Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 21/21

Masumi Sugiyama, Jacob B. Schroder, Ben S. Southworth, and Stephanie Friedhoff

Weighted Relaxation for Multigrid Reduction in Time

July 8, 2021

http://www.imacm.uni-wuppertal.de

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Weighted Relaxation for Multigrid Reduction in Time

Masumi Sugiyama*[1] | Jacob B. Schroder[2] | Ben S. Southworth[3] | Stephanie Friedhoff[4]

[1]Dept. of Mathematics, University of Tennessee, Chattanooga, USA
[2]Dept. of Mathematics and Statistics, University of New Mexico, New Mexico, USA
[3]Theoretical Division, Los Alamos National Laboratory, New Mexico, USA
[4]Dept. of Mathematics, University of Wuppertal, Germany

Correspondence
*Masumi Sugiyama, Dept. of Mathematics, University of Tennessee, Chattanooga, TN 37403, USA. Email: hgh889@mocs.utc.edu

Summary

Based on current trends in computer architectures, faster compute speeds must come from increased parallelism rather than increased clock speeds, which are currently stagnate. This situation has created the well-known bottleneck for sequential time-integration, where each individual time-value (i.e., time-step) is computed sequentially. One approach to alleviate this and achieve parallelism in time is with multigrid. In this work, we consider multigrid-reduction-in-time (MGRIT), a multilevel method applied to the time dimension that computes multiple time-steps in parallel. Like all multigrid methods, MGRIT relies on the complementary relationship between relaxation on a fine-grid and a correction from the coarse grid to solve the problem. All current MGRIT implementations are based on unweighted-Jacobi relaxation; here we introduce the concept of weighted relaxation to MGRIT. We derive new convergence bounds for weighted relaxation, and use this analysis to guide the selection of relaxation weights. Numerical results then demonstrate that non-unitary relaxation weights consistently yield faster convergence rates and lower iteration counts for MGRIT when compared with unweighted relaxation. In most cases, weighted relaxation yields a 10%–20% saving in iterations. For A-stable integration schemes, results also illustrate that under-relaxation can *restore convergence* in some cases where unweighted relaxation is not convergent.

KEYWORDS:
parallel-in-time, multigrid, multigrid-reduction-in-time, weighted relaxation, polynomial relaxation

1 | INTRODUCTION

Based on current trends in computer architectures, faster compute speeds must come from increased parallelism rather than increased clock speeds, which are stagnate. This situation has created a bottleneck for sequential time-integration [1–3], where each individual time-value (i.e., time-step) is computed sequentially. One approach to alleviate this is through parallelism in the time dimension, which goes back at least to Nievergelt [4] in 1964. For an introduction to parallel-in-time methods, see the review papers [1, 3], which give an overview of various approaches such as multiple shooting, waveform relaxation, domain decomposition, multigrid, and direct parallel-in-time methods.

In this work, we choose multigrid for parallelism in time for the same reasons that multigrid is often the method of choice for solving spatial problems [5, 6], i.e., a well-designed multigrid solver is an optimal method. In particular, we consider the multigrid-reduction-in-time (MGRIT) method [2], which has been applied in numerous settings, e.g., for nonlinear parabolic problems [7], compressible and incompressible Navier-Stokes [8, 9], elasticity [10], power-grid systems [11, 12], eddy current [13, 14], machine learning [15, 16], and more [3]. However, we note that there exist other powerful multigrid-like parallel-in-time methods such as the popular

parareal[17] and parallel full approximation scheme in space and time (PFASST)[18–20] methods. Parareal can be viewed as a two-level multigrid reduction method that coarsens in time[21]. PFASST can also be viewed as a multigrid method in time that utilizes a deferred correction strategy to compute multiple time-steps in parallel[22]. Unlike parareal, MGRIT is a full multilevel method applied to the time dimension, which allows for optimal scaling with respect to problem size. In contrast, for the two-level case, the coarsest temporal grid typically grows with problem size, yielding a potentially fast, but non-optimal method.

Like all multigrid methods, MGRIT relies on the complementary relationship between relaxation on a fine-grid, typically unweighted (block) Jacobi, and a correction from the coarse grid to solve the problem. In this work, we extend the use of weighted relaxation in multigrid[5, 6, 23, 24] to MGRIT, and analyze and select effective relaxation weights. With an appropriate choice of weight, MGRIT with weighted relaxation consistently offers faster convergence when compared with standard (unweighted) MGRIT, at almost no additional computational work[1]. Section 2 introduces a framework for weighted relaxation in MGRIT, and derives a new convergence analysis for linear two-grid MGRIT with degree-1 weighted-Jacobi relaxation. The theory is then verified with simple numerical examples in Section 3, and the utility of weighted relaxation is demonstrated on more complex problems in Section 4, including a 2D advection-diffusion problem and a 2D nonlinear eddy current problem. The new method consistently offers a 10–20% savings in iterations over standard unweighted MGRIT, and in some cases, (particularly A-stable integration schemes) yields convergence several times faster. Additional experiments are provided in the Supplemental Materials Appendix S2, exploring the effects of level-dependent relaxation weights for multilevel solvers and degree-2 weighted-Jacobi.

2 | MULTIGRID-REDUCTION-IN-TIME (MGRIT) AND WEIGHTED-JACOBI

2.1 | Two-level MGRIT method

This section derives the error-propagation operator for two-level linear MGRIT with weighted relaxation. Then, two-level convergence bounds are derived as a function of relaxation weight, providing insight on choosing the weight in practice. Although MGRIT uses full approximation storage (FAS) nonlinear multigrid cycling[25] to solve nonlinear problems, the linear two-grid setting makes analysis more tractable (e.g.,[26–30]), and MGRIT behavior for linear problems is often indicative of MGRIT behavior for related nonlinear problems[27]. Thus, consider a linear system of ordinary differential equations (ODEs) with N_x spatial degrees of freedom,

$$\frac{d\mathbf{u}}{dt} = G\mathbf{u}(t) + \mathbf{g}(t), \quad \mathbf{u}(0) = \mathbf{g}_0, \quad t \in [0, T], \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^{N_x}$ and $G \in \mathbb{R}^{N_x \times N_x}$ is a linear operator in space. For simplicity, define a uniform temporal grid as $t_j = j\delta_t$, for $j = 0, 1, .., N_t - 1$ where N_t refers to the number of points in time, with constant spacing $\delta_t = T/(N_t - 1) > 0$. Let $\mathbf{u}_j$ be an approximation to $\mathbf{u}(t_j)$ for $j = 1, 2, .., N_t - 1$ and $\mathbf{u}_0 = \mathbf{u}(0)$. Then, a general one-step time discretization for (1) is defined as

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{g}_0, \\ \mathbf{u}_j &= \Phi\mathbf{u}_{j-1} + \mathbf{g}_j, \quad j = 1, 2, ..., N_t - 1, \end{aligned} \tag{2}$$

where Φ is a one-step integration operator and $\mathbf{g}_j = \mathbf{g}(t_j)$. The solution to (2) for all time points is equivalent to solving the system of equations

$$\mathbf{Au} := \begin{bmatrix} I & & & \\ -\Phi & I & & \\ & \ddots & \ddots & \\ & & -\Phi & I \end{bmatrix} \begin{bmatrix} \mathbf{u_0} \\ \mathbf{u_1} \\ \vdots \\ \mathbf{u_{N_t-1}} \end{bmatrix} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \\ \vdots \\ \mathbf{g_{N_t-1}} \end{bmatrix} = \mathbf{g}. \tag{3}$$

While sequential time-stepping solves (3) directly with forward-substitution, MGRIT solves (3) iteratively by combining a block Jacobi relaxation with error corrections computed on a coarse-grid. Let the coarse temporal grid be $T_i = i\delta_T$, for $i = 0, 1, ..., N_{T-1}$ and $N_T = (N_t - 1)/m + 1$, which corresponds to a positive integer coarsening factor m and constant spacing $\delta_T = m\delta_t$. (Without loss of generality, we assume that $N_t - 1$ divides evenly by m in this description.) The original grid of points $\{t_j\}$ is then partitioned into C-points given by the set of coarse grid points $\{T_i\}$, and F-points given by $\{t_i\} \setminus \{T_i\}$ (see Figure 1). These C-points then induce a new coarser time-grid, with equivalent time-propagation problem

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{g}_0 \\ \mathbf{u}_{km} &= \Phi^m\mathbf{u}_{(k-1)m} + \tilde{\mathbf{g}}_{km}, \quad k = 1, 2, ..., N_T - 1, \end{aligned} \tag{4}$$

[1] Only one additional vector addition is performed.

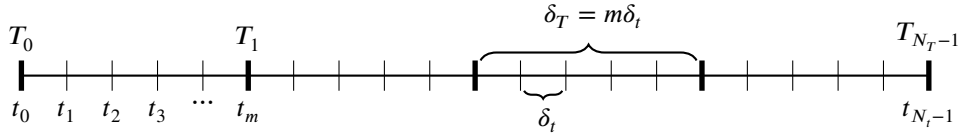


FIGURE 1 Uniform fine and coarse time-grid corresponding to coarsening factor m. The T_i are the C-points and form the coarse-grid, while the small hashmarks are F-points. Together, the F- and C-points form the fine-grid $\{t_j\}$.

(a) F-relaxation (b) C-relaxation

FIGURE 2 Schematic view of the action of (a) F-relaxation and (b) C-relaxation with a coarsening factor of $m = 4$.

where $\tilde{\mathbf{g}}_{km} = \mathbf{g}_{km} + \Phi\mathbf{g}_{km-1} + \cdots + \Phi^{m-1}\mathbf{g}_{(k-1)m+1}$. The solution to (4) is equivalent to solving the coarse system of equations

$$\mathbf{A}_\triangle \mathbf{u}_\triangle := \begin{bmatrix} I & & & \\ -\Phi^m & I & & \\ & \ddots & \ddots & \\ & & -\Phi^m & I \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_m \\ \vdots \\ \mathbf{u}_{(N_T-1)m} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \tilde{\mathbf{g}}_m \\ \vdots \\ \tilde{\mathbf{g}}_{(N_T-1)m} \end{bmatrix} = \mathbf{g}_\triangle, \tag{5}$$

where $\mathbf{A}_\triangle$ has N_T block rows and block columns. Unfortunately, solving equation (5) is as expensive as solving equation (3) because of the Φ^m operator. Thus, Φ^m is usually replaced with a cheap approximation $\Phi_\triangle$, which in turn induces a new operator on the coarse-grid, $\mathbf{B}_\triangle \approx \mathbf{A}_\triangle$. The operator $\mathbf{B}_\triangle$ has the exact same structure as $\mathbf{A}_\triangle$, only the Φ^m has been replaced by $\Phi_\triangle$. With the partition of F- and C-points as depicted in Figure 1, there are two fundamental types of relaxation: F- and C-relaxation. F-relaxation updates the F-point values based on the C-point values, i.e., one F-sweep updates each interval of F-points with

$$\mathbf{u}_i = \Phi\mathbf{u}_{i-1} + \mathbf{g}_i \quad \text{for } i = (km+1)\ldots((k+1)m-1), \tag{6}$$

and k is the F-interval index from 0 to $N_T - 2$. Similarly, C-relaxation updates each C-point value based on the preceding F-point value, i.e., the index i becomes km in equation (6). Each interval of F-points (T_{i-1}, T_i) for $i = 1, ..., N_T - 1$ can be updated simultaneously in parallel, and each C-point can also be updated simultaneously in parallel. Figure 2 illustrates the action of these relaxations in parallel. One application of F-relaxation followed by a C-relaxation updates each $\mathbf{u}_{km}$ based on $\mathbf{u}_{(k-1)m}$, which computes Φ^m applied to $\mathbf{u}_{(k-1)m}$ for $k = 1, ..., N_T - 1$. This FC-sweep corresponds to a block Jacobi iteration on the coarse-grid with $\mathbf{A}_\triangle$. Letting k denote the current relaxation iteration, this block Jacobi scheme can be written as

$$\begin{aligned} \mathbf{u}_\triangle^{(k+1)} &= \mathbf{u}_\triangle^{(k)} + D_\triangle^{-1}(\mathbf{g}_\triangle - \mathbf{A}_\triangle \mathbf{u}_\triangle^{(k)}) \\ &= \begin{bmatrix} \mathbf{u}_0^{(k)} \\ \mathbf{u}_m^{(k)} \\ \vdots \\ \mathbf{u}_{(N_T-1)m}^{(k)} \end{bmatrix} + D_\triangle^{-1} \begin{bmatrix} \mathbf{g}_0 - \mathbf{u}_0^{(k)} \\ \tilde{\mathbf{g}}_m + \Phi^m \mathbf{u}_0^{(k)} - \mathbf{u}_m^{(k)} \\ \vdots \\ \tilde{\mathbf{g}}_{(N_T-1)m} + \Phi^m \mathbf{u}_{(N_T-2)m}^{(k)} - \mathbf{u}_{(N_T-1)m}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \Phi^m \mathbf{u}_0^{(k)} + \tilde{\mathbf{g}}_m \\ \vdots \\ \Phi^m \mathbf{u}_{(N_T-2)m}^{(k)} + \tilde{\mathbf{g}}_{(N_T-1)m} \end{bmatrix}, \end{aligned} \tag{7}$$

where $D_\triangle$ is the diagonal of A_Δ and equal to the identity. The MGRIT algorithm performs either an F-relaxation or an FCF-relaxation, which consists of the initial F-relaxation, a C-relaxation, and a second F-relaxation.

2.1.1 | Weighted-Jacobi variant of FCF-relaxation

Here we introduce a *weighted* Jacobi relaxation to the MGRIT framework. Weighted-Jacobi relaxation with weight $\omega_C > 0$ applied to (7) takes the form

$$\mathbf{u}_\triangle^{(k+1)} = \omega_C\{(I - D_\triangle^{-1}A_\triangle)\mathbf{u}_\triangle^{(k)} + D_\triangle^{-1}\mathbf{g}_\triangle\} + (1-\omega_C)\mathbf{u}_\triangle^{(k)}, \quad k = 0, 1, 2, ... \tag{8}$$

We use ω_C to denote the weight in (8), because it will be shown that (8) is equivalent to applying a relaxation weight only during the C-relaxation step of an FC-sweep. Since the standard MGRIT FC-sweep corresponds to the block Jacobi method (7), it is thus natural to instead consider the weighted variant (8) inside of MGRIT.

In general, weighted relaxation has improved convergence for spatial multigrid methods applied to a variety of problems[5, 6, 23, 24], and so the remainder of this paper explores the application of weighted-Jacobi (8) in MGRIT. Regarding notation, the subscript $_F$ indicates the relaxation weight ω_F for F-relaxation, and subscript $_C$ indicates the weight ω_C for C-relaxation. Degree-two weighted-Jacobi will refer to two successive iterations of (8), possibly with different weights. The weight for the first C-relaxation, for example, is denoted ω_C, while the weight for the second is denoted ω_{CC}. It is called degree-two, because the resulting update to $\mathbf{u}_{\triangle}$ corresponds to a degree-two polynomial in $A_{\triangle}$.

2.2 | Convergence estimate for MGRIT with weighted-Jacobi relaxation

We now extend existing linear two-level MGRIT convergence bounds[27, 29] to account for the effects of weighted-Jacobi relaxation.

2.2.1 | MGRIT error propagator for unweighted FCF-relaxation

Let the fine-grid operator $\mathbf{A}$ in (3) be reordered so that F-points appear first and C-points second. Then by using the subscripts F and C to indicate the two sets of points, we have

$$\mathbf{A} = \begin{bmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{bmatrix}.$$

Define the ideal interpolation operator P [2], restriction by injection R_I, and a map to F-points S, respectively, as

$$P := \begin{bmatrix} -A_{FF}^{-1}A_{FC} \\ I_C \end{bmatrix}, \quad R_I := \begin{bmatrix} 0 & I_C \end{bmatrix}, \quad S := \begin{bmatrix} I_F \\ 0 \end{bmatrix}.$$

From[2], the two-level error propagator for linear MGRIT with unweighted FCF-relaxation is then given by

$$(I - PB_{\triangle}^{-1}R_I A)P(I - A_{\triangle})R_I = P(I - B_{\triangle}^{-1}A_{\triangle})(I - A_{\triangle})R_I. \tag{9}$$

2.2.2 | Two-level error propagator for weighted C-relaxation

Weighted-Jacobi for F-relaxation using the same structure as (8) can be written as

$$\begin{aligned} \mathbf{u}^{k+1} &= \omega_F\{(I - S(S^TAS)^{-1}S^TA)\mathbf{u}^k + D^{-1}\mathbf{g}\} + (1-\omega_F)\mathbf{u}^k \\ &= (I - \omega_F S(S^TAS)^{-1}S^TA)\mathbf{u}^k + \omega_F D^{-1}\mathbf{g}, \end{aligned} \tag{10}$$

where the first term (without $\mathbf{g}$) is the error propagator. Similarly, weighted-Jacobi for C-relaxation can be written as

$$\mathbf{u}^{k+1} = (I - \omega_C R_I^T(R_I A R_I^T)^{-1}R_I A)\mathbf{u}^k + D^{-1}\mathbf{g}, \tag{11}$$

where the first term (without $\mathbf{g}$) is the error propagator. Hence, the error propagator of FCF-relaxation with weighted-Jacobi is given by the product of F-, C-, and F-relaxation error-propagators:

$$(I - \omega_{FF}S(S^TAS)^{-1}S^TA)(I - \omega_C R_I^T(R_I A R_I^T)^{-1}R_I A)(I - \omega_F S(S^TAS)^{-1}S^TA), \tag{12}$$

where ω_{FF} denotes the weight for the second F-relaxation. Despite the above generality, moving forward we only consider $\omega_F = \omega_{FF} = 1.0$. If $\omega_{FF} \neq 1$, then MGRIT would no longer be an approximate reduction method. In other words, if the exact solution were given at C-points, the final F-relax using $\omega_{FF} \neq 1$ would no longer be guaranteed to yield the exact solution at F-points. We note that experiments also indicated $\omega_{FF} = 1$ performs best on model heat and advection problems. Similarly, letting $\omega_F \neq 1$ would restrict an inexact residual to the coarse grid problem, deviating from the principle of reduction methods.

[2] P is *ideal* because if an exact solution is available at C-points, then multiplication by P plus a right-hand-side contribution will yield the exact solution at all C- and F-points.

Thus, with this simplification, the error propagator for C-weighted FCF-relaxation takes the following block 2×2 form:

$$(I - S(S^T AS)^{-1} S^T A)(I - \omega_C R_I^T (R_I A R_I^T)^{-1} R_I A)(I - S(S^T AS)^{-1} S^T A) \tag{13a}$$

$$= \left(I - \begin{bmatrix} I_f & A_{ff}^{-1} A_{fc} \\ 0 & 0 \end{bmatrix} \right) \left(I - \omega_C \begin{bmatrix} 0 & 0 \\ A_{cc}^{-1} A_{cf} & I_c \end{bmatrix} \right) \left(I - \begin{bmatrix} I_f & A_{ff}^{-1} A_{fc} \\ 0 & 0 \end{bmatrix} \right) \tag{13b}$$

$$= \begin{bmatrix} 0 & -A_{ff}^{-1} A_{fc} \{ I_c - \omega_C A_{cc}^{-1} (A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}) \} \\ 0 & I_c - \omega_C A_{cc}^{-1} (A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}) \end{bmatrix} \tag{13c}$$

$$= \begin{bmatrix} -A_{ff}^{-1} A_{fc} \\ I_c \end{bmatrix} \left[I_c - \omega_C A_{cc}^{-1} (A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}) \right] \begin{bmatrix} 0 & I_c \end{bmatrix} \tag{13d}$$

$$= P(I - \omega_C \mathbf{A}_\triangle) R_I. \tag{13e}$$

Next, we take the two-level MGRIT error propagator with FCF-relaxation (9) and substitute in the new weighted variant (13e) to yield the following two-level error propagator for FCF-relaxation with weighted-C-Jacobi,

$$(I - P B_\triangle^{-1} R_I A) P (I - \omega_C \mathbf{A}_\triangle) R_I = P (I - \mathbf{B}_\triangle^{-1} \mathbf{A}_\triangle)(I - \omega_C \mathbf{A}_\triangle) R_I. \tag{14}$$

Lastly, to derive our convergence bound, we follow the convention from [27, 28] and examine the error propagator's effect only at C-points (i.e., drop the P and R_I from equation (14)). This simplification is typically made with the following motivation. If the solution at C-points is exact, then the final application of P in (14) will produce the exact solution at F-points, i.e., a zero residual. With this simplification, we denote the error propagator (14) at only C-points as $E_{\triangle,\omega_C}^{FCF}$, which takes the form

$$E_{\triangle,\omega_C}^{FCF} = (I - \mathbf{B}_\triangle^{-1} \mathbf{A}_\triangle)(I - \omega_C \mathbf{A}_\triangle) \tag{15a}$$

$$= \begin{bmatrix} 0 & & & \\ (1-\omega_C)(\Phi^m - \Phi_\triangle) & 0 & & \\ (1-\omega_C)\Phi_\triangle(\Phi^m - \Phi_\triangle) + \omega_C(\Phi^m - \Phi_\triangle)\Phi^m & (1-\omega_C)(\Phi^m - \Phi_\triangle) & 0 & \\ \vdots & \vdots & \ddots & 0 \\ (1-\omega_C)\Phi_\triangle^{N_T-1}(\Phi^m - \Phi_\triangle) + \omega_C \Phi_\triangle^{N_T-2}(\Phi^m - \Phi_\triangle)\Phi^m & \cdots & \cdots & (1-\omega_C)(\Phi^m - \Phi_\triangle)\ 0 \end{bmatrix}. \tag{15b}$$

2.2.3 | Two-grid eigenvalue convergence analysis

To guarantee convergence, ideally we bound (15) in some norm (e.g., see [29]). However, working in a norm can be difficult; thus we take the more tractable approach of considering convergence for individual eigenvectors [27, 29]. Thus, assume that Φ and Φ_Δ have the same set of eigenvectors, $\{v_\gamma\}$, as occurs when the same spatial discretization is used on the coarse and fine grid in time, and let $\{\lambda_\gamma\}$ be the eigenvalues of Φ and $\{\mu_\gamma\}$ be the eigenvalues of $\Phi_\triangle$. For instance, let $\kappa_\gamma \geq 0$ denote an eigenvalue of the linear operator G in (1); if backward Euler is used on the coarse and fine grid, we have

$$\lambda_\gamma = (1 - h_t \kappa_\gamma)^{-1}, \text{ and } \mu_\gamma = (1 - m h_t \kappa_\gamma)^{-1} \quad \text{for } \gamma = 1, 2, ..., N_x. \tag{16}$$

Define $\widetilde{U}$ as a block-diagonal operator, with diagonal blocks given by the eigenvector matrix for Φ and Φ_Δ. Following the discussion of Section 5 in [29], we can apply $\widetilde{U}$ to the left and $\widetilde{U}^{-1}$ to the right of (15). The resulting operator is then block diagonal, with diagonal blocks corresponding to a single pair of eigenvalues $\{\lambda_\gamma, \mu_\gamma\}$, and takes the following form:

$$\widetilde{E}_{\triangle,\omega_C}^{FCF} = \begin{bmatrix} 0 & & & \\ (1-\omega_C)(\lambda_\gamma^m - \mu_\gamma) & 0 & & \\ (1-\omega_C)\mu_\gamma(\lambda_\gamma^m - \mu_\gamma) + \omega_C(\lambda_\gamma^m - \mu_\gamma)\lambda_\gamma^m & (1-\omega_C)(\lambda_\gamma^k - \mu_\gamma) & 0 & \\ \vdots & \ddots & \ddots & \ddots \\ (1-\omega_C)\mu_\gamma^{N_T-1}(\lambda_\gamma^m - \mu_\gamma) + \omega_C \mu_\gamma^{N_T-2}(\lambda_\gamma^m - \mu_\gamma)\lambda_\gamma^m & \cdots & \cdots & (1-\omega_C)(\lambda_\gamma^m - \mu_\gamma)\ 0 \end{bmatrix}. \tag{17}$$

Following the analysis in [27, 29], we can provide bounds on (15) in a certain eigenvector-induced $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm by bounding (17) in norm and taking the maximum over γ (note, if the spatial matrix is SPD, $\widetilde{U}$ is unitary, and the $(\widetilde{U}\widetilde{U}^*)^{-1}$-norm is simply the

ℓ^2-norm). Note that (17) is a Toeplitz matrix, with asymptotic generating function

$$\begin{aligned}\mathcal{F}_\gamma(x) &:= (\lambda_\gamma^m - \mu_\gamma)\left[(1-\omega_C)\sum_{\ell=1}^{\infty}\mu_\gamma^{\ell-1}e^{i\ell x} + \omega_C\lambda_\gamma^m\sum_{\ell=2}^{\infty}\mu_\gamma^{\ell-2}e^{i\ell x}\right]\\ &= e^{ix}(\lambda_\gamma^m - \mu_\gamma)\left[(1-\omega_C)\sum_{\ell=0}^{\infty}(\mu_\gamma e^{ix})^\ell + e^{ix}\omega_C\lambda_\gamma^m\sum_{\ell=0}^{\infty}(\mu_\gamma e^{ix})^\ell\right]\\ &= e^{ix}\frac{(\lambda_\gamma^m - \mu_\gamma)}{1-e^{ix}\mu_\gamma}\left[1-\omega_C + e^{ix}\omega_C\lambda_\gamma^m\right].\end{aligned}$$

Noting that $\mathcal{F}_\gamma(x) \in L^1[-\pi,\pi]$, from[31] (see also[32, Th. 2.1]), we have that

$$\begin{aligned}\sigma_{max,\gamma}(\widetilde{E}^{FCF}_{\triangle,\omega_C}) &\leq \max_{x\in[0,2\pi]}|\mathcal{F}_\gamma(x)|\\ &= \max_{x\in[0,2\pi]}\frac{|\lambda_\gamma^m - \mu_\gamma|}{|1-e^{ix}\mu_\gamma|}|1-\omega_C + e^{ix}\omega_C\lambda_\gamma^m|.\end{aligned} \tag{18}$$

Taking the maximum over γ, corresponding to all (shared) eigenvectors of Φ and Φ_Δ yields the following final result.

Theorem 1. Assume that Φ and Φ_Δ have the same set of eigenvectors, with eigenvalues $\{\lambda_\gamma\}$ and $\{\mu_\gamma\}$, respectively, where $|\lambda_\gamma|, |\mu_\gamma| < 1$ for all $\gamma \in [1, N_x]$. Let $\widetilde{U}$ denote a block-diagonal operator, with diagonal blocks given by the eigenvector matrix of Φ and Φ_Δ. Then,

$$\|E^{FCF}_{\triangle,\omega_C}\|_{(\widetilde{U}\widetilde{U}^*)^{-1}} \leq \max_\gamma\max_{x\in[0,2\pi]}\frac{|\lambda_\gamma^m - \mu_\gamma|}{|1-e^{ix}\mu_\gamma|}|1-\omega_C + e^{ix}\omega_C\lambda_\gamma^m|. \tag{19}$$

Proof. The proof follows from the above discussion. □

For fixed γ, a closed form for the maximum over x in (19) to allow for easier computation is provided in the Supplemental materials.

We numerically verify the convergence bound (19) in Section 3.1 for model 1D heat and advection equations, respectively. In some cases, the bound is quite tight, while for others the general behavior is right, but bounds are not exact. This is likely due to Theorem 1 providing an *upper* bound on worst-case convergence; even if the upper bound is tight (which Theorem 1 is asymptotically in N_T), it is possible that better convergence can be observed in practice, depending on the problem and right-hand side.

Remark 1. We also note that one can approximate the maximum over x in Theorem 1 by assuming a fixed x rotates λ_γ and μ_γ to the real-axis. Experiments have indicated this to be a reasonable assumption for eigenvalues with dominant real-part, although less so for eigenvalues with large imaginary component. Nevertheless, it does yield a simpler measure to compute, and can be applied to weighted FCF- and FCFCF-relaxation (degree-two weighted-Jacobi), with approximate bounds

$$\begin{aligned}\|E^{FCF}_{\triangle,\omega_C}\|_{(\widetilde{U}\widetilde{U}^*)^{-1}} &\lessapprox \max_\gamma\frac{|\lambda_\gamma^m - \mu_\gamma|}{1-|\mu_\gamma|}|1-\omega_C + \omega_C|\lambda_\gamma^m||,\\ \|E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}\|_{(\widetilde{U}\widetilde{U}^*)^{-1}} &\lessapprox \max_\gamma\frac{|\lambda_\gamma^m - \mu_\gamma|}{1-|\mu_\gamma|}|1-\omega_C + \omega_C|\lambda_\gamma^m||\,|1-\omega_{CC} + \omega_{CC}|\lambda_\gamma^m||.\end{aligned} \tag{20}$$

For the derivation of the FCFCF-bound, see Appendix A.

3 | VERIFYING THE CONVERGENCE BOUND

3.1 | Numerical verification of the convergence bound

We focus our verification tests on three model problems with the following spatial discretizations, the 1D heat equation (second-order central differencing in space), the 1D advection equation with purely imaginary spatial eigenvalues (second-order central differencing in space), and the 1D advection equation with complex spatial eigenvalues (first-order upwinding in space). In all

cases, backward Euler is used in time.[3] We choose these model problems because the theoretical motivation of equation (19) indicates that it is the character of the spatial eigenvalues and the time-stepping method that determine the convergence of MGRIT, i.e., not the dimensionality of the problem, the complexity of the governing PDE, or the nature of the forcing term and boundary conditions. Thus, we choose these three representative cases, similar to[27, 28].

We consider the 1D heat equation subject to an initial condition and homogeneous Dirichlet boundary conditions,

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial^2 u}{\partial x^2} &= f(x,t), \quad \alpha > 0, && x \in \Omega = [0, L], \quad t \in [0, T], \\ u(x,0) &= u_0(x), && x \in \Omega, \\ u(x,t) &= 0, && x \in \partial\Omega, \quad t \in [0, T]. \end{aligned} \tag{21}$$

For numerical experiments, we use the space-time domain $[0,1] \times [0, 0.625]$, the diffusivity constant $\alpha = 1$, and the right-hand side $f(x,t) = \sin(\pi x)[\sin(t) - \pi^2 \cos(t)]$. Note that with these choices, the analytical solution is given by $u(x,t) = \sin(\pi x)\cos(t)$. A random initial guess and a residual norm halting tolerance of $10^{-10}/\sqrt{h_x h_t}$ are used. Reported convergence rates are taken as an average over the last five MGRIT iterations, where $\|r_k\|_2/\|r_{k-1}\|_2$ is the convergence rate at iteration k and r_k is the residual from equation (3) at iteration k. The combination of grid points in space N_x and time N_t are chosen so that $\frac{h_t}{h_x^2} = 12.8$. This value was chosen to be of moderate magnitude and consistent with other MGRIT literature, namely the work[27].

We also consider the 1D advection equation with purely imaginary spatial eigenvalues, subject to an initial condition and periodic spatial boundary conditions,

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial u}{\partial x} &= 0, \quad \alpha > 0, && x \in \Omega = [0, L], \quad t \in [0, T], \\ u(x,0) &= u_0(x), && x \in \Omega, \\ u(0,t) &= u(L,t), && t \in [0, T]. \end{aligned} \tag{22}$$

The space-time domain considered is $[0,1] \times [0,1]$, the velocity constant $\alpha = 1$, and the analytical solution $u(x,t) = e^{-25((x-t)-0.5)^2}$. The solution is chosen as a standard test problem that satisfies the spatially periodic boundary conditions. A random initial guess and a residual norm halting tolerance of $10^{-8}/\sqrt{h_x h_t}$ are used. The maximum allowed iterations is set to 70, because some cases will fail to quickly converge. Reported convergence rates are taken as $(\|r_k\|_2/\|r_0\|_2)^{1/k}$ at the final iteration k. The geometric average is used (as opposed to the heat equation case above) because the per iteration convergence rate here can vary significantly. The combination of grid points in space N_x and time N_t are chosen so that $\frac{h_t}{h_x} = 0.5$.

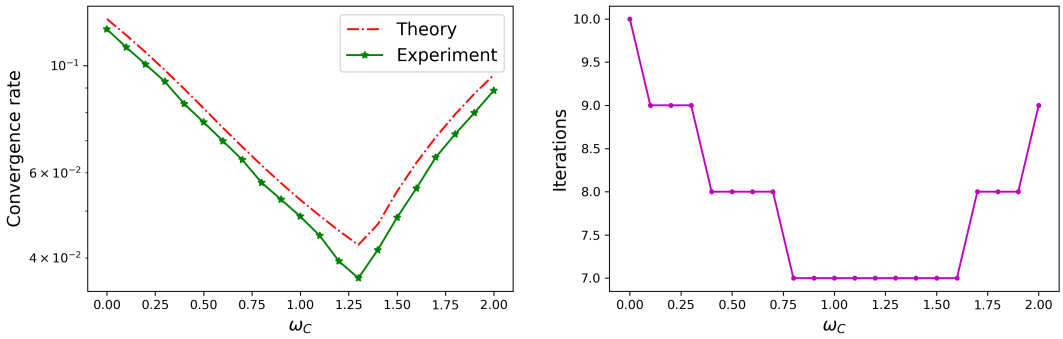
Figure 3 (a) and Figure 4 (a) depict the convergence bound (dashed line) and experimental convergence rates (solid line) against various relaxation weights ω_C for the 1D heat equation and the 1D advection equation with purely imaginary spatial eigenvalues, respectively. Figure 3 (b) and Figure 4 (b) show the iterations associated with the experimental convergence rates. For Figure 3, the theoretical bound is very tight and predicts the optimal ω_C. For the advective case in Figure 4, the bound is predictive, but not quite sharp enough to predict the best weight. The results for the 1D advection equation with complex spatial eigenvalues are similar to the 1D advection equation with purely imaginary spatial eigenvalues and, thus, are omitted.

Next, we summarize the experimentally best relaxation weights for the 1D heat equation and the 1D advection equation with purely imaginary spatial eigenvalues. For the full multilevel experiments, V-cycles are used and we coarsen down to a grid of size 4 or less in time. During searches in the weight-space for experimentally optimal weights, we use a step size of 0.1, and in these tables we report only the best weight in comparison to a unitary weight of 1.0. For expanded versions of these tables, please see Supplemental Materials S2, Tables S1, S2, S5, and S6. Regarding notation, ω_{CC} denotes the weight for the second weighted relaxation, if degree-two (FCFCF) weighted relaxation is used. If only ω_C is given, then only degree-one (FCF) weighted relaxation is used.

Tables 1 and 2 depict the results for the 1D heat equation for a two-level and multi-level solver, respectively. The best experimental weight for degree-one relaxation in both cases is $\omega_C = 1.3$ and saves 1 iteration on the largest problem, or approximately 10%–14%. The best weights (ω_C, ω_{CC}) for degree-two relaxation differ between two-level and multilevel, but similarly save 1 iteration. Other coarsening factors m were tested, but generated the same experimentally best weights (see Supplemental Results Section S2.1 for more details).

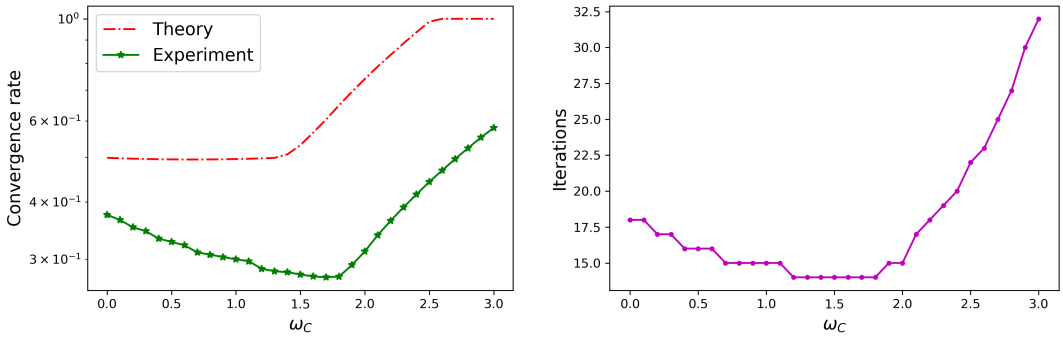
Tables 3 and 4 depict the results for the 1D advection equation with purely imaginary spatial eigenvalues for a two-level and multilevel solver, respectively. The best experimental weights for degree-one relaxation differ between not only two-level and

[3] For a complete description of these problems, see the Supplemental Materials for the heat equation in Section S2.1, the advection equation with purely imaginary spatial eigenvalues in Section S2.2, and the advection equation with complex spatial eigenvalues in Section S2.3.



(a) Convergence Rate (b) Iterations

FIGURE 3 Two-level MGRIT theoretical bound (dashed line in left plot), experimental convergence rates (solid line in left plot), and iteration counts (right plot) as a function of relaxation weights ω_C for the one-dimensional heat equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (291, 4097)$.



(a) Convergence Rate (b) Iterations

FIGURE 4 Two-level MGRIT theoretical bound (dashed line in left plot), experimental convergence rates (solid line in left plot) and iteration counts (right plot) as a function of relaxation weights ω_C for the one-dimensional linear advection equation with purely imaginary spatial eigenvalues, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (1025, 1025)$.

multilevel but also coarsening factors $m = 2$ and $m = 4$. The best experimental weight in the two-level case with $m = 4$ is $\omega_C = 1.5$ and saves 2–3 iterations on the larger problems, or approximately 5%-9%. The best experimental weight in the multilevel case with $m = 2$ is $\omega_C = 1.5$ and saves 15 iterations on the second largest problem, or approximately 22%. The best weights (ω_C, ω_{CC}) for degree-two relaxation have been omitted for brevity, but are in Supplemental Materials Section S2.2.

	$N_x \times N_t$	291×4097	411×8193	581×16385	821×32769
$m = 2$	$\omega_C = 1.0$	0.049 (7)	0.048 (7)	0.039 (7)	0.039 (7)
	1.3	0.036 (7)	0.036 (7)	0.034 (6)	0.034 (6)
	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.029 (6)	0.029 (6)	0.029 (6)	0.028 (6)
	$(1.7, 0.9)$	0.020 (6)	0.020 (6)	0.019 (6)	0.016 (5)

TABLE 1 1D heat equation, two-level MGRIT convergence rates (iterations) for weighted FCF- and FCFCF-relaxation with unitary weights and the experimentally best weights.

	$N_x \times N_t$	291×4097	411×8193	581×16385	821×32769
$m = 2$	$\omega_C = 1.0$	0.118 (9)	0.121 (9)	0.123 (9)	0.125 (9)
	1.3	0.092 (8)	0.095 (8)	0.096 (8)	0.096 (8)
	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.065 (7)	0.066 (7)	0.067 (7)	0.068 (7)
	(2.0, 0.9)	0.032 (6)	0.032 (6)	0.032 (6)	0.032 (6)

TABLE 2 1D heat equation, multilevel MGRIT convergence rates (iterations) for weighted FCF- and FCFCF-relaxation with unitary weights and the experimentally best weights.

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.304 (15)	0.307 (15)	0.308 (15)	0.309 (15)
	1.8	0.280 (14)	0.282 (14)	0.284 (14)	0.285 (14)
$m = 4$	$\omega_C = 1.0$	0.564 (30)	0.607 (34)	0.617 (35)	0.619 (35)
	1.5	0.568 (30)	0.581 (31)	0.591 (32)	0.596 (33)

TABLE 3 1D linear advection equation, two-level MGRIT convergence rates (iterations) for weighted FCF-relaxation with unitary weights and the experimentally best weights.

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.560 (30)	0.675 (44)	0.771 (67)	(> 100)
	1.5	0.495 (24)	0.606 (35)	0.718 (52)	0.810 (82)
$m = 4$	$\omega_C = 1.0$	0.581 (32)	0.666 (42)	0.757 (61)	0.838 (95)
	1.4	0.535 (27)	0.611 (34)	0.712 (50)	0.802 (77)

TABLE 4 1D linear advection equation, multilevel MGRIT convergence rates (iterations) for weighted FCF-relaxation with unitary weights and the experimentally best weights.

3.2 | Visualizing the convergence bound

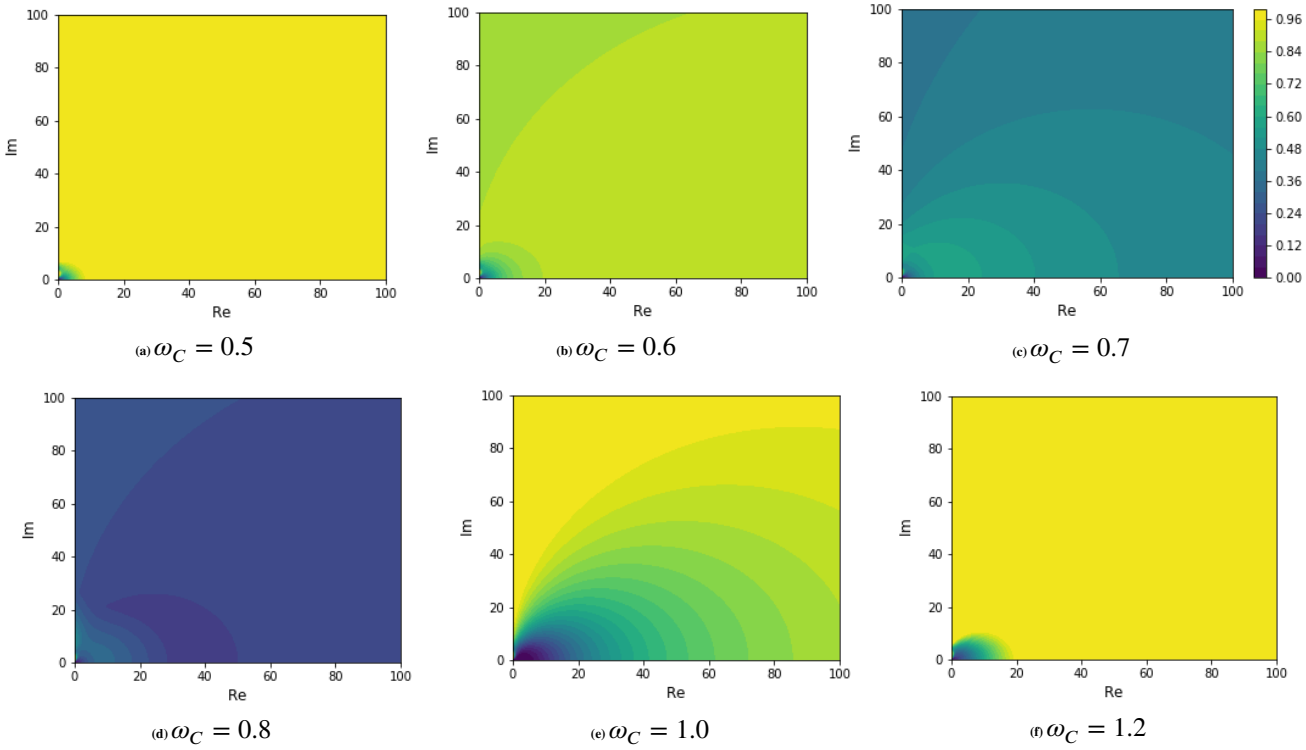
Recall that $\{\lambda_\gamma\}$ and $\{\mu_\gamma\}$ are the eigenvalues of Φ and $\Phi_\triangle$, respectively corresponding to the same set of eigenvectors $\{v_\gamma\}$. That is, Φ and $\Phi_\triangle$ are diagonalized by the eigenvectors $\{v_\gamma\}$. If $\kappa_\gamma \geq 0$ is an eigenvalue of the linear operator G in (1), the corresponding eigenvalue of Φ is given by

$$\lambda_\gamma = 1 + h_t\kappa_\gamma \mathbf{b}_0^T (I - h_t\kappa_\gamma A_0)^{-1}\mathbf{1}, \text{ and } \mu_\gamma = 1 + mh_t\kappa_\gamma \mathbf{b}_0^T (I - mh_t\kappa_\gamma A_0)^{-1}\mathbf{1} \tag{23}$$

where the Runge-Kutta matrix $A_0 = (a_{i,j})$ and weight vector $\mathbf{b}_0^T = (b_1, ..., b_s)^T$ are taken from the Butcher tableau of an s-stage Runge-Kutta method[30].

Here, we consider A-stable two-stage third-order SDIRK-23, L-stable two-stage second-order SDIRK-22, and L-stable three-stage third-order SDIRK-33 methods (see Appendix of[30] for coefficients), where SDIRK refers to singly diagonally implicit Runge-Kutta. Figures 5 – 7 depict the convergence bound (18) in the complex plane as a function of $h_t\kappa_\gamma$ over various ω_C for these methods, respectively. Overall, the L-stable schemes lead to significantly better MGRIT convergence bounds than the A-stable scheme, consistent with the discussion and results for unweighted relaxation in[30], and, more importantly, numerical results using weighted relaxation in Section 4.1.1. Additionally, note from Figure 5 that for unweighted relaxation ($\omega_C = 1$), two-level MGRIT is divergent in much of the complex plane (a known phenomenon[30]). However, applying under-relaxation with $\omega_C = 0.8$ restores reasonable convergence in much of the complex plane. This behavior is confirmed in practice in Section 4.1.2. Similarly, applying under-relaxation to L-stable SDIRK-33 in Figure 7 yields convergence, albeit slow, along the imaginary axis. Spatial eigenvalues on the imaginary axis are notoriously difficult for MGRIT to converge on, as can be seen with the theoretical bounds

for $\omega_C = 1$. To the best of our knowledge, backward Euler is the *only* one-step time-integration scheme that yields convergence on the imaginary axis.[4] Here, we see that weighted relaxation can yield convergence on higher-order integration schemes as well.



(a) $\omega_C = 0.5$ (b) $\omega_C = 0.6$ (c) $\omega_C = 0.7$

(d) $\omega_C = 0.8$ (e) $\omega_C = 1.0$ (f) $\omega_C = 1.2$

FIGURE 5 Two-level MGRIT theoretical convergence bound as a function of $\mathrm{Re}(h_t\kappa_\gamma)$ and $\mathrm{Im}(h_t\kappa_\gamma)$, for $m = 2$ and A-stable 2-stage SDIRK-23.

4 | RESULTS

This section demonstrates MGRIT with weighted relaxation on a 2D advection-diffusion problem and a nonlinear eddy current problem.

4.1 | 2D Convection-Diffusion with discontinuous Galerkin elements

To indicate generality of the proposed weighted relaxation scheme, we now consider the advection-diffusion problem

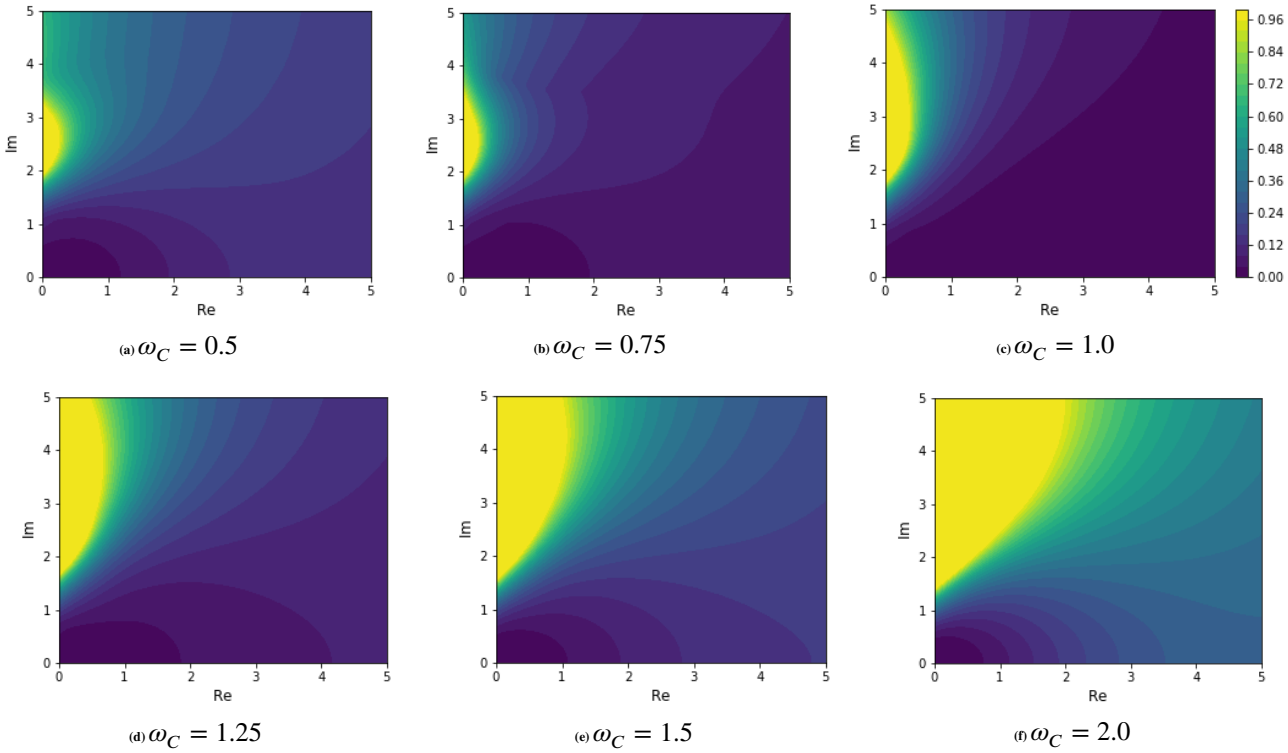
$$\frac{\partial u}{\partial t} + \mathbf{b}(t,\mathbf{x}) \cdot \nabla u - \epsilon \nabla \cdot \nabla u = 0, \quad \mathbf{x} \in \Omega, \quad t \in [0,T] \tag{24}$$

$$u(\mathbf{x},0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{25}$$

where $\epsilon > 0$ is the diffusion constant, Ω is a bounded convex domain in 2D, and the boundary conditions are periodic in space. The final time T is set to 20 and $\mathbf{b} = (\sqrt{2/3}, \sqrt{1/3})$. Letting $\mathbf{x} = (x_1, x_2)$, the initial condition is

$$u_0(\mathbf{x}) = \frac{1}{16}\,\mathrm{erfc}[w(x_1 - c_1 - r_1)]\,\mathrm{erfc}[-w(x_1 - c_1 + r_1)] \times \mathrm{erfc}[w(x_2 - c_2 - r_2)]\,\mathrm{erfc}[-w(x_2 - c_2 + r_2)],$$

[4] It is important to note that for unweighted relaxation, two-level convergence bounds are *necessary and sufficient* [29].



(a) $\omega_C = 0.5$ (b) $\omega_C = 0.75$ (c) $\omega_C = 1.0$

(d) $\omega_C = 1.25$ (e) $\omega_C = 1.5$ (f) $\omega_C = 2.0$

FIGURE 6 Two-level MGRIT theoretical convergence bound as a function of $\mathrm{Re}(h_t\kappa_\gamma)$ and $\mathrm{Im}(h_t\kappa_\gamma)$, for $m = 2$ and L-stable 2-stage SDIRK-22.

which defines a smooth rectangular hump with $\mathrm{erfc}(x)$ the complementary error function, $(c_1, c_2) = (0, -0.2)$, $(r_1, r_2) = (0.45, 0.25)$, and $w = 10$.

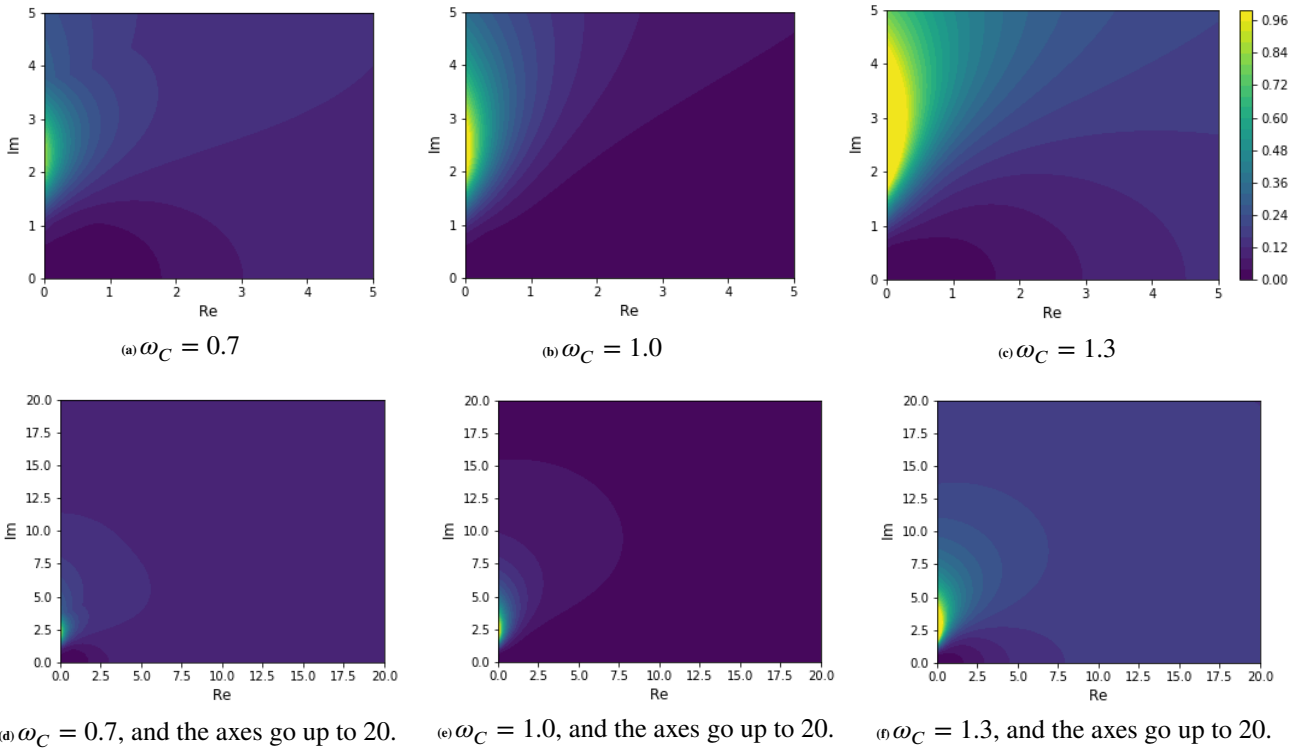
We use the MFEM library[33] to discretize over a regular quadrilateral grid on a hexagonal domain Ω, corresponding to the file `mfem/data/periodic-hexagon.mesh`. In space, we use Q_1 (bi-linear) or Q_3 (bi-cubic) discontinuous Galerkin (DG) elements with a standard upwind scheme for the advective term and the interior penalty (IP)[34] scheme for the diffusion term. In time, we consider backward Euler (L-stable), the A-stable two-stage third-order SDIRK-23 method, and the L-stable three-stage third-order SDIRK-33 method.

The numerical setup uses MGRIT V-cycles with a random initial guess and a residual halting tolerance of $10^{-10}/(h_x\sqrt{\delta_t})$. The iterations are capped at 125, with "125+" indicating that this maximum was reached. The value N_x represents the total number of spatial degrees-of-freedom, and grows by a factor of 4 each uniform refinement because space is now 2D. The number of time points grows by a factor of 2, so that $\delta_t/h_x = 0.477$ is fixed for all test problems, where h_x refers to the spatial mesh size. Regarding the diffusive term, the ratio δ_t/h_x^2 varies from 1.9245 for the smallest problem, to 15.396 on the largest problem, representing moderate ratios typical for an implicit scheme.

4.1.1 | Results for L-Stable Schemes

Tables 5 and 6 depict these results for the case of bilinear DG elements with backward Euler and bi-cubic DG elements with L-stable SDIRK-33, respectively. Three diffusion constants, $\epsilon = 0.1, 0.01$, and 0.001, are depicted to highlight the benefits of weighted relaxation for three different MGRIT convergence regimes. The first regime concerns sufficiently diffusive problems, where MGRIT convergence is bounded with growing problem size[27]. This is observed for the $\epsilon = 0.1$ case. For the next regime when $\epsilon = 0.01$, the problem is on the cusp of sufficient diffusiveness, as evidenced by the growing iteration counts for backward Euler in Table 5, but flat iteration counts in Table 6 for some weight values.[5] When $\epsilon = 0.001$, convergence is poor in both cases.

[5]Note that SDIRK-33 is a more favorable time-stepping scheme for MGRIT convergence and diffusive problems[27], thus it is not surprising that it provides better performance here. In fact, if these experiments are repeated with bi-cubic DG elements and backward Euler, the results are almost identical to Table 5 for bilinear DG elements and backward Euler, thus indicating that the use SDIRK-33 is the factor leading to the improved convergence.



(a) $\omega_C = 0.7$ (b) $\omega_C = 1.0$ (c) $\omega_C = 1.3$

(d) $\omega_C = 0.7$, and the axes go up to 20. (e) $\omega_C = 1.0$, and the axes go up to 20. (f) $\omega_C = 1.3$, and the axes go up to 20.

FIGURE 7 Two-level MGRIT theoretical convergence bound as a function of $\text{Re}(h_t\kappa_\gamma)$ and $\text{Im}(h_t\kappa_\gamma)$, for $m = 2$ and L-stable 3-stage SDIRK-33.

In all three regimes, the benefits of weighted relaxation can be observed and are similar to those benefits observed for the 1D model problems in the Supplemental Materials S2. For the first-order discretizations in Table 5, a weight choice of 1.6 is experimentally found to be best, saving 15%–20% of iterations, which aligns with the best weight choice for 1D advection in Appendix S2.[6] For the third-order discretizations in Table 6, a weight choice of 1.3 is experimentally found to be best, saving 10%–15% of iterations. This does not align with the best weight choice for 1D advection in Appendix S2, but instead aligns with the best weight choice for 1D diffusion. Thus, we can say that the simple 1D model problems from Appendix S2 provide a useful, but rough guide for choosing relaxation weights for more complicated problems. Lastly, we note that under-relaxation was not beneficial for these cases, as indicated by the $\omega_C = 0.7$ case.

4.1.2 | A-stable Results

Table 7 repeats the above experiments for the A-stable SDIRK-23 scheme with bi-cubic DG elements in space. We also consider larger ϵ (i.e., stronger diffusion) as this highlights the benefits of weighted-relaxation. Results for $\epsilon = 0.001$ are omitted because all test cases larger than the smallest took 125+ iterations. Weights larger than 1.0 are also omitted as they did not improve convergence.

Consistent with the discussion in Section 3.2, we find that under-relaxation ($\omega_C < 1.0$) is beneficial, with $\omega_C = 0.7$ providing the best performance. In fact, in most cases this under-relaxation even restores convergence compared with unweighted relaxation, where the 125+ label for $\omega_C = 1.0$ corresponds to a convergence rate larger than one. This divergence for $\omega_C = 1.0$ is not surprising, as the work[30] shows that A-stable schemes do not generally yield good MGRIT convergence and often lead to divergence, even for problems of a parabolic character.

Lastly, we compare Table 7 to the convergence plots in Figure 5. Convergence for $\omega_C = 0.7$ improves as the problem size increases. This is most likely due to increasing numerically diffusivity as the grid is refined, which results in the spectrum being pushed into the region of more rapid convergence close to the real axis in Figure 5. Additionally, overall performance degrades

[6] We note that while the tables only show a handful of weight choices, thorough experimentation with under- and over-relaxation using a weight step-size of 0.1 was done to find the experimentally best choices.

	$N_x \times N_t$	192×192	768×384	3072×768	12288×1536
$\epsilon = 0.001$	$\omega_C = 0.7$	29	39	56	125+
	1.0	25	32	47	65
	1.3	22	28	42	58
	1.6	29	38	40	52
	1.9	38	63	112	125+
$\epsilon = 0.01$	$\omega_C = 0.7$	28	34	45	53
	1.0	24	30	38	46
	1.3	21	27	32	41
	1.6	28	30	28	37
	1.9	37	58	81	76
$\epsilon = 0.1$	$\omega_C = 0.7$	16	19	21	23
	1.0	13	16	18	19
	1.3	12	14	16	17
	1.6	15	16	14	16
	1.9	24	29	26	26

TABLE 5 Multilevel MGRIT iterations for 2D advection-diffusion over various diffusion constants ϵ, with bilinear 1 DG elements, backward Euler in time, FCF-relaxation, and $m = 2$. For the cases labeled "125+", the solver is still diverging with a convergence rate over 1 at iteration 125.

	$N_x \times N_t$	768×192	3072×384	12288×768	49152×1536
$\epsilon = 0.01$	$\omega_C = 0.7$	32	31	29	29
	1.0	27	25	25	25
	1.3	25	22	22	22
	1.6	37	43	32	27
	1.9	52	66	73	68
$\epsilon = 0.1$	$\omega_C = 0.7$	11	10	10	10
	1.0	9	9	9	9
	1.3	9	8	8	8
	1.6	12	10	10	9
	1.9	19	17	18	15

TABLE 6 Multilevel MGRIT iterations for 2D advection-diffusion over various diffusion constants ϵ, with bi-cubic 3 DG elements, SDIRK-33 in time, FCF-relaxation, and $m = 2$. Results for $\epsilon = 0.001$ are omitted because all test cases larger than the smallest took 125+ iterations.

for larger ϵ, which is due to the spectrum being pushed out of the region of convergence (i.e., farther up the positive real axis) in Figure 5. Similarly, as ϵ decreases, the spectrum is pushed to the imaginary axis in Figure 5, and convergence eventually degrades, as is observed for $\epsilon = 0.001$. For this problem and time-discretization, MGRIT convergence is best for $\epsilon = 0.1$, and interestingly, the advection terms actually *help* MGRIT converge for this problem.

4.2 | Nonlinear Eddy Current Problem

The last example illustrates the performance of the new relaxation scheme for a nonlinear eddy current problem. The eddy current problem is an approximation of Maxwell's equations that is commonly used in the simulation of electrical machines, such as induction machines, transformers, or cables. Here, we consider a coaxial cable model. Let $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$ denote a 2D cross-section of the 3D cable model, as depicted in Figure 8.

	$N_x \times N_t$	192×192	768×384	3072×768	12288×1536
$\epsilon = 0.01$	$\omega_C = 0.6$	51	60	55	50
	0.7	47	54	49	45
	0.8	43	50	44	42
	1.0	43	85	125+	125+
$\epsilon = 0.1$	$\omega_C = 0.6$	38	38	32	27
	0.7	32	32	27	23
	0.8	36	47	47	42
	1.0	48*	96*	125+	125+
$\epsilon = 1.0$	$\omega_C = 0.6$	44	43	38	30
	0.7	38	38	33	26
	0.8	41	57	63	53
	1.0	48*	96*	125+	125+
$\epsilon = 100.0$	$\omega_C = 0.6$	52	59	60	59
	0.7	44	52	52	51
	0.8	44	66	90	98
	1.0	48*	96*	125+	125+

TABLE 7 Multilevel MGRIT iterations for 2D advection-diffusion over various diffusion constants ϵ, with bi-cubic DG elements, SDIRK-23 in time, FCF-relaxation, and $m = 2$. The asterisk * refers to convergence due only to the exactness property of FCF-relaxation, where FCF-relaxation reproduces sequential time-stepping in $(N_t - 1)/2m$ iterations[2]. For all cases labeled "125+", the solver is still diverging with a convergence rate over 1 at iteration 125.

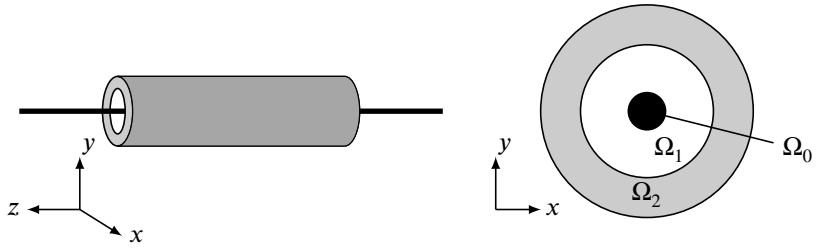


FIGURE 8 Coaxial cable model and its cross section. The inner, black region Ω_0 models the copper wire, the white region Ω_1 the air insulator and the outer, gray region Ω_2 the conducting shield[35].

For a voltage-driven system, the eddy current problem is coupled with an additional equation, resulting in the following system for unknown magnetic vector potential $A : \Omega \times (0, T] \to \mathbb{R}$ and the electric current $i_s : (0, T] \to \mathbb{R}$:

$$\sigma \partial_t A - \nabla \cdot \left(\nu \nabla A\right) - \chi_s i_s = 0, \tag{26}$$

$$\frac{d}{dt} \int_\Omega \chi_s \cdot A \, dV = v_s, \tag{27}$$

with homogeneous Dirichlet boundary condition $A = 0$ on $\partial\Omega$ and the initial value $A(\mathbf{x}, 0) = 0$, $\mathbf{x} \in \Omega$. The electrical conductivity $\sigma \geq 0$ is only non-zero in the tube region Ω_2 (here set to 10 MS/m), and the (isotropic, nonlinear) magnetic reluctivity $\nu(\mathbf{x}, |\nabla A|)$ is modeled by a vacuum $(1/\mu_0)$ in Ω_0 and Ω_1 and by a monotone cubic spline curve in Ω_2. The current distribution function $\chi_s : \Omega \to \mathbb{R}$ represents a stranded conductor in the model[36]. The relationship between the spatially integrated time derivative of the magnetic vector potential, called flux linkage, and the voltage v_s is modeled by Equation (27). The voltage is a pulsed voltage source, produced by comparing a reference wave with a triangular wave,

$$v_s(t) = 0.25 \text{sign}\left[r_s(t) - s_n(t)\right], \quad t \in (0, T],$$

with reference signal

$$r_s(t) = \sin\left(\frac{2\pi}{T}t\right)$$

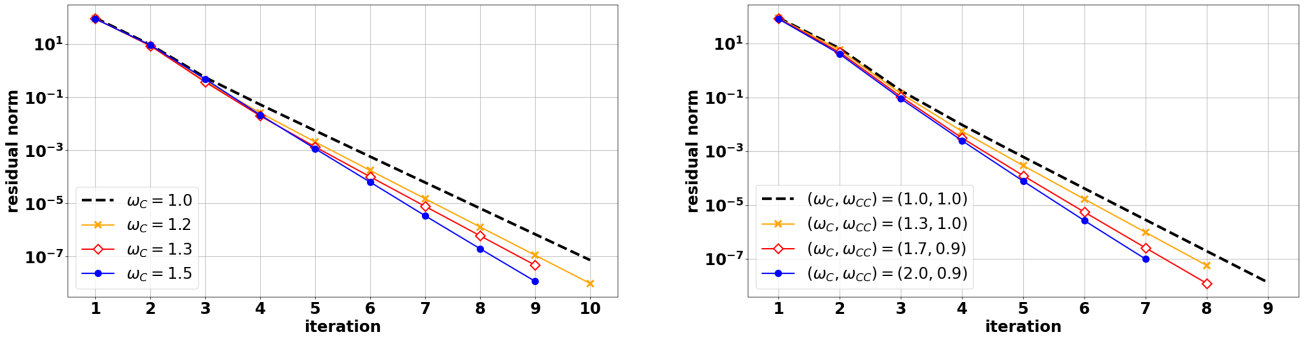
and bipolar trailing-edge modulation using a sawtooth carrier signal

$$s_n(t) = \frac{n}{T}t - \left\lfloor \frac{n}{T}t \right\rfloor,$$

with $n = 200$ teeth and electrical period $T = 0.02$ s[37].

We use linear edge shape functions with 2269 degrees of freedom in space to discretize (26)–(27). The resulting system of index-1 differential-algebraic equations (DAEs) is integrated on an equidistant time grid with 2^{14} intervals using the backward Euler method to resolve the pulses. For each time step t_j, we obtain a nonlinear system of the form $\Phi(\mathbf{u}_j) = \mathbf{g}_j$, with $\mathbf{u}_j^\top = (\mathbf{a}^\top, i)$ and where $\mathbf{a}$ is the vector of discrete vector potentials and i is an approximation of the current. Considering all time steps at once results in a space-time system of the form $\mathcal{A}(\mathbf{u}) = \mathbf{g}$, where each block row corresponds to one time step, i.e., the nonlinear extension of equation (3). This space-time system is solved using MGRIT V-cycles with a random initial guess, a residual halting tolerance of 10^{-7} and factor-4 coarsening ($m = 4$). The method is fully multilevel with the system on the coarsest grid consisting of four time points. For all spatial problems, Newton's method is used with a direct LU solver. For the experiments, we use the model `tube.fem` from the finite element package FEMM[35] and the Python framework PyMGRIT[38, 39].

Figure 9 shows MGRIT convergence for the eddy current problem and various relaxation weights for FCF- and FCFCF-relaxation[7]. The results show that non-unitary weights improve MGRIT convergence for both relaxation schemes. For this particular problem, the best weight choice for FCF-relaxation of $\omega_C = 1.5$ yields a saving of one iteration, or 10%, over a unitary weight choice. For degree-two relaxation, the experimentally optimal pair of weights $(\omega_C, \omega_{CC}) = (2.0, 0.9)$ even allows for a saving of two iterations, or 22%, over a unitary weight choice of $(\omega_C, \omega_{CC}) = (1.0, 1.0)$. Again, as for the 2D advection-diffusion problem, the benefits of weighted relaxation on MGRIT convergence for this problem are similar to the benefits observed for the 1D heat equation in Section 3.1. For FCF-relaxation, the best weight choice for 1D diffusion of $\omega_C = 1.3$ results in slightly slower convergence for the 2D eddy current problem, compared to the weight $\omega_C = 1.5$, but both weight choices allow for the same saving of one iteration over a unitary weight choice. For FCFCF-relaxation, the best weight choice of $(\omega_C, \omega_{CC}) = (2.0, 0.9)$ corresponds to the best weight choice for 1D diffusion. Thus again, the simple linear 1D model problem provides good guidance for choosing relaxation weights for a more complicated problem, particularly in choosing over- and/or under-relaxation. Lastly, comparing total runtimes of MGRIT with weighted FCF- and FCFCF-relaxation with the experimentally optimal weight choices of $\omega_C = 1.5$ and $(\omega_C, \omega_{CC}) = (2.0, 0.9)$, respectively, FCF-relaxation is about 4 % faster than FCFCF-relaxation. For this particular problem, MGRIT with weighted FCF-relaxation is the most efficient solver.



(a) FCF-relaxation (b) FCFCF-relaxation

FIGURE 9 Experimental MGRIT convergence using weighted FCF- (left) and FCFCF-relaxation (right), $m = 4$, and various relaxation weights ω_C and ω_{CC} for the eddy current problem.

[7] We note that also for this problem thorough experimentation with under- and over-relaxation using a weight step-size of 0.1 was done.

5 | CONCLUSIONS

In this work, we introduced the concept of weighted relaxation to MGRIT, which until now has used only unweighted relaxation. We derived a new convergence analysis for linear two-grid MGRIT with degree-1 weighted-Jacobi relaxation, and used this analysis to guide and explore the selection of relaxation weights. The theory was verified with simple numerical examples in Section 3, and the utility of weighted relaxation was demonstrated on more complex problems in Section 4, including a 2D advection-diffusion problem and a 2D nonlinear eddy current problem. The simple linear 1D model problems from Section 3.1 provide useful guidance when choosing relaxation weights for more complicated linear and nonlinear problems, and are intended in part to guide future weight choices.

With an appropriate choice of weight, the numerical results demonstrated that MGRIT with weighted relaxation consistently offers improved convergence rates and lower iteration counts when compared with standard (unweighted) MGRIT, at almost no additional computational work. In most cases, weighted relaxation yields a 10%–20% savings in iterations, while for the A-stable scheme, the results show that under-relaxation can *restore convergence* in some cases where unweighted relaxation does not converge.

ACKNOWLEDGMENTS

Los Alamos National Laboratory report number LA-UR-21-26114.

References

1. Gander MJ. 50 years of Time Parallel Time Integration. In: Carraro T, Geiger M, Körkel S, and Rannacher R, editors. Multiple Shooting and Time Domain Decomposition. Springer; 2015. p. 69–114.

2. Falgout RD, Friedhoff S, Kolev TV, MacLachlan SP, and Schroder JB. Parallel Time Integration with Multigrid. SIAM J Sci Comput. 2014;**36**(6):C635–C661.

3. Ong BW, and Schroder JB. Applications of Time Parallelization. Comput Vis Sci. 2020;**23**(1):1–15.

4. Nievergelt J. Parallel methods for integrating ordinary differential equations. Comm ACM. 1964;**7**:731–733.

5. Trottenberg U, Oosterlee C, and Schüller A. Multigrid. London, UK: Academic Press; 2001.

6. Briggs WL, Henson VE, and McCormick SF. A multigrid tutorial. 2nd ed. Philadelphia, PA, USA: SIAM; 2000.

7. Falgout RD, Manteuffel TA, O'Neill B, and Schroder JB. Multigrid reduction in time for nonlinear parabolic problems: A case study. SIAM Journal on Scientific Computing. 2017;**39**(5):S298–S322.

8. Falgout RD, Katz A, Kolev TV, Schroder JB, Wissink A, and Yang UM. 2015. *Parallel Time Integration with Multigrid Reduction for a Compressible Fluid Dynamics Application*. LLNL-JRNL-663416. Lawrence Livermore National Laboratory.

9. Christopher J, Gao X, Guzik SM, Falgout R, and Schroder J. Fully Parallelized Space-Time Adaptive Meshes for the Compressible Navier-Stokes Equations Using Multigrid Reduction in Time. Computing and Visualization in Science. 2020;**23**:1–19.

10. Hessenthaler A, Nordsletten D, Röhrle O, Schroder J, and Falgout R. Convergence of the multigrid-reduction-in-time algorithm for the linear elasticity equations. Numerical Linear Algebra with Applications. 2018;**25**(3):e2155.

11. Lecouvez M, Falgout RD, Woodward CS, and Top P. A parallel multigrid reduction in time method for power systems. Power and Energy Society General Meeting (PESGM). 2016;p. 1–5.

12. Günther S, Falgout RD, Top P, Woodward CS, and Schroder JB. Parallel-in-Time Solution of Power Systems with Unscheduled Events. Power and Energy Society General Meeting (PESGM). 2019;p. 1–5.

13. Friedhoff S, Hahne J, Kulchytska-Ruchka I, and Schöps S. Exploring parallel-in-time approaches for eddy current problems. In: Progress in Industrial Mathematics at ECMI 2018. Springer; 2019. p. 373–379.

14. Bolten M, Friedhoff S, Hahne J, and Schöps S. Parallel-in-time simulation of an electrical machine using MGRIT. Comput Vis Sci. 2020;**23**(1-4):Paper No. 14, 14.

15. Günther S, Ruthotto L, Schroder JB, Cyr EC, and Gauger NR. Layer-Parallel Training of Deep Residual Neural Networks. SIAM Journal on Data Science. 2019 (accepted);ArXiv preprint arXiv:1812.04352.

16. Cyr EC, Günther S, and Schroder JB. Multilevel Initialization for Layer-Parallel Deep Neural Network Training. International Journal of Computing and Visualization in Science and Engineering. 2021;**1**:1–9. ArXiv preprint arXiv:1912.08974.

17. Lions JL, Maday Y, and Turinici G. Résolution d'EDP par un schéma en temps pararéel. CRAcad Sci Paris Sér I Math. 2001;**332**:661–668.

18. Minion ML, and Williams SA. Parareal and Spectral Deferred Corrections. In: Simos TE, editor. Numerical Analysis and Applied Mathematics. AIP Conference Proceedings. AIP; 2008. p. 388–391.

19. Emmett M, and Minion ML. Toward an efficient parallel in time method for partial differential equations. Commun Appl Math Comput Sci. 2012;**7**(1):105–132.

20. Minion ML. A hybrid parareal spectral deferred corrections method. Comm App Math and Comp Sci. 2010;**5**(2):265–301.

21. Gander MJ, and Vandewalle S. Analysis of the parareal time-parallel time-integration method. SIAM Journal on Scientific Computing. 2007;**29**(2):556–578.

22. Bolten M, Moser D, and Speck R. A multigrid perspective on the parallel full approximation scheme in space and time. Numerical Linear Algebra with Applications. 2017;**24**(6):e2110.

23. Adams M, Brezina M, Hu J, and Tuminaro R. Parallel multigrid smoothing: polynomial versus Gauss-Seidel. J Comput Phys. 2003;**188**:593–610.

24. Baker AH, Falgout RD, and Yang TVKUM. Multigrid Smoothers for Ultraparallel Computing. SIAM J Sci Comput. 2011 Oct;**33**(5):2864–2887.

25. Brandt A. Multi–Level Adaptive Solutions to Boundary–Value Problems. Math Comp. 1977;**31**(138):333–390.

26. Friedhoff S, and MacLachlan S. A generalized predictive analysis tool for multigrid methods. Numerical Linear Algebra with Applications. 2015;**22**(4):618–647.

27. V Dobrev NAP Tz Kolev, and Schroder JB. Two-level Convergence Theory for Multigrid Reduction in Time (MGRIT). SIAM J Sci Comput. 2017;**39**(5):S501–S527.

28. Hessenthaler A, Southworth BS, Nordsletten D, Röhrle O, Falgout RD, and Schroder JB. Multilevel convergence analysis of multigrid-reduction-in-time. SIAM Journal on Scientific Computing. 2020;**42**:A771–A796.

29. Southworth BS. Necessary conditions and tight two-level convergence bounds for parareal and multigrid reduction in time. SIAM Journal on Matrix Analysis and Applications. 2019;**40**(2):564–608.

30. Friedhoff S, and Southworth BS. On "Optimal" h-independent convergence of Parareal and multigrid-reduction-in-time using Runge-Kutta time integration. Numerical Linear Algebra with Applications. 2020;p. e2301.

31. Widom H. On the singular values of Toeplitz matrices. Zeitschrift für Analysis und ihre Anwendungen. 1989;**8**(3):221–229.

32. Capizzano SS, and Tilli P. Extreme singular values and eigenvalues of non-Hermitian block Toeplitz matrices. Journal of computational and applied mathematics. 1999;**108**(1-2):113–130.

33. MFEM: Modular finite element methods library. http://mfem.org;.

34. Arnold DN, Brezzi F, Cockurn B, and Marini LD. Unified Analysis of Discontinuous Galerkin Methods for Elliptic Problems. SIAM J Numer Anal. 2002;**39**:1749–1779.

35. Meeker C DavidḞinite Element Method Magnetics, Version 4.2 (28Feb2018 Build);. Available from: http://www.femm.info.

36. Schöps S, Gersem HD, and Weiland T. Winding Functions in Transient Magnetoquasistatic Field-Circuit Coupled Simulations. The International Journal for Computation and Mathematics in Electrical and Electronic Engineering. 2013;(6):2063–2083. TEMF-Pub-DB TEMF002204.

37. Gander MJ, Kulchytska-Ruchka I, Niyonzima I, and Schöps S. A New Parareal Algorithm for Problems with Discontinuous Sources. SIAM Journal on Scientific Computing. 2019;(2):B375–B395.

38. Hahne J, and Friedhoff S. PyMGRIT: Multigrid-Reduction-in-Time in Python v1.0. Release 1.0; 2020. Available from: https://github.com/pymgrit/pymgrit.

39. Hahne J, Friedhoff S, and Bolten M. Algorithm 1016: PyMGRIT: A Python Package for the Parallel-in-Time Method MGRIT. New York, NY, USA: Association for Computing Machinery; 2021.

40. Sugiyama M. Optimal Relaxation Weights for Multigrid Reduction In Time (MGRIT). Dept. of Mathematics and Statistics, University of New Mexico; 2019. https://digitalrepository.unm.edu/math_etds/147.

41. Howse AJ, Sterck HD, Falgout RD, MacLachlan S, and Schroder J. Parallel-In-Time Multigrid with Adaptive Spatial Coarsening for The Linear Advection and Inviscid Burgers Equations. SIAM Journal on Scientific Computing. 2019;**41**(1):A538–A565.

□

APPENDIX

A BOUND WITH FCFCF-RELAXATION

The derivation of the theoretical convergence bound for weighted FCFCF-relaxation (degree-two weighted-Jacobi) is shown in this section. Remembering expression (13a), the error propagator for stand-alone weighted FCF-relaxation takes the form

$$\begin{aligned}&(I - S(S^TAS)^{-1}S^TA)(I - \omega_C R_I^T(R_IAR_I^T)^{-1}R_IA)(I - S(S^TAS)^{-1}S^TA)\\&= P(I - \omega_C\mathbf{A}_{\triangle})R_I.\end{aligned} \tag{A1}$$

Applying expression (A1) twice, once with weight ω_C and once with another weight ω_{CC}, the error propagator for stand-alone weighted FCFCF-relaxation can be expressed as

$$P(I - \omega_{CC}\mathbf{A}_{\triangle})(I - \omega_C\mathbf{A}_{\triangle})R_I. \tag{A2}$$

Combining the effect of FCFCF-relaxation (A2) with the previous two-level error propagator (9), yields the following two-level MGRIT error propagator for FCFCF-relaxation

$$\begin{aligned}&(I - PB_{\triangle}^{-1}R_IA)P(I - \omega_{CC}\mathbf{A}_{\triangle})(I - \omega_C\mathbf{A}_{\triangle})R_I\\&= P(I - \mathbf{B}_{\triangle}^{-1}\mathbf{A}_{\triangle})(I - \omega_{CC}\mathbf{A}_{\triangle})(I - \omega_C\mathbf{A}_{\triangle})R_I\end{aligned}. \tag{A3}$$

Simplifying the error propagator to consider only C-points yields

$$E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}} = (I - \mathbf{B}_{\triangle}^{-1}\mathbf{A}_{\triangle})(I - \omega_{CC}\mathbf{A}_{\triangle})(I - \omega_C\mathbf{A}_{\triangle}). \tag{A4}$$

Similar to Section 2.2.3, we next use the set of eigenvectors $\{v_\gamma\}$ and corresponding eigenvalues $\{\lambda_\gamma\}$ of Φ and $\{\mu_\gamma\}$ of $\Phi_{\triangle}$ to diagonalize $E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}$ with the block diagonal eigenvector matrix $\widetilde{U}$. The resulting matrix $\widetilde{E}^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}$ is Toeplitz with the

following asymptotic generating function,

$$\begin{aligned}\mathcal{F}(x) &:= (\lambda_\gamma^m - \mu_\gamma)\left[(1-\omega_{CC})(1-\omega_C)\sum_{\ell=1}^{\infty}\mu_\gamma^{\ell-1}e^{i\ell x} + \{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}\lambda_\gamma^m\sum_{\ell=2}^{\infty}\mu_\gamma^{\ell-2}e^{i\ell x} + \omega_{CC}\omega_C\lambda_\gamma^{2m}\sum_{\ell=3}^{\infty}\mu_\gamma^{\ell-3}e^{i\ell x}\right]\\ &= e^{ix}(\lambda_\gamma^m - \mu_\gamma)\left[(1-\omega_{CC})(1-\omega_C)\sum_{\ell=0}^{\infty}(\mu_\gamma e^{ix})^\ell + e^{ix}\{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}\lambda_\gamma^m\sum_{\ell=0}^{\infty}(\mu_\gamma e^{ix})^\ell + e^{i2x}\omega_{CC}\omega_C\lambda_\gamma^{2m}\sum_{\ell=0}^{\infty}(\mu_\gamma e^{ix})^\ell\right]\\ &= e^{ix}\frac{(\lambda_\gamma^m - \mu_\gamma)}{1-e^{ix}\mu_\gamma}\left[(1-\omega_{CC})(1-\omega_C) + e^{ix}\{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}\lambda_\gamma^m + e^{i2x}\omega_{CC}\omega_C\lambda_\gamma^{2m}\right]. \qquad \text{(A5)}\end{aligned}$$

Again following Section 2.2.3, we bound the maximum singular value of $E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}$ with

$$\begin{aligned}\sigma_{max,\gamma}(\widetilde{E}^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}) &\leq \max_{x\in[0,2\pi]}|\mathcal{F}(x)|\\ &= \max_{x\in[0,2\pi]}\frac{|\lambda_\gamma^m-\mu_\gamma|}{|1-e^{ix}\mu_\gamma|}|(1-\omega_{CC})(1-\omega_C)+e^{ix}\{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}\lambda_\gamma^m+e^{i2x}\omega_{CC}\omega_C\lambda_\gamma^{2m}|.\end{aligned} \qquad \text{(A6)}$$

Next by taking the maximum over γ, we have the following result, similar to Theorem 1,

$$||E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}||_{(\widetilde{U}\widetilde{U}^*)^{-1}} \leq \max_\gamma\max_{x\in[0,2\pi]}\frac{|\lambda_\gamma^m-\mu_\gamma|}{|1-e^{ix}\mu_\gamma|}|(1-\omega_{CC})(1-\omega_C)+e^{ix}\{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}\lambda_\gamma^m+e^{i2x}\omega_{CC}\omega_C\lambda_\gamma^{2m}|.$$

Finally, the approximation of the maximum over x yields the theoretical convergence bound for weighted FCFCF-relaxation given in equation (20),

$$\begin{aligned}||E^{FCFCF}_{\triangle,\{\omega_C,\omega_{CC}\}}||_{(\widetilde{U}\widetilde{U}^*)^{-1}} &\lessapprox \max_\gamma\frac{|\lambda_\gamma^m-\mu_\gamma|}{1-|\mu_\gamma|}|(1-\omega_{CC})(1-\omega_C)+\{\omega_{CC}(1-\omega_C)+\omega_C(1-\omega_{CC})\}|\lambda_\gamma^m|+\omega_{CC}\omega_C|\lambda_\gamma^{2m}||\\ &= \max_\gamma\frac{|\lambda_\gamma^m-\mu_\gamma|}{1-|\mu_\gamma|}|1-\omega_C+\omega_C|\lambda_\gamma^m||\,|1-\omega_{CC}+\omega_{CC}|\lambda_\gamma^m||. \qquad \text{(A7)}\end{aligned}$$

SUPPLEMENTAL MATERIALS

S1 MAX OVER X

Here we derive a closed form for the maximum over x that arises in theoretical bounds to allow easier computation. Consider

$$\max_{x\in[0,2\pi]} \frac{|\lambda^k-\mu|}{1-e^{ix}\mu}|1-\omega+e^{ix}\omega\lambda^k|. \tag{A8}$$

This function is not differentiable due to the absolute values, but the maximum is obtained at the same x if we square the underlying function. Noting that for complex f, $|f|^2 = ff^*$; thus, consider

$$\begin{aligned}
&\max_{x\in[0,2\pi]} |\lambda^k-\mu|^2 \frac{(1-\omega+e^{ix}\omega\lambda^k)(1-\omega+e^{-ix}\omega(\lambda^*)^k)}{(1-e^{ix}\mu)(1-e^{-ix}\mu^*)} \\
&\quad = |\lambda^k-\mu|^2 \max_{x\in[0,2\pi]} \frac{(1-\omega+e^{ix}\omega\lambda^k)(1-\omega+e^{-ix}\omega(\lambda^*)^k)}{(1-e^{ix}\mu)(1-e^{-ix}\mu^*)} \\
&\quad = |\lambda^k-\mu|^2 \max_{x\in[0,2\pi]} \frac{(\omega-1)^2+\omega^2|\lambda^k|^2-2\omega(\omega-1)\,\mathrm{Re}(\lambda^k)\cos(x)+2\omega(\omega-1)\,\mathrm{Im}(\lambda^k)\sin(x)}{1+|\mu|^2-2\,\mathrm{Re}(\mu)\cos(x)+2\,\mathrm{Im}(\mu)\sin(x)} \\
&\quad := |\lambda^k-\mu|^2 \max_{x\in[0,2\pi]} \frac{C_\lambda-2a\cos(x)+2b\sin(x)}{C_\mu-2c\cos(x)+2d\sin(x)}.
\end{aligned} \tag{A9}$$

Note that by assumption $|\mu|<1$, which implies $|1-|\mu||>0$, and the denominator of (A9) is necessarily nonzero. Thus the function we are maximizing is well-defined at all x (i.e., has non zero denominator). To find the maximum, we differentiate in x, where

$$\frac{\partial}{\partial x}\frac{C_\lambda-2a\cos(x)+2b\sin(x)}{C_\mu-2c\cos(x)+2d\sin(x)} = \frac{2\sin(x)(aC_\mu-cC_\lambda)+2\cos(x)(bC_\mu-dC_\lambda)+4(ad-bc)}{(C_\mu-2c\cos(x)+2d\sin(x))^2}.$$

To set the derivative equal to zero, we only need to worry about the numerator, so we seek x such that

$$\sin(x)(aC_\mu-cC_\lambda)+\cos(x)(bC_\mu-dC_\lambda)+2(ad-bc)=0. \tag{A10}$$

Note if $\omega=1$ (unweighted relaxation),

$$ad-bc=\omega(1-\omega)(\mathrm{Re}(\lambda^k)\,\mathrm{Im}(\mu)-\mathrm{Im}(\lambda^k)\,\mathrm{Re}(\mu))=0, \tag{A11}$$

in which case we can directly compute the solution x_0 to (A10) via the arctangent. The perturbation term in (A11) arises for $\omega\neq 1$. If μ and λ^k have the same angle in the complex plane (i.e., $\mu = C\lambda^k$ for some constant C), (A11) is also zero, and we arrive at the same solution x_0 as when $\omega=1$. More generally, we need to account for the case that μ and λ^k are not the same direction in the complex plane. *Mathematica* provides the root as

$$\begin{aligned}
x_0 &:= 2\arctan\left(\frac{aC_\mu-cC_\lambda\pm\sqrt{a^2C_\mu^2-4a^2d^2+8abcd-2acC_\lambda C_\mu-4b^2c^2+b^2C_\mu^2-2bdC_\lambda C_\mu+c^2C_\lambda^2+d^2C_\lambda^2}}{-2(ad-bc)+bC_\mu-dC_\lambda}\right) \\
&= 2\arctan\left(\frac{aC_\mu-cC_\lambda\pm\sqrt{(aC_\mu-cC_\lambda)^2+(bC_\mu-dC_\lambda)^2-4(ad-bc)^2}}{-2(ad-bc)+bC_\mu-dC_\lambda}\right).
\end{aligned} \tag{A12}$$

Now we want to evaluate (A9) at our maximum, x_0. Note that the maximum in (A12) takes the form $x_0 = 2\arctan(r)$ for a certain r, and recall the identities

$$\cos(2\arctan(r)) = \frac{1-r^2}{1+r^2}, \qquad \sin(2\arctan(r)) = \frac{2r}{1+r^2}.$$

Then from (A9),

$$\begin{aligned}\frac{C_\lambda - 2a\cos(2\arctan(r)) + 2b\sin(2\arctan(r))}{C_\mu - 2c\cos(2\arctan(r)) + 2d\sin(2\arctan(r))} &= \frac{C_\lambda - \frac{2a(1-r^2)}{1+r^2} + \frac{4br}{1+r^2}}{C_\mu - \frac{2c(1-r^2)}{1+r^2} + \frac{4dr}{1+r^2}} \\ &= \frac{C_\lambda(1+r^2) - 2a(1-r^2) + 4br}{C_\mu(1+r^2) - 2c(1-r^2) + 4dr} \\ &= \frac{(C_\lambda + 2a)r^2 + 4br + C_\lambda - 2a}{(C_\mu + 2c)r^2 + 4dr + C_\mu - 2c}. \end{aligned} \tag{A13}$$

Thus to compute the bound in (A8), we first evaluate r from (A12),

$$r := \frac{aC_\mu - cC_\lambda \pm \sqrt{(aC_\mu - cC_\lambda)^2 + (bC_\mu - dC_\lambda)^2 - 4(ad-bc)^2}}{-2(ad-bc) + bC_\mu - dC_\lambda}, \tag{A14}$$

where

$$\begin{aligned} a &= \omega(\omega-1)\,\mathrm{Re}(\lambda^k), \\ b &= \omega(\omega-1)\,\mathrm{Im}(\lambda^k), \\ c &= \mathrm{Re}(\mu), \\ d &= \mathrm{Im}(\mu), \\ C_\mu &= 1 + |\mu|^2 = 1 + c^2 + d^2, \\ C_\lambda &= (\omega-1)^2 + \omega^2|\lambda^k|^2 = (\omega-1)^2 + \omega^2(\mathrm{Re}(\lambda^k)^2 + \mathrm{Im}(\lambda^k)^2). \end{aligned}$$

We then plug r into (A13) and take the square root to map from (A9) to (A8).

S2 ONE-DIMENSIONAL MODEL PROBLEM RESULTS

This section thoroughly examines weighted-relaxation and MGRIT for three model problems, the 1D heat equation, the 1D advection equation with purely imaginary spatial eigenvalues, and the 1D advection equation with complex spatial eigenvalues. For full multilevel experiments, V-cycles are used and we coarsen down to a grid of size 4 or less in time. During searches in the weight-space for experimentally optimal weights, we use a step size of 0.1. Other testing parameters are discussed below on a case-by-case basis.

Regarding notation, we introduce a level subscript to allow for level-dependent weights, i.e., $\omega_{C,\ell=k}$ is the weight used on level k. If the level subscript is omitted, then the weight is uniform across all levels. For example, $\omega_{C,\ell=0}$ represents the relaxation weight for the first application of C-relaxation on the finest level 0, and $\omega_{CC,\ell=1}$ represents the relaxation weight for the second application of C-relaxation (degree two weighted-Jacobi) on the first coarse level 1.

S2.1 One-dimensional heat equation

We consider the one-dimensional heat equation subject to an initial condition and homogeneous Dirichlet boundary conditions,

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial^2 u}{\partial x^2} &= f(x,t), \quad \alpha > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T], \\ u(x,0) &= u_0(x), \quad x \in \Omega, \\ u(x,t) &= 0, \quad x \in \partial\Omega, \quad t \in [0, T]. \end{aligned} \tag{B15}$$

We transform the model problem to a system of ODEs of the form (1) by using second-order central differencing for discretizing the spatial derivative and then a standard one-step method (backward Euler) of the form (2) for discretizing the time derivative. We call this the *Backward Time, Central Space* or BTCS scheme, which yields

$$\mathbf{u}_j = (I - \delta_t G)^{-1}\mathbf{u}_{j-1} + (I - \delta_t G)^{-1}\delta_t \mathbf{f}_j, \quad j = 1, 2, ..., N_t, \tag{B16}$$

where the linear operator G in (1) is the three-point stencil $\frac{\alpha}{h_x^2}[1, -2, 1]$. In the form of (2), $\Phi = (I - \delta_t G)^{-1}$ and $\mathbf{g}_j = (I - \delta_t G)^{-1}\delta_t \mathbf{f}_j$. The eigenvalues of Φ and Φ^m are computed using the eigenvalues of G, i.e.,

$$\kappa_\gamma = -\frac{4}{h_x^2}\sin^2\left(\frac{\gamma\pi}{2(N_x+1)}\right),$$

for $\gamma = 1, 2, ..., N_x$, which in turn allows for the computation of the theoretical convergence estimate (19). For more details on our computation of κ_γ, see the work[27].

The following functions with the given domains are used for numerical experiments,

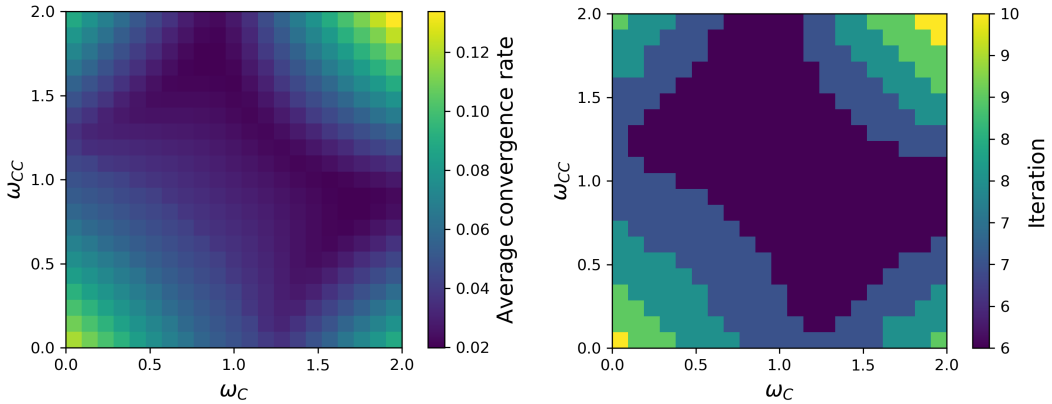
$$\begin{aligned} u(x,t) &= \sin(\pi x)\cos(t), \\ f(x,t) &= \sin(\pi x)[\sin(t) - \pi^2\cos(t)], \\ \alpha &= 1, \quad x \in [0,1], \quad t \in [0, 0.625]. \end{aligned}$$

The residual norm halting tolerance for MGRIT is set to $10^{-10}/\sqrt{h_x\delta_t}$. Reported convergence rates are taken as an average over the last 5 MGRIT iterations, where $\|r_k\|_2/\|r_{k-1}\|_2$ is the convergence rate at iteration k and r_k is the residual from equation (3) at iteration k. The combination of grid points in space N_x and time N_t are chosen so that a $\frac{\delta_t}{h_x^2} = 12.8$. This value was chosen to be of moderate magnitude and consistent with other MGRIT literature, namely the work[27].

S2.1.1 Weighted FCF- and FCFCF-relaxation

We start by considering the two-level method for weighted FCF- and FCFCF-relaxation, i.e., degree-one and degree-two relaxation, respectively. Here, the search for the experimentally optimal pair of weights for FCFCF-relaxation and $m = 2$ is depicted in Figure S1, where $(\omega_C, \omega_{CC}) = (1.7, 0.9)$ is the point corresponding to the minimal experimental convergence rate. The search space of possible weights is $0 \leq \omega_C, \omega_{CC} \leq 2.0$, and is based on a more expansive preliminary search. A similar study was done in the thesis[40] for FCF-relaxation and found that $\omega_C = 1.3$ is the point where the minimal convergence rate is reached.

Table S1 depicts the convergence rate and iterations for the two-level case. Each table entry is formatted as *convergence rate (iterations)*. The experimentally optimal weights for FCFCF-relaxation $(\omega_C, \omega_{CC}) = (1.7, 0.9)$, found using $(N_x, N_t) = (291, 4097)$ and $m = 2$ above, is highlighted in bold. This weight choice leads to a saving of 1 MGRIT iteration, or 16%, over unitary weights and FCFCF-relaxation on the largest problem. The best weight choice for FCF-relaxation of $\omega_C = 1.3$ yields a saving of 1 iteration, or 14%,



(a) Convergence Rate (b) Iterations

FIGURE S1 Two-level MGRIT experimental convergence rates (left) and iteration counts (right) using FCFCF-relaxation and various relaxation weights ω_C and ω_{CC} for the 1D heat equation, coarsening factor $m = 2$, and grid sizes $(N_x, N_t) = (291, 4097)$.

over a unitary weight choice (i.e., $\omega_C = 1.0$) on the largest problem. At the bottom of the table, we examine whether the experimentally optimal weights for FCF- and FCFCF-relaxation carry over to another coarsening factor choice, $m = 16$, and find that this is largely the case.

Table S2 repeats these experiments for a full multilevel method. We see that the best two-level choice for FCFCF-relaxation of $(1.7, 0.9)$ still performs well, but no longer yields the fastest convergence. Another search of the weight-space for the multilevel case yielded the experimentally optimal pair of weights $(\omega_C, \omega_{CC}) = (2.0, 0.9)$ when $m = 2$, which allows for saving 1 iteration. The uniform weight choice of $\omega_C = 1.3$ for FCF-relaxation continues to save 1 iteration.

Regarding cost, we can say that the cost of relaxation is the dominant cost of each V-cycle[2], thus a V-cycle with $m = 2$ and FCFCF-relaxation has a cost of about 1.66× when compared to a V-cycle using FCF-relaxation. Furthermore, we can then say that the use of weighted relaxation with FCF-relaxation is the most efficient solver depicted, as the number of iterations (8) for the largest problem size in Table S2 and weighted FCF-relaxation is noticeably less than 1.66 times the number of iterations for weighted FCFCF-relaxation ($1.66 * 6 \approx 10$).

S2.1.2 Multilevel weights for C-relaxation

We now consider the effect of level-dependent FCF-relaxation weights on MGRIT. Weighted FCFCF-relaxation is not considered because it is not as efficient as FCF, as discussed in Section S2.1.1, and the search space quickly becomes prohibitive. Thus, the search for the experimentally optimal pair of weights for three-level MGRIT with FCF-relaxation and $m = 2$ is depicted in Figure S2, where $(\omega_{C,\ell=0}, \omega_{C,\ell=1}) = (1.0, 2.0)$ is the point corresponding to the minimal convergence rate.

Next, we move to a four-level method while keeping fixed the experimentally optimal weights found in Figure S2 and search only for the weight on level three (the second coarse grid), $\omega_{C,\ell=2}$. The search for $\omega_{C,\ell=2}$ is depicted in Figure S3, and the trio of experimentally optimal weights is found to be $(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 2.0, 1.7)$ when $m = 2$.

Table S3 depicts the convergence rate and iterations for level dependent weights, comparing the experimentally "best" choice of $(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 2.0, 1.7)$ against unitary weights and the best uniform weight choice of $\omega_C = 1.3$. Level dependent weights provide only a very modest improvement in convergence rate with $m = 2$ and no benefit in iteration count over the best uniform weight choice of $\omega_C = 1.3$. Additionally, the selected level dependent weights do not translate to improved performance for $m = 16$,

	$N_x \times N_t$	291×4097	411×8193	581×16385	821×32769
$m = 2$	$\omega_C = 1.0$	0.049 (7)	0.048 (7)	0.039 (7)	0.039 (7)
	1.3	0.036 (7)	0.036 (7)	0.034 (6)	0.034 (6)
	1.5	0.048 (7)	0.049 (7)	0.049 (7)	0.049 (7)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.029 (6)	0.029 (6)	0.029 (6)	0.028 (6)
	$(1.3, 1.0)$	0.025 (6)	0.024 (6)	0.024 (6)	0.023 (6)
	(1.7, 0.9)	0.020 (6)	0.020 (6)	0.019 (6)	0.016 (5)
	$(2.0, 0.9)$	0.023 (6)	0.023 (6)	0.023 (6)	0.023 (6)
$m = 16$	$\omega_C = 1.0$	0.101 (9)	0.099 (8)	0.099 (8)	0.099 (8)
	1.3	0.074 (8)	0.075 (8)	0.075 (8)	0.074 (8)
$m = 16$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.056 (7)	0.060 (7)	0.060 (7)	0.060 (7)
	$(1.3, 1.0)$	0.049 (7)	0.053 (7)	0.053 (7)	0.053 (7)
	$(1.7, 0.9)$	0.041 (6)	0.042 (6)	0.041 (6)	0.040 (6)
	$(2.0, 0.9)$	0.042 (6)	0.042 (6)	0.042 (6)	0.042 (6)

TABLE S1 Two-level MGRIT convergence rates (iterations) for the 1D heat equation and weighted FCF- and FCFCF-relaxation.

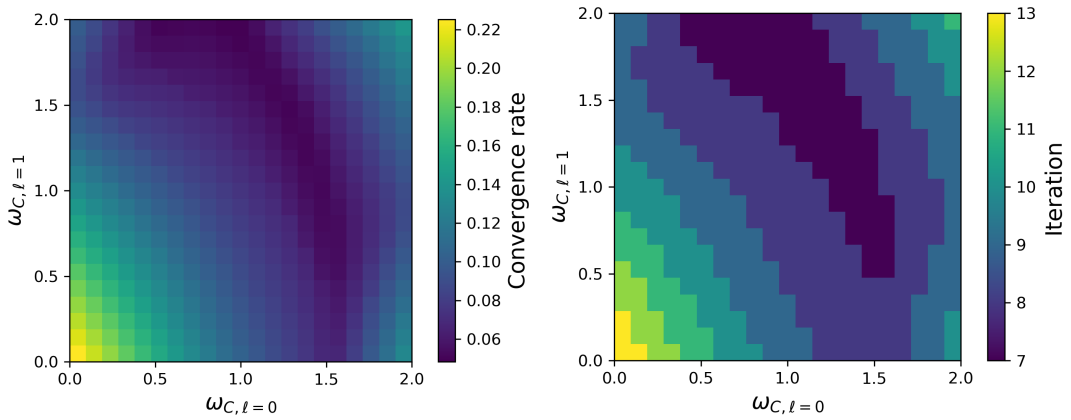
	$N_x \times N_t$	291×4097	411×8193	581×16385	821×32769
$m = 2$	$\omega_C = 1.0$	0.118 (9)	0.121 (9)	0.123 (9)	0.125 (9)
	1.3	0.092 (8)	0.095 (8)	0.096 (8)	0.096 (8)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.065 (7)	0.066 (7)	0.067 (7)	0.068 (7)
	$(1.3, 1.0)$	0.057 (7)	0.058 (7)	0.059 (7)	0.059 (7)
	$(1.7, 0.9)$	0.048 (7)	0.049 (7)	0.049 (7)	0.049 (7)
	(2.0, 0.9)	0.032 (6)	0.032 (6)	0.032 (6)	0.032 (6)
$m = 16$	$\omega_C = 1.0$	0.101 (9)	0.099 (8)	0.098 (8)	0.098 (8)
	1.3	0.071 (8)	0.068 (7)	0.067 (7)	0.067 (7)
$m = 16$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.056 (7)	0.060 (7)	0.060 (7)	0.060 (7)
	$(1.3, 1.0)$	0.048 (7)	0.053 (7)	0.052 (7)	0.052 (7)
	$(1.7, 0.9)$	0.037 (6)	0.040 (6)	0.039 (6)	0.038 (6)
	$(2.0, 0.9)$	0.041 (6)	0.041 (6)	0.041 (6)	0.041 (6)

TABLE S2 Multilevel MGRIT convergence rates (iterations) for the 1D heat equation and weighted FCF- and FCFCF-relaxation.

as shown at the bottom of the table. Thus, we conclude that level independent weights for problems similar to the heat equation are likely sufficient.

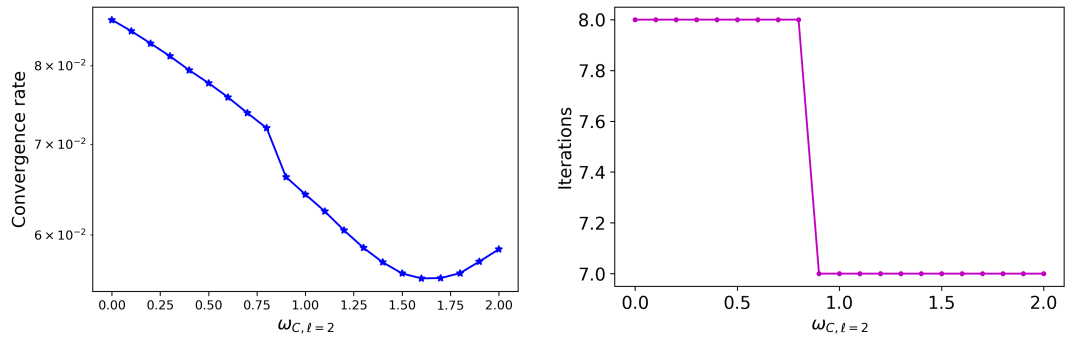
S2.1.3 Varying δ_t experiment

Lastly, for the one-dimensional heat equation, we explore the question of why weighted relaxation offers a significantly larger convergence benefit for multilevel MGRIT than for two-level MGRIT (compare Tables S1 and S2). In particular, we are interested if the progressively larger δ_t on coarse grids drives the improved performance for weighted relaxation in a multilevel setting. Thus, Table S4 depicts the two-level MGRIT convergence rate for various fine-grid δ_t values that mimic the δ_t values encountered with $m = 2$ on coarse MGRIT levels, when a final time of 0.625 is used and $N_t = 16385$ (i.e., the largest problem from Tables S1 and S2). To further mimic the coarse levels in MGRIT, N_t adapts with δ_t, so that the final time is



(a) Convergence Rate (b) Iterations

FIGURE S2 Three-level MGRIT experimental convergence rates (left) and iteration counts (right) using level-dependent FCF-relaxation weights $\omega_{C,\ell=0}$ and $\omega_{C,\ell=1}$ for the 1D heat equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (291, 4097)$.



(a) Convergence Rate (b) Iterations

FIGURE S3 Four-level MGRIT convergence rates (left) and iteration counts (right) using FCF-relaxation, as we search for the best level-three relaxation weight $\omega_{C,\ell=2}$, with the fixed values of $(\omega_{C,\ell=0}, \omega_{C,\ell=1}) = (1.0, 2.0)$ on the first two levels. The problem is the 1D heat equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (291, 4097)$.

	$N_x \times N_t$	291×4097	411×8193	581×16385	821×32769
$m = 2$	$(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 1.0, 1.0)$	0.090 (8)	0.090 (8)	0.090 (8)	0.090 (8)
	(1.0, 2.0, 1.7)	0.056 (7)	0.056 (7)	0.056 (7)	0.056 (7)
	(1.3, 1.3, 1.3)	0.069 (8)	0.069 (8)	0.063 (7)	0.062 (7)
$m = 16$	$(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 1.0, 1.0)$	0.101 (9)	0.099 (8)	0.098 (8)	0.098 (8)
	(1.0, 2.0, 1.7)	0.087 (8)	0.086 (8)	0.087 (8)	0.087 (8)
	(1.3, 1.3, 1.3)	0.071 (8)	0.068 (7)	0.067 (7)	0.067 (7)

TABLE S3 Four-level MGRIT convergence rates (iterations) for the 1D heat equation with level-dependent weights.

unchanged, e.g., when δt has been multiplied by 16 in Table S4, N_t decreases by a factor of 16 from 4096 to 256. However, as evidenced in the table, no MGRIT dependence on δ_t for weighted-relaxation is found, so we conclude that a more complication multilevel interaction is driving the improved benefit of weighted-relaxation in the multilevel case.

δ_t	$3.81e^{-5}$	$2 \cdot 3.81e^{-5}$	$4 \cdot 3.81e^{-5}$	$8 \cdot 3.81e^{-5}$	$16 \cdot 3.81e^{-5}$
Iterations	6	7	7	7	7
Convergence Rate	0.034	0.036	0.036	0.036	0.036

TABLE S4 Two-level MGRIT with $\omega_C = 1.3$ and $m = 2$ for various fine-grid δ_t values for the 1D heat equation.

S2.2 One-dimensional linear advection equation with purely imaginary spatial eigenvalues

We now consider the one-dimensional linear advection equation subject to an initial condition and periodic boundary conditions,

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial u}{\partial x} = 0, \quad \alpha > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T], \\ u(x, 0) = u_0(x), \quad x \in \Omega, \\ u(0, t) = u(L, t), \quad t \in [0, T]. \end{aligned} \tag{B18}$$

If we apply the BTCS scheme, we obtain

$$\mathbf{u}_j = (I - \delta_t G)^{-1} \mathbf{u}_{j-1}, \quad j = 1, 2, ..., N_t,$$

where the linear operator G from (1) is the two-point stencil $\frac{\alpha}{2h_x}[-1, 0, 1]$. Here, $\Phi = (I - \delta_t G)^{-1}$ and $\mathbf{g}_j = 0$. Similar to the heat equation, the eigenvalues of Φ and Φ^m are computed from the eigenvalues of G, i.e.,

$$\kappa_\gamma = \frac{i}{h_x} \sin\left(\frac{2\pi\gamma}{N_x}\right),$$

for $\gamma = 1, 2, ..., N_x$, which in turn allows for the computation of the theoretical convergence estimate (19). The following function with the given domain is used for numerical experiments,

$$u(x, t) = e^{-25((x-t)-0.5)^2}, \tag{B19a}$$

$$\alpha = 1, \quad x \in [0, 1], \quad t \in [0, 1]. \tag{B19b}$$

The function is chosen as a standard test problem that satisfies the spatially periodic boundary conditions. The MGRIT residual norm halting tolerance is set to $10^{-8}/\sqrt{h_x \delta_t}$ and the maximum allowed iterations is set to 70, because some cases will fail to quickly converge. Reported convergence rates are taken as $(\|r_k\|_2/\|r_0\|_2)^{1/k}$ at the final iteration k, where r_i is the residual from equation (3) at iteration i. The combination of grid points in space N_x and time N_t are chosen so that $\frac{\delta_t}{h_x} = 0.5$.

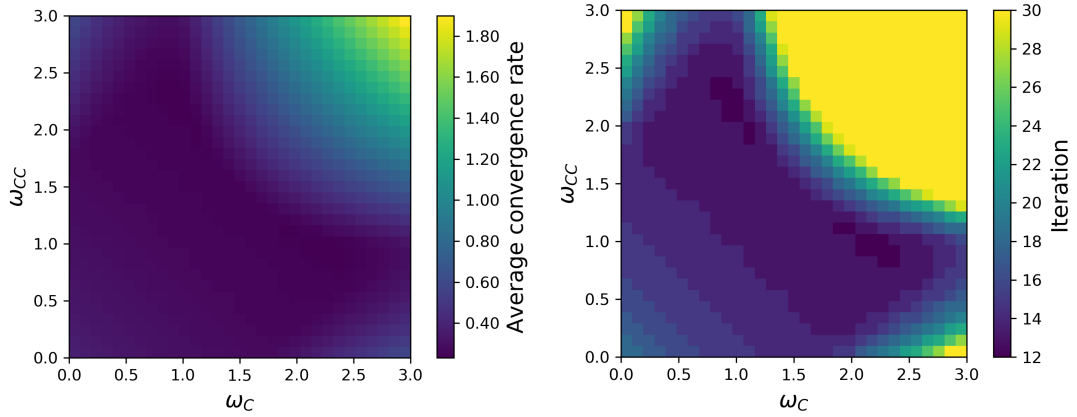
S2.2.1 Weighted FCF- and FCFCF-relaxation

We again start by considering the two-level method for weighted FCF- and FCFCF-relaxation. The search for the experimentally optimal pair of weights for FCFCF-relaxation and $m = 2$ is depicted in Figure S4, where $(\omega_C, \omega_{CC}) = (1.0, 2.3)$ is the point corresponding to the minimal convergence rate. The search space of weights is widened to $0 \leq \omega_C, \omega_{CC} \leq 3$, because a more expansive preliminary search indicated this was a reasonable range. A similar study was done in the thesis[40] for FCF-relaxation and found that $\omega_C = 1.8$ is the point where the minimal convergence rate is reached.

Table S5 depicts the convergence rate and iterations for the two-level case. The experimentally optimal pair of weights for FCFCF-relaxation $(\omega_C, \omega_{CC}) = (1.0, 2.3)$, found in Figure S4, is highlighted in bold, and this choices leads to saving 1 iteration, or 7% over unitary weights and FCFCF-relaxation on the largest problem. The best weight choice for FCF-relaxation of $\omega_C = 1.8$ yields a saving of 1 iteration, or 7%, over a unitary weight choice on the largest problem. At the bottom of the table, we examine whether the experimentally optimal weights carry over to another coarsening factor, $m = 4$, and find that this is not the

case, in contrast to the heat equation. MGRIT for advection problems is typically sensitive to changes in m (as opposed to the heat equation)[27, 41], hence we do not consider $m = 16$ or other large coarsening factors. Table S6 repeats these experiments for a full multilevel method. We see that the best two-level choice for FCFCF-relaxation of $(\omega_C, \omega_{CC}) = (1.0, 2.3)$ fails to provide a benefit for larger problems in the multilevel setting. Thus, we carry out another search in the weight-space and find that $(\omega_C, \omega_{CC}) = (2.3, 0.6)$ (in bold) yields the fastest convergence when $m = 2$, saving 25% of the iterations over unitary weights $(\omega_C, \omega_{CC}) = (1.0, 1.0)$ on the largest problem. A search in the weight-space for FCF-relaxation yielded the best convergence rate when $\omega_C = 1.5$, saving 22% of the iterations on the second largest problem. At the bottom of the table, we show that the best weight choices for $m = 2$ do not carry over to $m = 4$. The choice of $\omega_C = 1.4$ for FCF-relaxation is depicted to illustrate the performance for the best weight choice found in that case.

Overall, we note that linear advection is traditionally difficult for MGRIT[27, 41], so while these iteration counts with experimentally optimal weights are not scalable, we view any significant improvement in convergence as an important step.



(a) Convergence Rate (b) Iterations

FIGURE S4 Two-level MGRIT experimental convergence rates (left) and iteration counts (right) using FCFCF-relaxation and various relaxation weights ω_C and ω_{CC} for the 1D linear advection equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (1025, 1025)$.

S2.2.2 Multilevel weights for C-relaxation

We again consider the effect of level-dependent FCF-relaxation weights on MGRIT, similar to the heat equation. Weighted FCFCF-relaxation is again not considered due to its cost and size of search space. Thus, the search for the experimentally optimal pair of weights for three-level MGRIT with FCF-relaxation and $m = 2$ is depicted in Figure S5, where $(\omega_{C,\ell=0}, \omega_{C,\ell=1}) = (1.3, 2.0)$ is the point corresponding to the minimal convergence rate.

Next, we move to a four-level method while keeping fixed the experimentally optimal weights found in Figure S5 and search only for the weight on level three (the second coarse grid), $\omega_{C,\ell=2}$. This search is depicted in Figure S6 and the trio of experimentally optimal weights is found to be $(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.3, 2.0, 1.7)$ when $m = 2$.

Table S7 depicts the convergence rate and iterations for level dependent weights, comparing the experimentally "best" choice of $(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.3, 2.0, 1.7)$ against unitary weights and the best uniform weight choice of $\omega_C = 1.5$. Level dependent weights provide only a modest improvement in convergence,

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.304 (15)	0.307 (15)	0.308 (15)	0.309 (15)
	1.8	0.280 (14)	0.282 (14)	0.284 (14)	0.285 (14)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.263 (13)	0.266 (13)	0.268 (13)	0.278 (14)
	$(1.8, 1.0)$	0.249 (13)	0.254 (13)	0.257 (13)	0.257 (13)
	(1.0, 2.3)	0.237 (12)	0.250 (13)	0.251 (13)	0.252 (13)
	$(2.3, 0.6)$	0.238 (12)	0.254 (13)	0.256 (13)	0.256 (13)
$m = 4$	$\omega_C = 1.0$	0.564 (30)	0.607 (34)	0.617 (35)	0.619 (35)
	1.8	0.763 (63)	0.777 (67)	0.780 (68)	0.780 (68)
	1.5	0.568 (30)	0.581 (31)	0.591 (32)	0.596 (33)
$m = 4$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.473 (23)	0.537 (27)	0.557 (29)	0.566 (30)
	$(1.5, 1.0)$	0.448 (21)	0.511 (25)	0.537 (27)	0.546 (28)
	$(1.0, 2.3)$	0.655 (40)	0.675 (43)	0.679 (44)	0.680 (44)
	$(2.3, 0.6)$	0.643 (38)	0.660 (41)	0.663 (41)	0.664 (41)

TABLE S5 Two-level MGRIT convergence rates (iterations) for the 1D linear advection equation and weighted FCF- and FCFCF-relaxation.

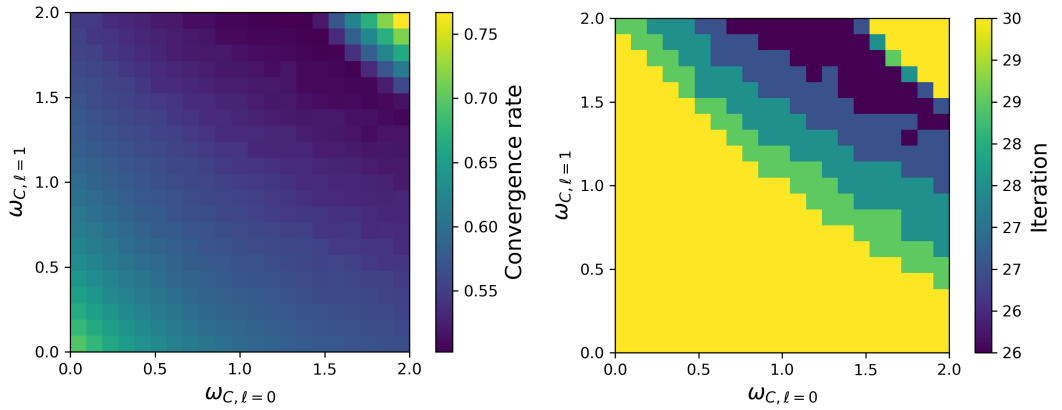
	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.560 (30)	0.675 (44)	0.771 (67)	0.854 (> 100)
	1.5	0.495 (24)	0.606 (35)	0.718 (52)	0.810 (82)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.464 (23)	0.576 (32)	0.678 (45)	0.765 (64)
	$(1.5, 1.0)$	0.423 (20)	0.542 (29)	0.646 (40)	0.738 (57)
	$(1.0, 2.3)$	0.452 (22)	0.605 (35)	0.744 (59)	0.858 (>100)
	(2.3, 0.6)	0.390 (19)	0.492 (25)	0.603 (34)	0.696 (48)
$m = 4$	$\omega_C = 1.0$	0.581 (32)	0.666 (42)	0.757 (61)	0.838 (95)
	1.4	0.535 (27)	0.611 (34)	0.712 (50)	0.802 (77)
$m = 4$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.476 (23)	0.577 (31)	0.677 (43)	0.774 (66)
	$(1.4, 1.0)$	0.448 (22)	0.544 (28)	0.643 (39)	0.752 (60)
	$(1.0, 2.3)$	0.658 (41)	0.683 (44)	0.761 (63)	0.884 (>100)
	$(2.3, 0.6)$	0.607 (34)	0.640 (38)	0.758 (62)	0.860 (>100)

TABLE S6 Multilevel MGRIT convergence rates (iterations) for the 1D linear advection equation and weighted FCF- and FCFCF-relaxation.

but it is a larger improvement than observed for the heat equation, where no iterations were saved. Here, only 3 iterations (4.7%) are saved for $m = 2$, when compared to the best uniform weight choice of $\omega_C = 1.5$. At the bottom of the table, we show how this expensive weight optimization procedure does not carry over to another coarsening factor of $m = 4$, and instead show that a uniform weight choice of $\omega_C = 1.4$ still provides a substantial improvement in convergence. We conclude that for this problem, level-dependent weights do not offer much improvement for convergence and come at the high cost of finding weights.

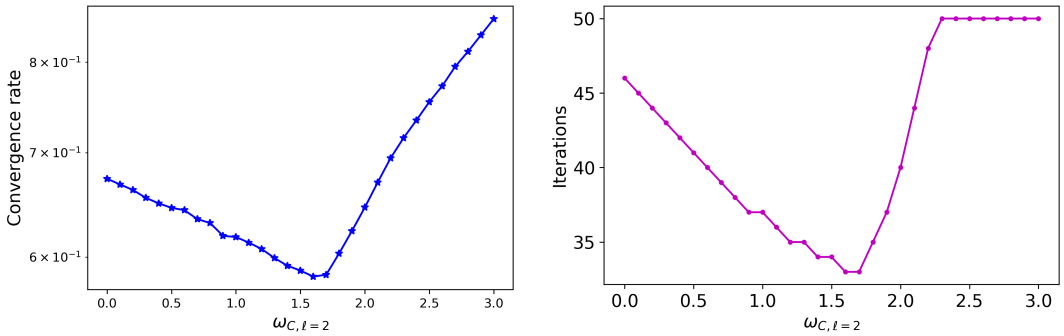
S2.2.3 Varying δ_t experiment

Lastly, similar to the heat equation, we explore the question of why weighted relaxation offers a significantly larger benefit for multilevel MGRIT than for two-level MGRIT (compare Tables S6 and S5). Thus, we



(a) Convergence Rate (b) Iterations

FIGURE S5 Three-level MGRIT experimental convergence rates (left) and iteration counts (right) using various level-dependent FCF-relaxation weights $\omega_{C,\ell=0}$ and $\omega_{C,\ell=1}$ for the 1D linear advection equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (1025, 1025)$.



(a) Convergence Rate (b) Iterations

FIGURE S6 Four-level MGRIT convergence rates (left) and iteration counts (right) using FCF-relaxation, as we search for the best level-three relaxation weight $\omega_{C,\ell=2}$, with the fixed values of $(\omega_{C,\ell=0}, \omega_{C,\ell=1}) = (1.3, 2.0)$ on the first two levels. The problem is the 1D linear advection equation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (1025, 1025)$. The maximum allowed iterations is 50.

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 1.0, 1.0)$	0.562 (31)	0.670 (43)	0.749 (60)	0.788 (72)
	(1.3, 2.0, 1.7)	0.584 (32)	0.591 (33)	0.695 (47)	0.754 (61)
	(1.5, 1.5, 1.5)	0.485 (24)	0.609 (35)	0.710 (51)	0.764 (64)
$m = 4$	$(\omega_{C,\ell=0}, \omega_{C,\ell=1}, \omega_{C,\ell=2}) = (1.0, 1.0, 1.0)$	0.579 (31)	0.670 (42)	0.755 (61)	0.838 (96)
	(1.3, 2.0, 1.7)	0.545 (28)	0.673 (44)	0.794 (76)	0.983 (>100)
	(1.4, 1.4, 1.4)	0.535 (27)	0.613 (35)	0.711 (50)	0.803 (77)

TABLE S7 Four-level MGRIT convergence rates (iterations) for the 1D linear advection equation with level-dependent weights.

explore whether increasing the δ_t value has a discernible impact on MGRIT convergence. Table S8 depicts the two-level MGRIT convergence rate for various fine-grid δ_t values that mimic the δ_t values encountered

with $m = 2$ on coarse MGRIT levels, when a final time of 1.0 is used and $N_t = 4097$ (i.e., the largest problem in Tables S5 and S6). The value N_t also adapts with δ_t so that the final time remains unchanged, similar to coarse MGRIT levels, e.g., when δ_t is multiplied by 16 in Table S8, N_t decreases by a factor or 16 from 4097 to 257. The table shows that only a weak potential dependence exists between δ_t and MGRIT convergence, with a slight improvement in convergence rate as δ_t increases, but no decrease in iterations. This leads us to believe that a more complicated multilevel interaction is driving the improved benefit of weighted-relaxation in the multilevel case.

δ_t	$2.44e^{-4}$	$2 \cdot 2.44e^{-4}$	$4 \cdot 2.44e^{-4}$	$8 \cdot 2.44e^{-4}$	$16 \cdot 2.44e^{-4}$
Iterations	14	14	14	14	14
Convergence Rate	0.285	0.284	0.282	0.280	0.274

TABLE S8 One-dimensional linear advection equation and two-level MGRIT with $\omega_C = 1.8$ and $m = 2$ for various fine-grid δ_t values.

S2.3 One-dimensional advection equation with grid-dependent dissipation

The final one-dimensional model problem considered is the one-dimensional advection equation with grid-dependent dissipation, which yields complex spatial eigenvalues. For initial condition $u_0(x)$ and periodic spatial boundary condition, we have

$$\begin{aligned} &\frac{\partial u}{\partial t} - \alpha \frac{\partial u}{\partial x} - \epsilon h_x \frac{\partial^2 u}{\partial x^2} = 0, \\ &\alpha > 0, \quad \epsilon > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T], \\ &u(x, 0) = u_0(x), \quad x \in \Omega, \\ &u(0, t) = u(L, t), \quad t \in [0, T]. \end{aligned} \tag{B20}$$

By applying standard central differencing for discretizing the spatial derivatives, we obtain the classic first-order upwind difference scheme with $\epsilon = 0.5$. Next, using backward Euler for discretizing the temporal derivative results in

$$\mathbf{u}_j = (I - \delta_t G)^{-1} \mathbf{u}_{j-1}, \quad j = 1, 2, ..., N_t, \tag{B21}$$

where the linear operator G from (1) is the two-point upwinding stencil $\frac{\alpha}{h_x}[-1, 1, 0]$. The eigenvalues of G are then computed from the combination of the previously described eigenvalues for the heat equation and linear advection equations (see Sections S2.1 and S2.2, respectively), yielding

$$\kappa_\gamma = \frac{i}{h_x} \sin\left(\frac{2\pi\gamma}{N_x}\right) - \frac{4\epsilon}{h_x} \sin^2\left(\frac{\gamma\pi}{2(N_x + 1)}\right),$$

for $\gamma = 1, 2, ..., N_x$. These values for κ_γ allow for the computation of the theoretical convergence estimate (19).

The same function, domains, and boundary conditions are used as in equations (B19a) and (B19b). Likewise, the same MGRIT residual norm tolerance, convergence rate measurements, and maximum iterations are used as in Section S2.2. The combination of grid points in space N_x and time N_t are chosen so that $\frac{\delta_t}{h_x} = 1.0$.

S2.3.1 Weighted FCF- and FCFCF-relaxation

We again start by considering the two-level method for weighted FCF- and FCFCF-relaxation. The search for the experimentally optimal pair of weights for FCFCF-relaxation and $m = 2$ is depicted in Figure S7b,

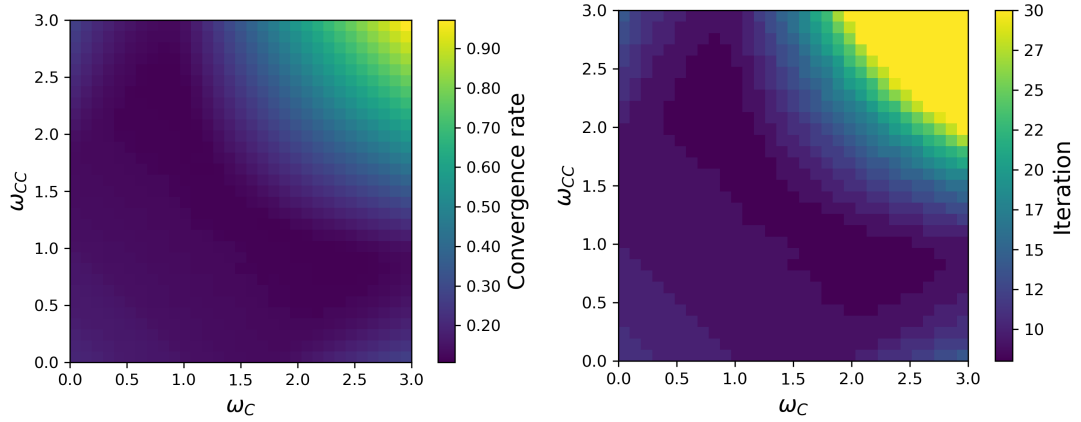
where $(\omega_C, \omega_{CC}) = (2.4, 1.0)$ is the point corresponding to the minimal convergence rate. The search space of weights is the same as that for Section S2.2, $0 \leq \omega_C, \omega_{CC} \leq 3$, because a more expansive preliminary search indicated this was a reasonable range.

A similar study was done in the thesis[40] for FCF-relaxation and found that $\omega_C = 1.9$ is the point where the minimal convergence rate is reached.

Table S9 depicts the convergence rate and iterations for the two-level case. The experimentally optimal pair of weights found in Figure S7b for FCFCF-relaxation $(\omega_C, \omega_{CC}) = (2.4, 1.0)$ is in bold, and this choice leads to saving 1 iteration, or 11%, over unitary weights and FCFCF-relaxation on the largest problem. The best weight choice for FCF-relaxation of $\omega_C = 1.9$ yields only a marginal improvement in convergence and no reduction in iterations when compared to a unitary weight and FCF-relaxation on the largest problem. At the bottom of the table, we examine whether the experimentally optimal weights carry over to $m = 4$ and find that they do not, e.g., $(\omega_C, \omega_{CC}) = (2.4, 1.0)$ is slightly out-performed by $(\omega_C, \omega_{CC}) = (2.2, 0.5)$. Additionally, the experimentally best weight for FCF-relaxation and $m = 4$ was found to be 1.7 (not 1.9).

Table S10 repeats these experiments for a full multilevel method. We see that the best two-level choice for FCFCF-relaxation of $(\omega_C, \omega_{CC}) = (2.4, 1.0)$ fails to provide a benefit for larger problems. Thus, we carry out another search for FCFCF-relaxation and find that the weights $(\omega_C, \omega_{CC}) = (2.2, 0.5)$ yield the fastest multilevel convergence when $m = 2$, saving 9 iterations, or 22%, when compared to unitary weights and FCFCF-relaxation on the largest problem. A search in the weight-space for FCF-relaxation yielded the best convergence rate when $\omega_C = 1.6$, saving 14 iterations or 21%, over a unitary weight choice on the largest problem. At the bottom of the table, we show that the best weight choices for $m = 2$ do not carry over to $m = 4$. We depict the results for an experimentally best weight of 1.4 for FCF-relaxation in order to show that, curiously, MGRIT with FCF-relaxation performs better for $m = 4$ than for $m = 2$.

We again note that linear advection is traditionally difficult for MGRIT, so we view this improved convergence when using experimentally optimal weights to be an important step.



(a) Convergence Rate (b) Iterations

FIGURE S7 Two-level MGRIT experimental convergence rates (left) and iteration counts (right) using FCFCF-relaxation and various relaxation weights ω_C and ω_{CC} for the 1D linear advection equation with dissipation, coarsening factor $m = 2$, and grid size $(N_x, N_t) = (1025, 1025)$.

Remark 2. To avoid repetition, we omit our experiments for level-dependent weights and for varying δ_t, because the results are similar to that seen in Sections S2.2.2 and S2.2.3 for the linear advection equation with purely imaginary spatial eigenvalues. That is, optimized level-dependent weights saved 2 iterations, or 7%, in the four-level setting and FCF-relaxation, and little MGRIT dependence on the size of δ_t was found.

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.147 (9)	0.150 (9)	0.151 (9)	0.151 (9)
	1.9	0.140 (9)	0.141 (9)	0.142 (9)	0.142 (9)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.133 (9)	0.134 (9)	0.135 (9)	0.136 (9)
	(2.2, 0.5)	0.115 (8)	0.117 (8)	0.117 (8)	0.118 (8)
	(2.4, 1.0)	0.114 (8)	0.115 (8)	0.116 (8)	0.116 (8)
$m = 4$	$\omega_C = 1.0$	0.366 (17)	0.339 (18)	0.332 (18)	0.394 (18)
	1.7	0.343 (16)	0.352 (16)	0.363 (17)	0.366 (17)
$m = 4$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.304 (14)	0.329 (15)	0.346 (16)	0.349 (16)
	(1.7, 1.0)	0.273 (13)	0.304 (14)	0.323 (15)	0.326 (15)
	(2.2, 0.5)	0.314 (15)	0.323 (15)	0.330 (15)	0.338 (16)
	(2.4, 1.0)	0.328 (15)	0.334 (16)	0.337 (16)	0.338 (16)

TABLE S9 Two-level MGRIT convergence rates (iterations) for the 1D advection equation with dissipation and weighted FCF- and FCFCF-relaxation.

	$N_x \times N_t$	513×513	1025×1025	2049×2049	4097×4097
$m = 2$	$\omega_C = 1.0$	0.438 (21)	0.560 (30)	0.667 (43)	0.772 (66)
	1.6	0.388 (18)	0.488 (23)	0.613 (35)	0.719 (52)
$m = 2$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.344 (16)	0.432 (21)	0.559 (29)	0.660 (41)
	(1.6, 1.0)	0.293 (14)	0.412 (20)	0.520 (26)	0.638 (38)
	(2.2, 0.5)	0.295 (14)	0.363 (17)	0.482 (24)	0.585 (32)
	(2.4, 1.0)	0.388 (19)	0.564 (32)	0.725 (53)	0.834 (94)
$m = 4$	$\omega_C = 1.0$	0.428 (20)	0.549 (28)	0.657 (40)	0.746 (57)
	1.4	0.375 (18)	0.496 (24)	0.607 (34)	0.694 (46)
$m = 4$	$(\omega_C, \omega_{CC}) = (1.0, 1.0)$	0.336 (16)	0.449 (21)	0.562 (29)	0.677 (43)
	(1.4, 1.0)	0.301 (14)	0.416 (20)	0.542 (28)	0.653 (39)
	(2.2, 0.5)	0.454 (22)	0.582 (31)	0.682 (44)	0.712 (49)
	(2.4, 1.0)	0.404 (19)	0.559 (30)	0.672 (42)	0.689 (45)

TABLE S10 Multilevel MGRIT convergence rates (iterations) for the 1D linear advection equation with dissipation and weighted FCF- and FCFCF-relaxation.