0.294	0.298	0.300	0.333
F-relaxation	4.0	0.133	0.134	0.134	0.134	0.140
	0	0.537	0.647	0.774	0.872	-
F-cycles	0.5	0.230	0.311	0.399	0.506	-
	4.0	0.138	0.138	0.163	0.198	-
	0	0.275	0.398	0.443	0.461	0.500
Two-level	0.5	0.213	0.264	0.288	0.297	0.333
FCF-relaxation	4.0	0.083	0.086	0.087	0.087	0.100
	0	0.260	0.367	0.441	0.524	-
F-cycles	0.5	0.204	0.249	0.327	0.396	-
	4.0	0.088	0.100	0.127	0.150	-

Table 4.3: One-dimensional advection equation with artificial dissipation, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation; $m = 2$, $N_t = N_x$, $\delta_t/h = 1$, for various ε. The theoretical bound appears in the final column.

case. For $m = 2$ diffusion-only like behavior is quickly attained, however the grids considered for $m = 32$ are not large enough to show such asymptotic behavior.

FCF-relaxation offers similar benefits over F-relaxation to that seen for the heat equation. Regarding other parameters, increasing δ_t/h will worsen the MGRIT convergence, while decreasing δ_t/h will improve it. Increasing η will also improve the convergence.

Table 4.4 validates numerically the lower-order plots (left-side) of Figure 4.2, analogously to Section 4.2.1. For $m = 2$, the results match well the one-dimensional heat equation results from Table 4.1, while for $m = 32$, the grid-sizes are not quite large enough to show asymptotic behavior similar to the heat equation. A final column showing $\|E_\Delta\|_2 \leq$ is not practical because each h-value leads to a different convergence bound.

4.2.3. Advection-diffusion with finite elements and SDIRK time stepping. To show generality for our analysis, we now consider a 2D discontinuous Galerkin (DG) discretization, using the MFEM library, of the advection-diffusion problem (4.2) with periodic boundary conditions over a regular hexagonal domain Ω, specifically, the file `data/periodic-hexagon.mesh` from MFEM. We set the final time $T = 10$, the velocity field $\mathbf{b} = (\sqrt{2/3}, \sqrt{1/3})$ and let the initial condition be a

(a) SDIRK-1, low-order γ_ω, F-relaxation

(b) SDIRK-3, high-order γ_ω, F-relaxation

(c) SDIRK-1, low-order γ_ω, FCF-relaxation

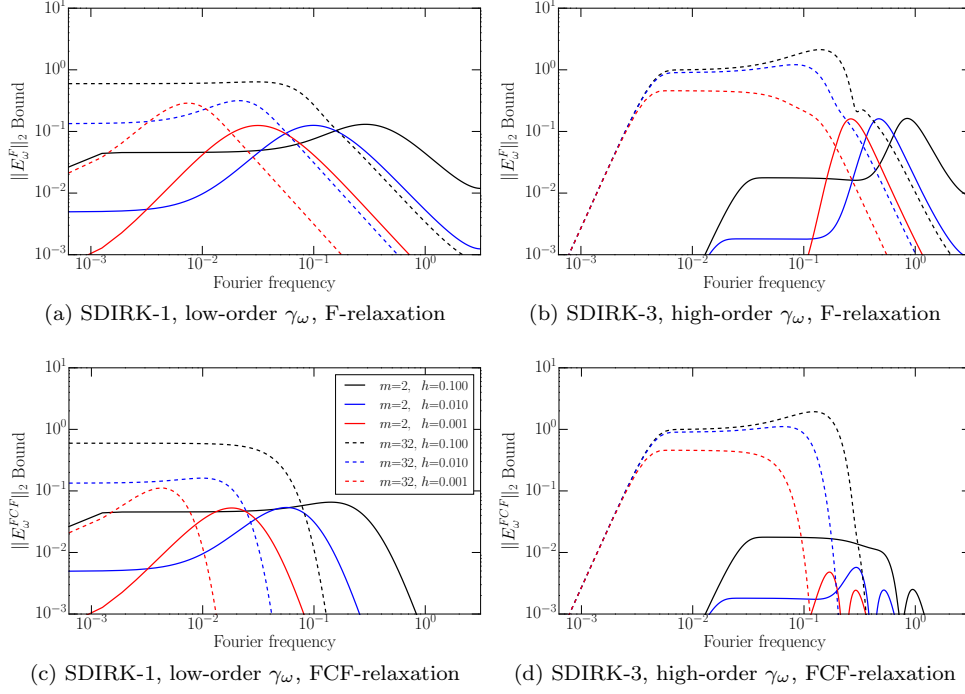(d) SDIRK-3, high-order γ_ω, FCF-relaxation

Fig. 4.2: $\|E_\omega\|_2$ as function of the scaled Fourier frequency $\xi = \omega h$ for $\xi \in [0, \pi]$, $m = 2, 32$, $\delta_t/h = 1$ and physical diffusion $\eta = 1$, for different h values. Note that $\|E_\omega\|_2$ is a symmetric function of ξ and reflects around the origin.

smooth rectangular bump:

$$u_0(\mathbf{x}) = \frac{1}{16}\operatorname{erfc}[w(x_1 - c_1 - r_1)]\operatorname{erfc}[-w(x_1 - c_1 + r_1)]$$

$$\times \operatorname{erfc}[w(x_2 - c_2 - r_2)]\operatorname{erfc}[-w(x_2 - c_2 + r_2)], \quad \mathbf{x} = (x_1, x_2),$$

where $\operatorname{erfc}(x)$ is the complementary error function, and the parameters are: $(c_1, c_2) = (0, -0.2)$, $(r_1, r_2) = (0.45, 0.25)$, and $w = 10$. A sample mesh and initial condition are depicted in Figure S3.4 in Supplemental Materials. For the spatial discretization, we use Q_3 (bi-cubic) discontinuous elements with a standard upwind DG discretization for the advection operator and the symmetric interior penalty method (IP or SIPG) for the diffusion operator, when $\eta > 0$. The semi-discrete problem is:

$$(4.4) \qquad M\dot{\mathbf{u}}_h = K\mathbf{u}_h - \eta S\mathbf{u}_h, \quad \eta \geq 0,$$

where M is the DG mass matrix, K, S are the discretizations of $-\mathbf{v} \cdot \nabla u$ and $-\Delta u$, respectively, and $G = M^{-1}(K - \eta S)$. For the temporal discretization we use the SDIRK-3 method defined earlier. We consider three cases:

(4.5a) $\qquad \eta = 0$, (pure advection),

(4.5b) $\qquad \eta = 0.1$, (advection-diffusion with constant diffusion),

(4.5c) $\qquad \eta = 0$, (pure advection with artificial dissipation).

23

	δ_t/h	m / Size	$2^8 \times 2^8$	$2^9 \times 2^9$	$2^{10} \times 2^{10}$	$2^{11} \times 2^{11}$
Two-level F-relaxation	1	2	0.116	0.118	0.118	0.118
	4	2	0.111	0.120	0.118	0.119
	1	32	N/A	N/A	0.181	0.259
	4	32	N/A	N/A	0.181	0.251
F-cycles	1	2	0.113	0.114	0.114	0.115
	4	2	0.108	0.117	0.114	0.115
Two-level FCF-relaxation	1	2	0.042	0.048	0.049	0.048
	4	2	0.042	0.051	0.043	0.049
	1	32	N/A	N/A	0.057	0.091
	4	32	N/A	N/A	0.001	0.023
F-cycles	1	2	0.042	0.047	0.048	0.048
	4	2	0.042	0.050	0.043	0.048

Table 4.4: One-dimensional advection-diffusion equation with physical diffusion coefficient $\eta = 1$ in (4.2), asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation, for different coarsening factors m and δ_t/h.

For the case (4.5c), the term

$$(4.6) \qquad -\varepsilon h^3 S M^{-1} S = -10^{-3}(h/h_0)^3 S M^{-1} S,$$

is added to the right-hand side of the ODE (4.4). The values h and h_0 are the mesh spacings for the current mesh and the smallest (coarsest) mesh, respectively. In other words on the coarsest mesh we use 10^{-3} dissipation and as the mesh is refined by a factor of 2, we decrease the dissipation by a factor of 8. This scaling of ε is chosen to yield a dissipation similar to the one-dimensional case with high-order dissipation. Note that the inverse M^{-1} is computable because it is block-diagonal with blocks corresponding to the mesh elements.

Since some of cases (4.5a)-(4.5c) make MGRIT diverge, we cap the number of iterations at 30, and only generate results at larger grid sizes for the cases that do converge. If no results were generated, a '-' is used. For this set of tests, the problem size is given as "number of degrees of freedom" $\times$ "number of time steps".

Our theory requires diagonalizability of G, which we cannot guarantee in all of the cases. For the pure advection case (4.5a), we did numerically verify for all three grid sizes that the eigenvalues of G are separated by a distance bounded from below by $\approx 5 \times 10^{-5}$. Thus, we only claim that our theory applies for the pure advection case. Nonetheless we find it worthwhile to investigate if the theory also has predictive properties for the cases (4.5b) and (4.5c) (which it does).

The results for the pure advection case are presented in Table 4.5. Here MGRIT only converged for the two-level method and $m = 2, 4$; however, the convergence rate deteriorated as the grid was refined. MGRIT diverged (denoted by "div") for all other parameter combinations. We conjecture that the small numerical diffusion added by

m / Size	$3 \cdot 2^8 \times 2^{10}$	$3 \cdot 2^{10} \times 2^{11}$	$3 \cdot 2^{12} \times 2^{12}$	$3 \cdot 2^{14} \times 2^{13}$
Two-level	2	0.069	0.098	0.127	0.154
F-relaxation	4	0.338	0.479	0.584	0.682
	32	div	div	-	-
F-cycles	2,4,32	div	div	-	-
Two-level	2	0.054	0.087	0.123	0.159
FCF-relaxation	4	0.289	0.392	0.523	0.648
	32	div	div	-	-
F-cycles	2,4,32	div	div	-	-

Table 4.5: Asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation, for the case of pure advection with $\eta = 0$, Q_3 elements and SDIRK-3.
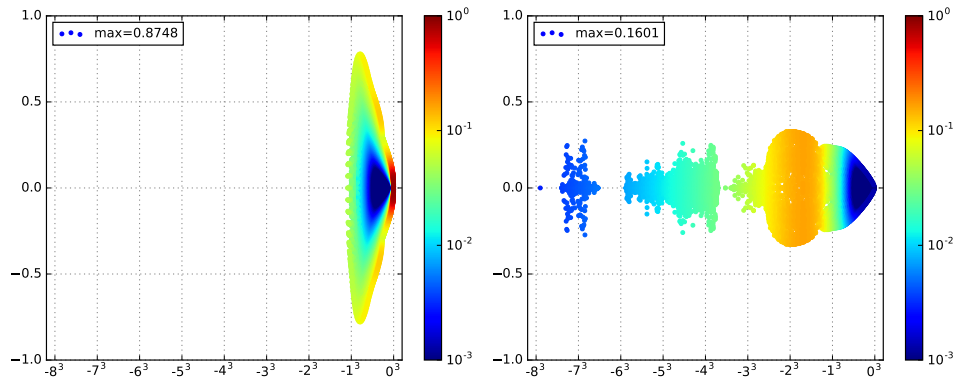

Fig. 4.3: Convergence estimates for $\|E_\omega\|_2$ using SDIRK-3, F-relaxation and $m = 2$. The point clouds indicate numerical eigenvalues of $\delta_t G$ colored by $\|E_\omega\|_2$ for the problem size $3 \cdot 2^{12} \times 2^{12}$. **Left:** Pure advection, $\eta = 0$. **Right:** Advection with third-order artificial dissipation according to (4.6). Note the *power law* scaling of the horizontal axis.

the DG upwinding process is insufficient to make the pure advection case work well together with MGRIT. Improving this behavior is a topic for future investigation.

To compare our experimental observations with our theory, we plot for the largest problem size the numerically computed eigenvalues of $\delta_t G$ in the complex plane, using a small circle, colored by the value of the estimating function $\|E_\omega\|_2$. Figure 4.3 (left) depicts this for the case of F-relaxation and $m = 2$. For the other problem sizes, see Figure S3.5 in Supplemental Materials. The plots indicate a worst case convergence factor of ≈ 0.87 for a small set of eigenvalues near the origin, which is pessimistic when compared to the results in Table 4.5.

Next we consider the diffusion-dominated case (4.5b). The results, shown in Table 4.6, are similar to those for the pure diffusion case (with SDIRK-3) from Section 4.1.2.

m / Size		$3 \cdot 2^8 \times 2^{10}$	$3 \cdot 2^{10} \times 2^{11}$	$3 \cdot 2^{12} \times 2^{12}$
	2	0.148	0.152	0.151
F-relaxation, F-cycles	4	0.142	0.142	0.142
	32	0.192	0.156	0.146
	2	0.005	0.004	0.004
FCF-relaxation, F-cycles	4	0.006	0.005	0.004
	32	0.034	0.018	0.008

Table 4.6: Asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation, for the case of advection with constant diffusion with $\eta = 0.1$, Q_3 elements and SDIRK-3. The omitted two-level results are virtually identical.

The two-level results are omitted because they are virtually identical to F-cycles.

In Table 4.7 we give the results for the case of third order artificial dissipation (4.5c)-(4.6). In particular, we wish to see if the results from one-dimensional advection with high-order artificial dissipation carry-over, i.e., if scalable MGRIT results are possible. We give both the average and (maximal) convergence factors because they differ significantly in this case, possibly because the theory no longer applies. When the average convergence factor is less than one (even when the maximal is greater than one), the method converged. The two-level results for $m = 2, 4, 8$ appear to be scalable and fast, especially for FCF-relaxation.

In order to compare our experimental results with our theory, we plot the numerically computed eigenvalues of $\delta_t G$ for the largest problem size in Figure 4.3 (right). See Figure S3.6 in Supplemental Materials for the other problem sizes. The theoretically predicted convergence value of ≈ 0.16 for $m = 2$ is accurate.

Moving to the multi-level F-cycle results in Table 4.7, convergence deteriorates, but still occurs. While improving this case is future work, we note that there is promise for a scalable solver because some of the numerical results may indicate an asymptote. While memory constraints limited us from running more grid sizes (our spatial solver is serial), we were able to run four grid sizes for FCF, F-cycles and $m = 16$. Here, we observed the experimental average convergence rates of 0.151, 0.194, 0.202 and 0.2383, i.e., almost scalable results.

5. Conclusions. In conclusion, we have developed a sharp two-grid convergence theory for linear problems and MGRIT, where the F-relaxation case is equivalent to parareal. The theory assumes that the temporal grid is uniform and that the coarse and fine time-grid propagators can be diagaonalized by the same set of eigenvectors, which is often the case.

The theory provides a novel description of how FCF-relaxation improves the convergence and stability of MGRIT and how the coarsening factor m impacts the method. For parabolic problems (i.e, problems where the negative real part of the spatial eigenvalues dominate), the convergence is excellent. For the case of FCF-relaxation, convergence is robust regarding changes in m, allowing for large coarsening factors which reduce memory overhead and communication. The benefit of FCF-relaxation grows with the order of the implicit Runge-Kutta method, e.g., the

m / Size		$3 \cdot 2^8 \times 2^{10}$	$3 \cdot 2^{10} \times 2^{11}$	$3 \cdot 2^{12} \times 2^{12}$
	2	0.123 (0.150)	0.130 (0.151)	0.130 (0.151)
Two-level	4	0.120 (0.141)	0.120 (0.141)	0.120 (0.141)
	32	0.468 (0.833)	0.777 (0.957)	0.856 (1.001)
F-relaxation	2	0.130 (0.287)	0.130 (5.607)	0.521 (32.14)
F-cycles	4	0.120 (1.750)	0.232 (3.841)	0.586 (9.421)
	32	0.468 (0.833)	0.777 (0.957)	0.856 (1.005)
	2	0.002 (0.004)	0.002 (0.004)	0.002 (0.004)
Two-level	4	0.004 (0.007)	0.004 (0.007)	0.004 (0.007)
	32	0.184 (0.689)	0.421 (0.795)	0.700 (0.864)
FCF-relaxation	2	0.002 (0.007)	0.010 (0.065)	0.043 (0.575)
F-cycles	4	0.021 (0.043)	0.071 (1.123)	0.161 (1.619)
	32	0.184 (0.689)	0.421 (0.795)	0.700 (0.864)

Table 4.7: The average and (maximal) convergence rates for MGRIT with F-relaxation and then with FCF-relaxation, for the case of advection and artificial dissipation (4.5c)-(4.6). Here, $\varepsilon = (10h_0)^{-3}$, with Q_3 elements and SDIRK-3.

convergence rate for SDIRK-3 is below 0.01.

For the case when the imaginary part of the spatial eigenvalue dominates, the results are less satisfactory. Some cases, such as $m = 2$ and backward Euler (SDIRK-1) converge, but many other cases diverge. For the general case of complex spatial eigenvalues, the results are mixed. Considering the example of 1D advection with artificial dissipation, convergence is restored in some cases for SDIRK-2 and SDIRK-3 when $m = 2$. In particular, the use of FCF-relaxation is critical for the two-level convergence to carry over to the multi-level F-cycle results. We note that showing scalable behavior for advective problems, even in these limited cases, is a novel contribution.

Lastly, this work explores how the convergence of MGRIT compares to the stability of the chosen time-stepping scheme. Overall, a stable time-stepping scheme does not necessarily imply convergence of MGRIT, although MGRIT with FCF-relaxation always converges for the diffusion dominated problems considered here.

REFERENCES

[1] GUILLAUME BAL, *On the convergence and the stability of the parareal algorithm to solve partial differential equations*, in Domain Decomposition Methods in Science and Engineering, TimothyJ. Barth, Michael Griebel, DavidE. Keyes, RistoM. Nieminen, Dirk Roose, Tamar Schlick, Ralf Kornhuber, Ronald Hoppe, Jacques Priaux, Olivier Pironneau, Olof Widlund, and Jinchao Xu, eds., vol. 40 of Lecture Notes in Computational Science and Engineering, Springer Berlin Heidelberg, 2005, pp. 425–432.
[2] A. BRANDT, *Multi-level adaptive computations in fluid dynamics*, 1979. Technical Report AIAA-79-1455, AIAA, Williamsburg, VA.
[3] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, SIAM Books, Philadelphia, 2000. Second edition.

[4] X. Dai and Y. Maday, *Stable parareal in time method for first- and second-order hyperbolic systems*, SIAM J. Sci. Comput., 35 (2013), pp. A52–A78.

[5] H. De Sterck, Thomas A. Manteuffel, Stephen F. McCormick, and Luke Olson, *Numerical conservation properties of H(div)-conforming least-squares finite element methods for the Burgers equation*, SIAM J. Sci. Comput., 26 (2005), pp. 1573–1597.

[6] 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.

[7] R. D. Falgout, S. Friedhoff, Tz. V. Kolev, S. P. MacLachlan, and J. B. Schroder, *Parallel time integration with multigrid*, SIAM Journal on Scientific Computing, 36 (2014), pp. C635–C661.

[8] R. D. Falgout, A. Katz, Tz. V. Kolev, J. B. Schroder, A. Wissink, and U. M. Yang, *Parallel time integration with multigrid reduction for a compressible fluid dynamics application*, Tech. Report LLNL-JRNL-663416, Lawrence Livermore National Laboratory, 2015.

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

[10] M. J. Gander, *50 years of time parallel time integration*, in Multiple Shooting and Time Domain Decomposition, T. Carraro, M. Geiger, S. Krkel, and R. Rannacher, eds., Springer, 2015, pp. 69–114.

[11] M. J. Gander and E. Hairer, *Nonlinear convergence analysis for the parareal algorithm*, in Domain decomposition methods in science and engineering XVII, Ulrich Langer, Marco Discacciati, David E. Keyes, Olof B. Widlund, and Walter Zulehner, eds., vol. 60 of Lecture Notes in Computational Science and Engineering, Springer Berlin Heidelberg, 2008, pp. 193–200.

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

[13] B. Gustafsson, H.-O. Kreiss, and J. Oliger, *Time dependent problems and difference methods*, Wiley–Interscience, 1995.

[14] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 848–864.

[15] J.-L. Lions, Y. Maday, and G. Turinici, *Résolution d'EDP par un schéma en temps "pararéel"*, C. R. Acad. Sci. Paris Sér. I Math., 332 (2001), pp. 661–668.

[16] Y. Maday, *The "parareal in time" algorithm*, in Sub-Structuring Techniques and Domain Decomposition Methods, F. Magoulès, ed., Computational Science, Engineering & Technology, Saxe-Coburg Publications, Stirlingshire, UK, 2010, ch. 2, pp. 19–44.

[17] *MFEM: Modular finite element methods.* `http://mfem.org`.

[18] M. L. Minion, *A hybrid parareal spectral deferred corrections method*, Comm. App. Math. and Comp. Sci., 5 (2010), pp. 265–301.

[19] M. L. Minion and S. A. Williams, *Parareal and spectral deferred corrections*, in Numerical Analysis and Applied Mathematics, T. E. Simos, ed., no. 1048 in AIP Conference Proceedings, AIP, 2008, pp. 388–391.

[20] J. Nievergelt, *Parallel methods for integrating ordinary differential equations*, Comm. ACM, 7 (1964), pp. 731–733.

[21] M. Ries, U. Trottenberg, and G. Winter, *A note on MGR methods*, J. Lin. Alg. Applic., 49 (1983), pp. 1–26.

[22] R. Speck, D. Ruprecht, R. Krause, M. Emmett, M. 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, IEEE Computer Society Press, 2012, p. 92.

[23] Gunnar Andreas Staff and Einar M Rønquist, *Stability of the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering, TimothyJ. Barth, Michael Griebel, DavidE. Keyes, RistoM. Nieminen, Dirk Roose, Tamar Schlick, Ralf Kornhuber, Ronald Hoppe, Jacques Priaux, Olivier Pironneau, Olof Widlund, and Jinchao Xu, eds., vol. 40 of Lecture Notes in Computational Science and Engineering, Springer Berlin Heidelberg, 2005, pp. 449–456.

[24] S. Vandewalle and G. Horton, *Fourier mode analysis of the multigrid waveform relaxation and time-parallel multigrid methods*, Computing, 54 (1995), pp. 317–330.

[25] *XBraid: Parallel multigrid in time.* `http://llnl.gov/casc/xbraid`.

[26] Irad Yavneh, Cornelis H. Venner, and Achi Brandt, *Fast multigrid solution of the advection problem with closed characteristics*, SIAM J. Sci. Comput., 19 (1998), pp. 111–125.

Supplemental Materials S1. Error propagation from a reduction point-of-view.

We now derive Lemmas 3.1 and 3.2 from a matrix reduction (Schur complement) point-of-view. This alternate derivation shows the relationship of MGRIT with multi-grid reduction and also standard error analyses of iterative methods.

The solution to (2.1) may also be written as the solution to the block $(N_t + 1) \times (N_t + 1)$ system

$$
\text{(S1.1)} \qquad A\mathbf{u} = \begin{pmatrix} I & & & \\ -\Phi & I & & \\ & \ddots & \ddots & \\ & & -\Phi & I \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N_t} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{N_t} \end{pmatrix} = \mathbf{g}.
$$

Sequential time stepping solves (S1.1) with a forward solve, which is $\mathcal{O}(N_t)$, but also sequential. MGRIT, instead, achieves parallelism in time by iteratively solving (S1.1) through multigrid reduction (MGR) [21] applied to the time dimension.

We start by reordering, with the F-block-rows first followed by the C-block-rows,

$$
\begin{bmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{bmatrix} \begin{bmatrix} \mathbf{u}_f \\ \mathbf{u}_c \end{bmatrix} = \begin{bmatrix} \mathbf{g}_f \\ \mathbf{g}_c \end{bmatrix}.
$$

In this ordering, we can consider the simple Schur-complement decomposition

$$
A = \begin{bmatrix} I_f & 0 \\ A_{cf} A_{ff}^{-1} & I_c \end{bmatrix} \begin{bmatrix} A_{ff} & 0 \\ 0 & A_{cc} - A_{cf} A_{ff}^{-1} A_{fc} \end{bmatrix} \begin{bmatrix} I_f & A_{ff}^{-1} A_{fc} \\ 0 & I_c \end{bmatrix}.
$$

This decomposition naturally implies operators R and P, known as ideal restriction and interpolation, and S,

$$
\text{(S1.2)} \qquad R = \begin{bmatrix} -A_{cf} A_{ff}^{-1} & I_c \end{bmatrix}, \quad P = \begin{bmatrix} -A_{ff}^{-1} A_{fc} \\ I_c \end{bmatrix}, \quad S = \begin{bmatrix} I_f \\ 0 \end{bmatrix}.
$$

Noting that $S^T A S = A_{ff}$ and $RAP = A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}$, we have

$$
A^{-1} = P(RAP)^{-1} R + S(S^T A S)^{-1} S^T.
$$

This gives the error propagator for the exact iterative method

$$
\text{(S1.3)} \qquad 0 = I - A^{-1} A = (I - P(RAP)^{-1} RA)(I - S(S^T A S)^{-1} S^T A),
$$

which is equivalent to the above formula for A^{-1} because $RAS = 0$. The first term corresponds to the error propagator for coarse grid correction using the ideal Petrov-Galerkin coarse grid operator, $RAP = A_\Delta$, and the second term, to the error propagator for F-relaxation. F-relaxation corresponds to setting the residual equal to zero at all F-points, i.e., $(S^T A S)^{-1} = A_{ff}^{-1}$. FCF-relaxation can be similarly defined as

$$
\text{(S1.4)} \qquad P(I - A_{cc}^{-1}(A_{cc} - A_{cf} A_{ff}^{-1} A_{fc})) R_I = P(I - A_\Delta) R_I,
$$

which is Jacobi-relaxation on the coarse-grid with the ideal coarse-grid operator. [5] Here $R_I = [0, I_c]^T$ is the simple injection operator.

To produce an iterative multigrid reduction method we approximate the ideal coarse-grid operator RAP. It is ideal because the method is exact and inverting RAP is as expensive as inverting A. To make a useful method, we introduce a cheaper approximation $B_\Delta \approx A_\Delta$. Making this substitution in (S1.3), the two grid error propagator for F-relaxation is

$$(S1.5) \qquad (I - PB_\Delta^{-1}RA)(I - S(S^T AS)^{-1}S^T A) = P(I - B_\Delta^{-1}A_\Delta)R_I.$$

Equivalence in (S1.5) holds because PR_I is equivalent to F-relaxation, i.e., P inverts the ff-block. See also equation (S1.8) for a representation of P as an application of F-relaxation.

Adding the effects of FCF-relaxation from (S1.4), yields the two grid error propagation operator

$$(S1.6) \qquad (I - PB_\Delta^{-1}RA)(P(I - A_\Delta)R_I) = P(I - B_\Delta^{-1}A_\Delta)(I - A_\Delta)R_I.$$

While the above is abstract, let us now revert back to the original unknown ordering of (S1.1) and consider the illuminating structure of R, P, A_Δ and B_Δ,

$$(S1.7) \qquad A_\Delta = \begin{bmatrix} I & & & \\ -\Phi^m & I & & \\ & \ddots & \ddots & \\ & & -\Phi^m & I \end{bmatrix}, \qquad B_\Delta = \begin{pmatrix} I & & & \\ -\Phi_\Delta & I & & \\ & \ddots & \ddots & \\ & & -\Phi_\Delta & I \end{pmatrix},$$

where the coarse-grid time stepper Φ_Δ is used to approximate m steps of the fine-grid time stepper. [6] Also, inverting A_Δ provides the exact solution at C-points, i.e., it simply applies Φ m-times. Lastly without loss of generality, we assume that $(N_t + 1)$ divides evenly by m and note that restriction and prolongation have the form

$$(S1.8) \qquad R = \begin{bmatrix} I & & & \\ \Phi^{m-1}\ldots\Phi & I & & \\ & & \ddots & \\ & & \Phi^{m-1}\ldots\Phi & I \end{bmatrix}, P = I \otimes \begin{bmatrix} I \\ \Phi \\ \vdots \\ \Phi^{m-1} \end{bmatrix}$$

where R is an F-relaxation, followed by injection from C-points to the coarse-grid, and P injects from the coarse-grid to C-points, followed by an F-relaxation to yield values for F-points. If the coarse-grid solution is exact, then one application of P yields the exact solution on the fine-grid, i.e., P leaves the residual zero at all F-points.

To derive Lemmas 3.1 and 3.2, we apply the original unknown ordering of (S1.1) to the two-grid error propagators (S1.5) and (S1.6). Since we are in a reduction framework with ideal interpolation that leaves zero residuals at F-points, the error

---

[5]This representation can be easily verified because FC-relaxation propagates each value at a C-point to the next adjacent C-point, i.e., it is propagation with Φ^m. See equation (S1.7) and [7].
   [6]Note that $R_I AP = RAP = A_\Delta$, hence we use R_I in the algorithm for computational savings.

can be analyzed only with the coarse-grid. The error propagators for F- and FCF-relaxation are then

with F-relaxation:
$$E_\Delta^F = I - B_\Delta^{-1} A_\Delta,$$

and with FCF-relaxation:
$$E_\Delta^{FCF} = (I - B_\Delta^{-1} A_\Delta)(I - A_\Delta).$$

The blocks of B_Δ^{-1} are

$$\left(B_\Delta^{-1}\right)_{ij} = \begin{cases} 0 & i < j \\ \Phi_\Delta^{i-j} & i \geq j, \end{cases} \quad i, j = 0, \dots, N_T.$$

In the following we dispense with the $\mathbf{e}_c$ notation, which was only used in this subsection to make our presentation consistent with other reduction-based methods. Instead, let an overbar denote a space-time vector containing a grid function at every point in space and at every C-point in time For example, the space-time error is represented by the vector $\bar{\mathbf{e}} = [\mathbf{e}_0^T, \mathbf{e}_m^T, \mathbf{e}_{2m}^T, \dots, \mathbf{e}_{N_T m}^T]^T$. This vector is blocked and $\bar{\mathbf{e}}_k = \mathbf{e}_{km}$ refers to the error at every point in space at the kth C-point in time.

The definition of B_Δ^{-1} yields the vector blocks of the product $E_\Delta^F \bar{\mathbf{e}}$ as

(S1.9)
$$(E_\Delta^F \bar{\mathbf{e}})_k = \begin{cases} 0, & k = 0, \\ \sum_{q=0}^{k-1} \Phi_\Delta^{k-1-q}(\Phi^m - \Phi_\Delta)\bar{\mathbf{e}}_q, & k = 1, 2, \dots, N_T, \end{cases}$$

which agrees with Lemma 3.1. Lemma 3.2 can be similarly derived.

Supplemental Materials S2. Fourier symbols of centered difference approximations.

To analyze spatially periodic solutions of partial differential equations with constant coefficients it is convenient to use the Fourier basis. For simplicity we restrict the presentation to one space dimension. The material in this section is essentially a summary of well-known results that, for example, can be found in the book by Gustafsson, Kreiss and Oliger [13].

S2.1. Fourier series expansion of a periodic grid function. Let $w(x)$: $\mathbb{R} \to \mathbb{C}$ be a 2π-periodic function of x, i.e., $w(x + 2\pi) = w(x)$. We introduce a spatial grid with constant step size $h = 2\pi/N_x$ and grid points $x_j = jh$, $j = 0, \pm 1, \pm 2, \ldots$. Let $w_j = w(x_j)$ be a grid function where $w_j = w_{j+N_x}$ because $w(x)$ is 2π-periodic. Assume for simplicity that N_x is even. Then the (normalized) Fourier series expansion of w_j is given by

$$(\text{S2.1}) \qquad w_j = \sum_{\omega=-N_x/2+1}^{N_x/2} \hat{w}_\omega s_j^{(\omega)}, \quad s_j^{(\omega)} = \frac{1}{\sqrt{2\pi}} e^{i\omega x_j}.$$

We define the discrete scalar product between grid functions u_j and v_j,

$$(u, v)_h := h \sum_{j=0}^{N_x-1} \bar{u}_j v_j.$$

It is straightforward to verify $(s^{(p)}, s^{(r)})_h = \delta_{pr}$, where p and r are integers and δ_{pr} is the Kronecker delta, i.e., $\delta_{pr} = 0$ if $p \neq r$ and $\delta_{rr} = 1$. Hence, the grid functions $\{s^{(-N_x/2+1)}, s^{(-N_x/2+2)}, \ldots, s^{(N_x/2)}\}$ form an orthonormal basis of the form (3.13) and the Fourier coefficients satisfy

$$\hat{w}_\omega = (s^{(\omega)}, w)_h.$$

S2.2. Difference approximations of d/dx. The centered finite difference approximations of d/dx, for accuracy orders $2p = 2, 4, 6$ and grid function q, are given by

$$(\text{S2.2}) \qquad D_1^{\{2\}} q_j = D_0 q_j,$$

$$(\text{S2.3}) \qquad D_1^{\{4\}} q_j = D_0 \left(I - \frac{h^2}{6} D_+ D_- \right) q_j,$$

$$(\text{S2.4}) \qquad D_1^{\{6\}} q_j = D_0 \left(I - \frac{h^2}{6} D_+ D_- + \frac{h^4}{30} D_+^2 D_-^2 \right) q_j,$$

where $D_+ q_j = (q_{j+1} - q_j)/h$ is the upwind difference operator, $D_- q_j = D_+ q_{j-1}$ is the downwind difference operator and $D_0 = 0.5(D_+ + D_-)$ is the classic centered difference operator.

A dissipative approximation of d/dx can be obtained by combining the centered formulae with an artificial dissipation term,

$$(\text{S2.5}) \qquad D_1^{\{2p,\varepsilon\}} q_j := D_1^{\{2p\}} q_j + \varepsilon h^{2p-1} R^{\{2p\}} q_j, \quad \varepsilon \geq 0.$$

The artificial dissipation terms satisfy

$$(\text{S2.6}) \qquad R^{\{2\}} q_j = D_+ D_- q_j,$$

$$(\text{S2.7}) \qquad R^{\{4\}} q_j = -D_+^2 D_-^2 q_j,$$

$$(\text{S2.8}) \qquad R^{\{6\}} q_j = D_+^3 D_-^3 q_j.$$

The truncation error in the formula (S2.5) is $\mathcal{O}(h^{2p})$ when $\varepsilon = 0$, but is reduced to $\mathcal{O}(h^{2p-1})$ for $\varepsilon > 0$.

The difference formula (S2.5) can be analyzed by Fourier decomposition. Note that

$$(S2.9) \qquad D_0 s_j^{(\omega)} = \frac{1}{2h}\left(e^{i\omega h} - e^{-i\omega h}\right)e^{i\omega x_j} = \frac{i}{h}\sin(\xi)s_j^{(\omega)},$$

$$(S2.10) \qquad D_+ D_- s_j^{(\omega)} = \frac{1}{h^2}\left(e^{i\omega h} - 2 + e^{-i\omega h}\right)e^{i\omega x_j} = -\frac{4}{h^2}\sin^2(\xi/2)s_j^{(\omega)},$$

where $\xi = \omega h$ is the scaled wave number. Because $-N_x/2 + 1 \le \omega \le N_x/2$, we have $-\pi < \xi \le \pi$. For a periodic grid function w_j with Fourier expansion (S2.1), this means that the periodic grid function $D_0 w_j$ has Fourier expansion

$$D_0 w_j = \sum_{\omega = -N_x/2+1}^{N_x/2} \hat{w}_\omega D_0 s_j^{(\omega)} = \sum_{\omega = -N_x/2+1}^{N_x/2} \left(\frac{i}{h}\sin(\xi)\hat{w}_\omega\right)s_j^{(\omega)}, \quad \xi = \omega h.$$

We call $\frac{i}{h}\sin(\xi)$ the Fourier symbol, which acts like an eigenvalue for wave number ω. We conclude that the basic difference operators D_0 and $D_+ D_-$ have Fourier symbols

$$(S2.11) \qquad \widehat{D_0}(\xi) = \frac{i}{h}\sin(\xi), \quad \widehat{D_+ D_-}(\xi) = -\frac{4}{h^2}\sin^2(\xi/2).$$

The Fourier symbol of (S2.5) can be written

$$(S2.12) \qquad \widehat{D}_1^{\{2p,\varepsilon\}}(\xi) = \widehat{D}_1^{\{2p\}}(\xi) + \varepsilon h^{2p-1}\widehat{R}^{\{2p\}}(\xi),$$

and the basic Fourier symbols (S2.11) can be combined to obtain the Fourier symbols $\widehat{D}_1^{\{2p\}}$ and $\widehat{R}^{\{2p\}}$. This results in

$$(S2.13) \quad \widehat{D}_1^{\{2,\varepsilon\}}(\xi) = \frac{1}{h}\left[i\sin(\xi) - 4\varepsilon\sin^2(\xi/2)\right],$$

$$(S2.14) \quad \widehat{D}_1^{\{4,\varepsilon\}}(\xi) = \frac{1}{h}\left[i\sin(\xi)\left(1 + \frac{2}{3}\sin^2(\xi/2)\right) - 16\varepsilon\sin^4(\xi/2)\right],$$

$$(S2.15) \quad \widehat{D}_1^{\{6,\varepsilon\}}(\xi) = \frac{1}{h}\left[i\sin(\xi)\left(1 + \frac{2}{3}\sin^2(\xi/2) + \frac{8}{15}\sin^4(\xi/2)\right) - 64\varepsilon\sin^6(\xi/2)\right].$$

The above Fourier symbols can be used to analyze the spectrum of the operator G in (3.23). For example, if the hyperbolic PDE $u_t = u_x$ is discretized by the dissipative finite difference method of order $2p$, it leads to a system of ODEs for the periodic grid function $q_j(t) \approx u(x_j, t)$ where the Fourier symbol is given by (S2.13)-(S2.15). In other words, the eigenvalues of the operator G in (3.23) satisfy

$$(S2.16) \qquad \gamma_\omega^{\{2p,\varepsilon\}} = \widehat{D}_1^{\{2p,\varepsilon\}}((\omega - N_x/2)h), \quad \omega = 1, 2, \ldots, N_x.$$

S2.3. Difference approximations of d^2/dx^2. A compact centered finite difference approximation of d^2/dx^2 of order $2p = 2, 4, 6$ is given by

$$(S2.17) \qquad D_2^{\{2\}}q_j = D_+ D_- q_j,$$

$$(S2.18) \qquad D_2^{\{4\}}q_j = D_+ D_-\left(I - \frac{h^2}{12}D_+ D_-\right)q_j,$$

$$(S2.19) \qquad D_2^{\{6\}}q_j = D_+ D_-\left(I - \frac{h^2}{12}D_+ D_- + \frac{h^4}{90}(D_+ D_-)^2\right)q_j.$$

Applying the same technique as before leads to the symbols

(S2.20) $$\widehat{D}_2^{\{2\}}(\xi) = \frac{-4}{h^2} \sin^2(\xi/2),$$

(S2.21) $$\widehat{D}_2^{\{4\}}(\xi) = \frac{-4}{h^2} \sin^2(\xi/2) \left[1 + \frac{4}{3} \sin^2(\xi/2) \right],$$

(S2.22) $$\widehat{D}_2^{\{6\}}(\xi) = \frac{-4}{h^2} \sin^2(\xi/2) \left[1 + \frac{4}{3} \sin^2(\xi/2) + \frac{8}{45} \sin^4(\xi/2) \right].$$

Supplemental Materials S3. Additional numerical results.

We now give our supplemental numerical results, as referenced from the text.

Effect of the Coarsening Factor on Stability Regions. In Figure S3.1 we present convergence contours of $\|E_\omega\|_2$ for the SDIRK-3 method and coarsening factors $m = 4$, $m = 8$, and $m = 16$.

(a) SDIRK-3, F-relaxation, $m = 4$

(b) SDIRK-3, FCF-relaxation, $m = 4$

(c) SDIRK-3, F-relaxation, $m = 8$

(d) SDIRK-3, FCF-relaxation, $m = 8$

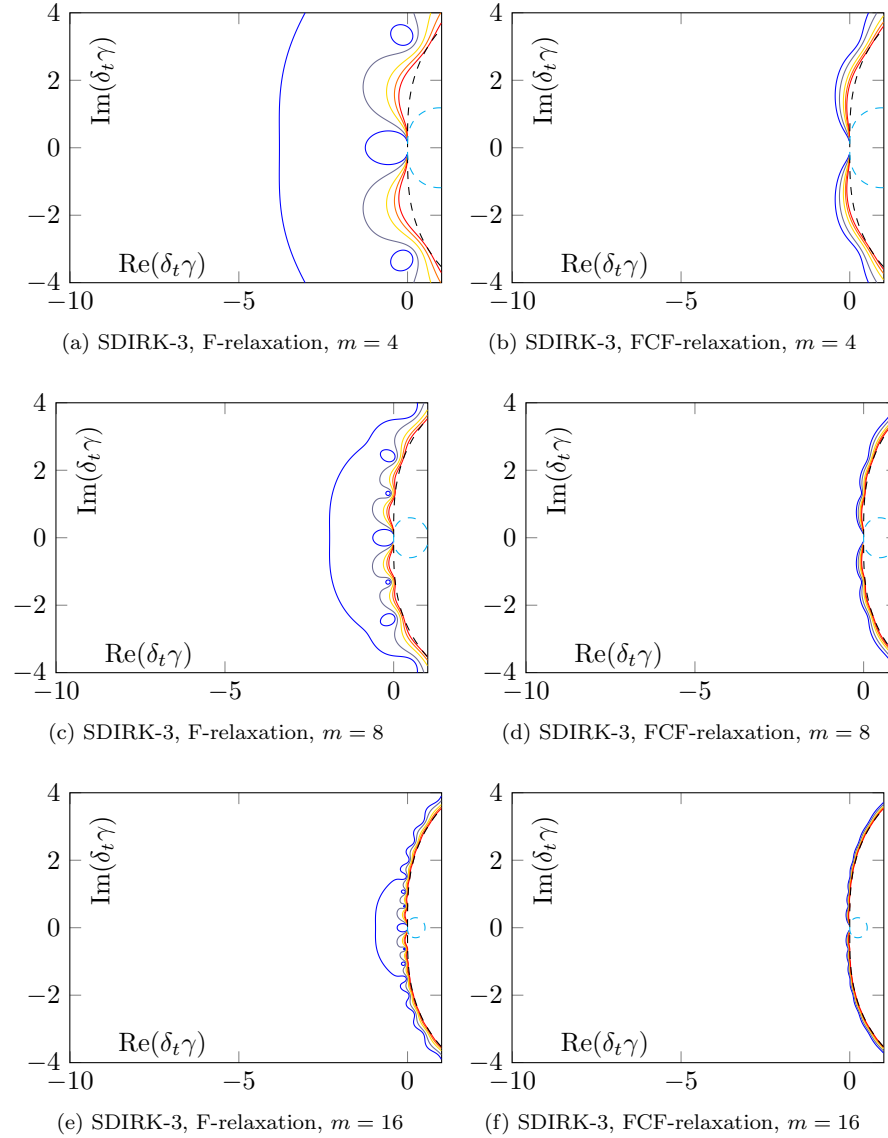(e) SDIRK-3, F-relaxation, $m = 16$

(f) SDIRK-3, FCF-relaxation, $m = 16$

Fig. S3.1: $\|E_\omega\|_2$ as function of the real and imaginary parts of $\delta_t\gamma$ for the SDIRK-3 method with different coarsening factors. See Figure 3.1, sub-figures (e) and (f), for the $m = 2$ case. The contour levels are the same as in Figure 3.1.

Two-dimensional heat equation results. Table S3.1 gives results for the two-dimensional heat equation, see Section 4.1.1. The domain is $([0, \pi])^2 \times [0, 49.35]$, with an initial condition of $u(t{=}0) = \sin(x)\sin(y)$. The ratio δ_t/h^2 remains 10.0. Standard central second-order finite-differencing in space is used, otherwise, the tests are the same as for one-dimension.

	m / Size	$(2^5)^2 \times 2^{10}$	$(2^6)^2 \times 2^{12}$	$(2^7)^2 \times 2^{14}$	$(2^8)^2 \times 2^{16}$
	4	0.203	0.197	0.199	0.199
Two-level	8	0.203	0.240	0.244	0.244
	32	0.241	0.224	0.274	0.280
F-relaxation	4	0.332	0.362	0.393	0.406
V-cycles	8	0.251	0.307	0.318	0.314
	32	0.241	0.225	0.275	0.286
	4	0.073	0.078	0.078	0.078
Two-level	8	0.093	0.087	0.094	0.093
	32	0.013	0.103	0.095	0.106
FCF-relaxation	4	0.097	0.102	0.105	0.107
V-cycles	8	0.093	0.086	0.092	0.091
	32	0.013	0.103	0.095	0.106

Table S3.1: Two-dimensional heat equation, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation.

Three-dimensional heat equation results. Table S3.2 gives results for the three-dimensional heat equation using Q_1 elements and SDIRK-2, see Section 4.1.2.

	m / Size	$2^9 \times 2^{10}$	$2^{12} \times 2^{11}$	$2^{15} \times 2^{12}$	$\|E_\Delta\|_2 \leq$
Two-level	2	0.004	0.012	0.037	0.295
	4	0.016	0.046	0.123	0.261
	8	0.051	0.129	0.239	0.261
F-relaxation	32	0.237	0.248	0.250	0.261
V-cycles	2	0.300	0.319	0.322	-
	4	0.227	0.270	0.269	-
	8	0.225	0.240	0.250	-
	32	0.237	0.248	0.250	-
Two-level	2	0.002	0.005	0.007	0.012
	4	0.007	0.009	0.009	0.010
	8	0.009	0.009	0.009	0.011
FCF-relaxation	32	0.009	0.009	0.010	0.011
V-cycles	2	0.006	0.007	0.007	-
	4	0.007	0.009	0.008	-
	8	0.009	0.009	0.009	-
	32	0.009	0.009	0.010	-

Table S3.2: Three-dimensional heat equation with Q_1 elements and SDIRK-2, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation. The theoretical bound appears in the final column.

S3.1. Example solution for the finite-element heat equation problems.
Figure S3.2 gives an example solution at an intermediate time for the heat equation considered in Section 4.1.2 and the mesh from the MFEM data file `pipe-nurbs.mesh`.
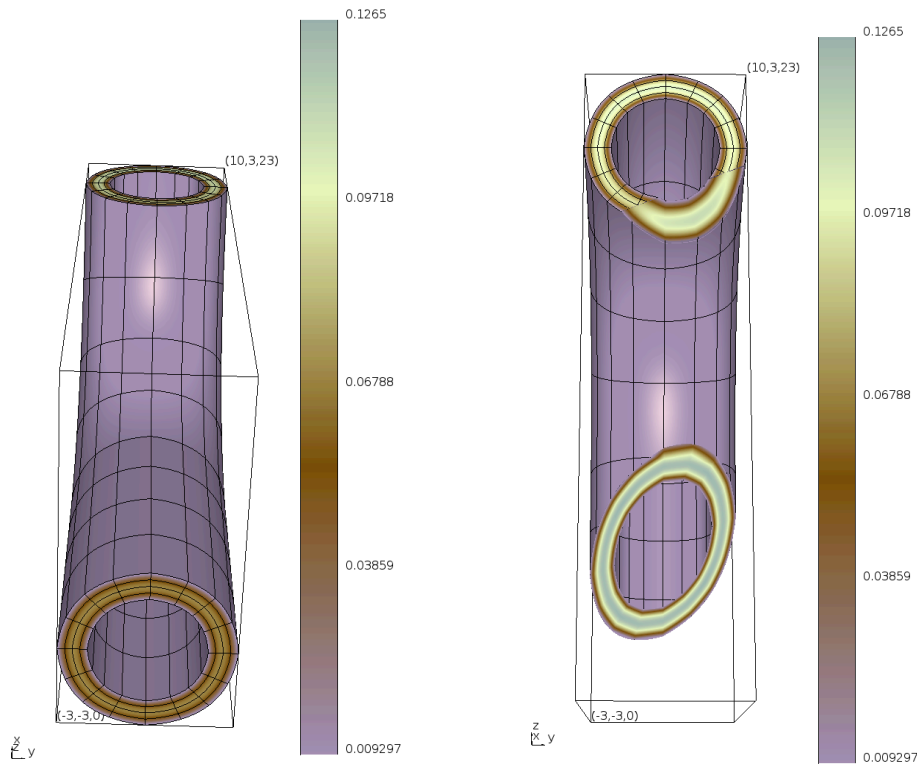

Fig. S3.2: Example solution for an intermediate t using `pipe-nurbs.mesh`. The right image depicts a cut-away of the left image.

S3.2. Convergence estimates for one-dimensional advection. In Figure S3.3 we give convergence estimates for one-dimensional advection with artificial dissipation and F-relaxation, see Section 4.2.1.

(a) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 0.0$

(b) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 0.0$

(c) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 0.5$

(d) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 0.5$

(e) SDIRK-1, low-order $\gamma_\omega^{2,\varepsilon}$, $\varepsilon = 4$

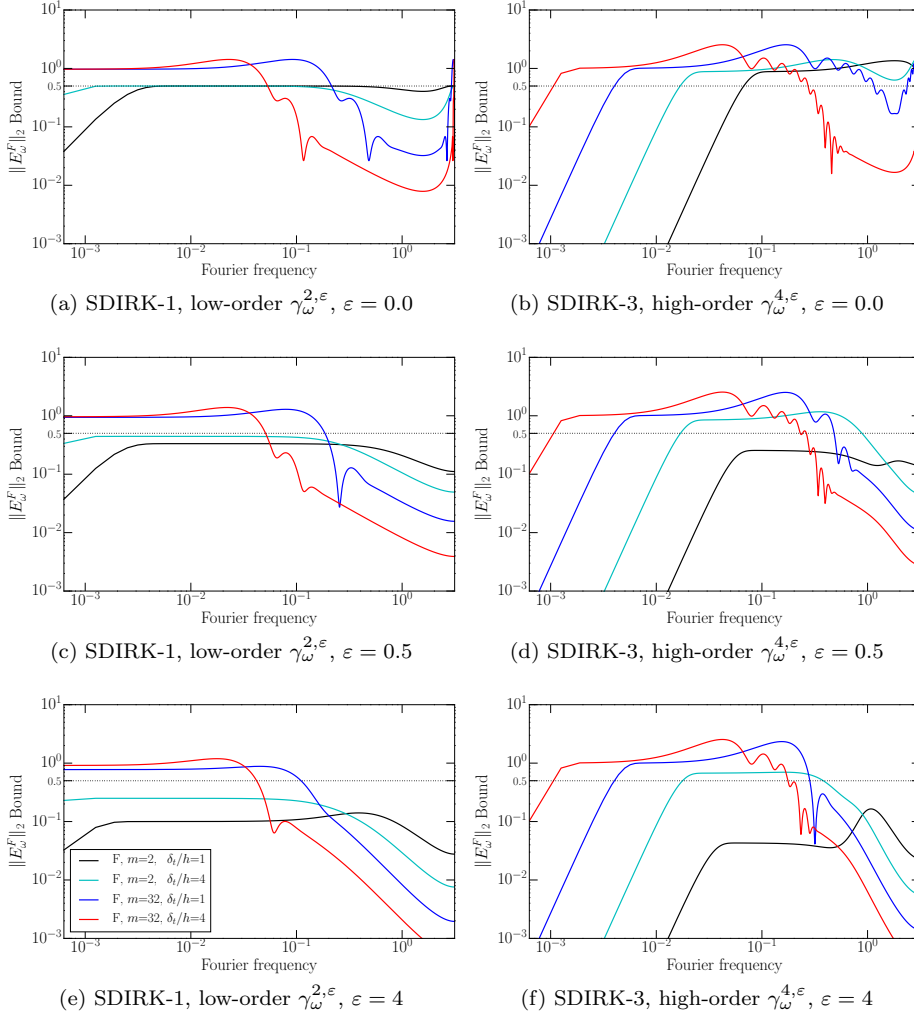(f) SDIRK-3, high-order $\gamma_\omega^{4,\varepsilon}$, $\varepsilon = 4$

Fig. S3.3: $\|E_\omega^F\|_2$ as function of the scaled Fourier frequency $\xi = \omega h$ for $\xi \in [0 < \xi \leq \pi]$, $m = 2, 32$ and $\delta_t/h = 1, 4$. Note that $\|E_\omega^F\|_2$ is a symmetric function of ξ and reflects around the origin.

One-dimensional advection equation with artificial dissipation. Table S3.3 gives results for the one-dimensional advection equation with artificial dissipation, see Section 4.1.2.

	ε / Size	$2^8 \times 2^8$	$2^9 \times 2^9$	$2^{10} \times 2^{10}$	$2^{11} \times 2^{11}$
Two-level	0	0.452	0.458	0.461	0.462
	0.5	0.390	0.400	0.405	0.406
F-relaxation	4.0	0.219	0.221	0.225	0.226
	0	0.546	0.666	0.778	0.861
F-cycles	0.5	0.432	0.598	0.650	0.811
	4.0	0.262	0.292	0.350	0.555
Two-level	0	0.276	0.383	0.441	0.462
	0.5	0.243	0.352	0.392	0.408
FCF-relaxation	4.0	0.166	0.203	0.215	0.219
	0	0.261	0.355	0.451	0.523
F-cycles	0.5	0.231	0.367	0.401	0.450
	4.0	0.149	0.201	0.262	0.316

Table S3.3: One-dimensional advection equation with artificial dissipation, asymptotic convergence rates for MGRIT with F-relaxation and then with FCF-relaxation; $m = 2$, $N_t = N_x$, $\delta_t/h = 4$ for various ε.

Initial condition for the finite-element advection-diffusion problems.
Figure S3.4 gives an example initial condition for the advection(-diffusion) problem
considered in Section 4.2.3 and the mesh from the MFEM data file
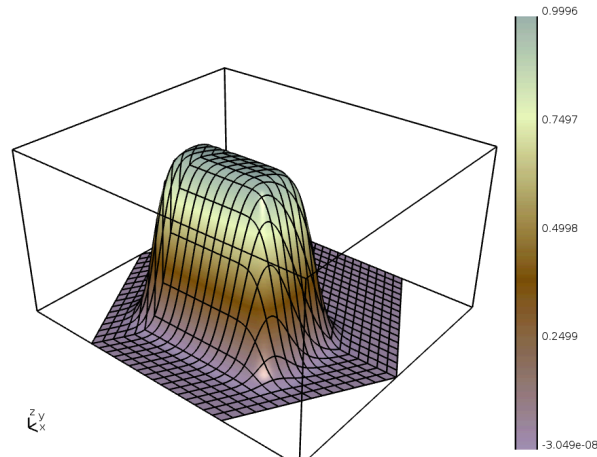`data/periodic-hexagon.mesh`.

Fig. S3.4: Initial condition for the advection(-diffusion) problem using
`data/periodic-hexagon.mesh`.

Convergence estimates for finite-element discretizations of advection and artificial dissipation. Figures S3.5 and S3.6 give the full version of Figure 4.3 from Section 4.2.3. Note that the contour plots in both Figures are variants of Figure 3.1 (e).
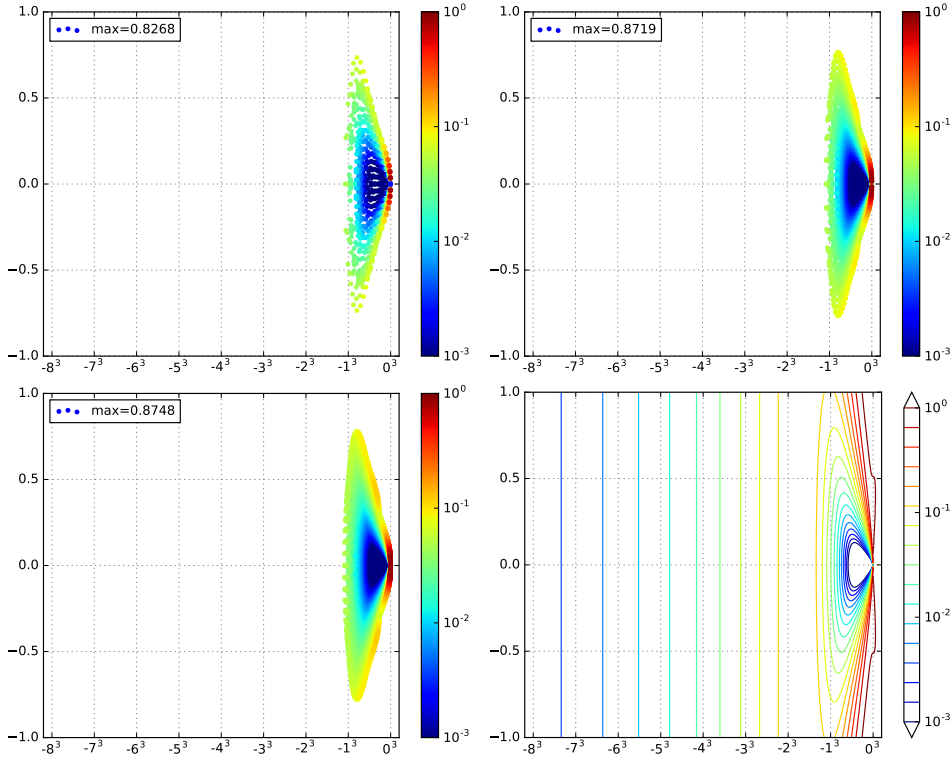

Fig. S3.5: Convergence estimates $\|E_\omega\|_2$ for the pure advection case $\eta = 0$ using SDIRK-3, F-relaxation and $m = 2$. The point clouds use numerical eigenvalues of $\delta_t G$ colored by $\|E_\omega\|_2$ for the problem sizes $3 \cdot 2^8 \times 2^{10}$ (top left), $3 \cdot 2^{10} \times 2^{11}$ (top right), and $3 \cdot 2^{12} \times 2^{12}$ (bottom left). The contour plot (bottom right) plots $\|E_\omega\|_2$ as a function in the complex plane. Note that the x-axis uses *power law* scaling.
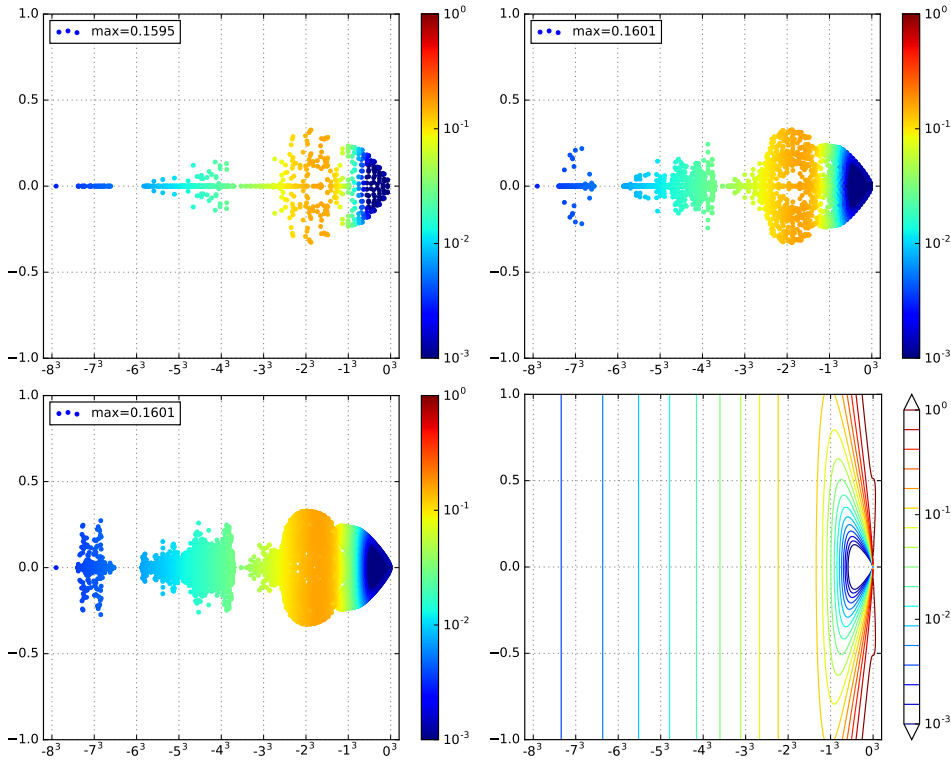
Fig. S3.6: Convergence estimates $\|E_\omega\|_2$ for the advection case with third-order arti-
ficial dissipation from equation (4.6) using SDIRK-3, F-relaxation and $m = 2$. The
point clouds use numerical eigenvalues of $\delta_t G$ colored by $\|E_\omega\|_2$ for the problem sizes
$3 \cdot 2^8 \times 2^{10}$ (top left), $3 \cdot 2^{10} \times 2^{11}$ (top right), and $3 \cdot 2^{12} \times 2^{12}$ (bottom left). The
contour plot (bottom right) plots $\|E_\omega\|_2$ as a function in the complex plane. Note
that the x-axis uses *power law* scaling.