Fall 11-15-2019

Optimal Relaxation Weights for Multigrid Reduction In Time (MGRIT)

Masumi Sugiyama

Masumi Sugiyama

Candidate

Mathematics and Statistics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Jacob Schroder , Chairperson

Jehanzeb Chaudhry

Stephanie Friedhoff

Optimal Relaxation Weights for Multigrid Reduction In Time (MGRIT)

by

Masumi Sugiyama

B.S., Applied Mathematics, University of California, Davis, 2015

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Mathematics

The University of New Mexico

Albuquerque, New Mexico

December, 2019

Dedication

Dedicated to my family and friends.

Acknowledgments

I would first like to thank my thesis advisor Dr. Jacob Schroder for his guidance and support through each stage of the process. I would also like to thank Dr. Jehanzeb Chaudhry and Dr. Stephanie Friedhoff for their support and advice.

Optimal Relaxation Weights for Multigrid Reduction In Time (MGRIT)

by

Masumi Sugiyama

B.S., Applied Mathematics, University of California, Davis, 2015

M.S., Mathematics, University of New Mexico, 2019

Abstract

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 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 the scheme known as multigrid-reduction-in-time (MGRIT), but note that there exist other parallel-in-time methods such as parareal and the parallel full approximation scheme in space and time (PFASST). MGRIT is a full multi-level method applied to the time dimension and 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. In this work, we analyze and select relaxation weights for MGRIT using a convergence analysis and find that this is beneficial since it improves the convergence rate and consequently improves the efficiency of computation. We note that choosing appropriate weights for relaxation (here weighted-Jacobi) has a long history for

improving the convergence of spatial multigrid methods, and thus it is no surprise that such weight selection can be beneficial for MGRIT, too. Our numerical results demonstrate an improved convergence rate and lower iteration count for MGRIT when non-unitary weights are used for weighted-Jacobi.

Contents

List of Figures

List of Tables

Glossary

ω	Unitary relaxation weights $\omega_1 = 1,\ \omega_2 = 1$ used in the standard MGRIT algorithm
$\tilde{\omega}$	Non-unitary relaxation weights $\tilde{\omega_1},\ \tilde{\omega_2}$
$\tilde{\omega}_{op}$	Experimentally optimal relaxation weights
$\|\|r\|\|$	Residual norm for unitary relaxation weights

Chapter 1

Introduction

1.1 Overview

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 the well-known bottleneck for sequential time-integration [7, 5], 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 the scheme known as multigrid-reduction-in-time (MGRIT) [5], but note that there exist other parallel-in-time methods such as parareal [10] and the parallel full approximation scheme in space and time (PFASST) [12, 4, 11]. Paraeal is a two-level multigrid method that allows computations in parallel based on a decomposition of the interval in time dimensions [8]. PFASST can be thought of as a space-time multigrid method that utilizes a deferred correction strategy to computes multiple time-steps in parallel [1].

Unlike parareal, MGRIT is a full multi-level method applied to the time dimension and allows parallel computations in time. Like all multigrid methods, MGRIT

relies on the complementary relationship between relaxation on a fine-grid (e.g., with Jacobi or Gauss-Seidel) and a correction from the coarse grid to solve the problem. For a gentle introduction to parallel-in-time methods, in general, see this review paper [7].

In this work, we analyze and select optimal relaxation weights for the Jacobi method inside of MGRIT and find that this is beneficial since it increases the convergence rate and consequently improves the efficiency of computation. Weighted-Jacobi is considered because it is parallelizable, unlike Gauss-Seidel. We note that choosing appropriate weights for weighted-Jacobi has a long history for improving the convergence of spatial multigrid methods, and thus it is no surprise that such weight selection can be beneficial for MGRIT, too. Our numerical results demonstrate an improved convergence rate and lower iteration count for MGRIT when non-unitary weights are used for weighted-Jacobi. To our knowledge, these are the first results exploring such weighted relaxation with MGRIT.

Chapter 2

MGRIT algorithm

2.1 Two-level multigrid reduction method

In practice, MGRIT uses FAS nonlinear multigrid cycling [2] to solve general nonlinear problems; however, MGRIT convergence analysis [9, 13, 6] to date has focused on linear problems. This is due to the fact that convergence analysis of FAS nonlinear multigrid is largely an open problem, and that MGRIT behavior of linear problems is typically indicative of MGRIT behavior for related nonlinear problems [13].

Thus, for our analysis we consider a linear system of Ordinary Differential Equations (ODEs) of the form

$$\frac{d\mathbf{u}}{dt} = G\mathbf{u}(t) + \mathbf{h}(t), \quad \mathbf{u}(0) = \mathbf{g}_0, \quad t \in [0, T] \tag{2.1}$$

where $\mathbf{u} \in \mathbb{R}^{N_x}$ where N_x refers to the number of points in space, and G is the linear operator. Let $t_j = j\delta_t,\ j = 0, 1, .., N_t$, be a temporal grid with a constant spacing $\delta_t = \frac{T}{N_t} > 0$, and let $\mathbf{u}_j$ be an approximation to $\mathbf{u}(t_j)$ for $j = 1, 2, .., N_t$ with $\mathbf{u}(\mathbf{0}) = \mathbf{u}_0$. Then, a general one-step time discretization method for (2.1) is defined

as

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{g}_0 \\ \mathbf{u}_j &= \Phi \mathbf{u}_{j-1} + \mathbf{g}_j, \quad j = 1, 2, ..., N_t \end{aligned} \tag{2.2}$$

This method is equivalent to solve the system of equations

$$\mathbf{A}\mathbf{u} = \begin{bmatrix} I & & & \\ -\Phi & I & & \\ & \ddots & \ddots & \\ & & -\Phi & I \end{bmatrix} \begin{bmatrix} \mathbf{u_0} \\ \mathbf{u_1} \\ \vdots \\ \mathbf{u_{N_t}} \end{bmatrix} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \\ \vdots \\ \mathbf{g_{N_t}} \end{bmatrix} = \mathbf{g}. \tag{2.3}$$

While sequential time-stepping solves (2.3) directly with forward-substitution, MGRIT algorithm solves (2.3) iteratively by combining a block Jacobi relaxation with error corrections from a coarser approximate representation. Let $T_i = i\delta_T$ for $i = 0, 1, ..., N_T = \frac{N_t}{m}$ be a coarse temporal grid point with a positive integer coarsening factor m and a constant spacing $\delta_T = m\delta_t$. The original grid 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 2.1. These C-points then induce a new coarser time discretization for a new coarse time-grid problem

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

where $\tilde{\mathbf{g}}_{km} = \mathbf{g}_{km} + \Phi \mathbf{g}_{km-1} + \cdots + \Phi^{m-1} \mathbf{g}_{(k-1)m+1}$. This method satisfies the coarse system of equations

$$\mathbf{A}_\triangle \mathbf{u}_\triangle = \begin{bmatrix} I & & & \\ -\Phi^m & I & & \\ & \ddots & \ddots & \\ & & -\Phi^m & I \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_m \\ \vdots \\ \mathbf{u}_{N_T m} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \tilde{\mathbf{g}}_m \\ \vdots \\ \tilde{\mathbf{g}}_{N_T m} \end{bmatrix} = \mathbf{g}_\triangle \tag{2.5}$$

where $\mathbf{A}_\triangle$ has $N_T = \frac{N_t}{m}$ block rows and columns. Unfortunately, solving the equation (2.5) is as expensive as solving (2.3), thus Φ^m is usually approximated with something

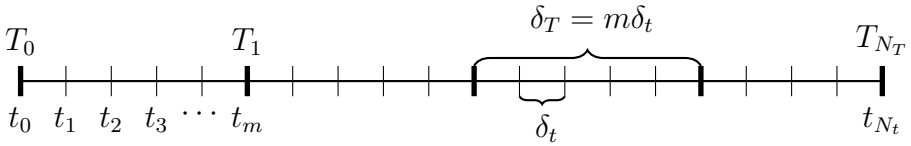


Figure 2.1: Uniformly spaced fine-grid points and coarse-grid points with coarsening factor m. The T_i are the C-points and form the coarse-grid, while the small hashmarks are F-points. Together, the F- and C-points form the fine-grid.

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

Figure 2.2: Schematic view of the action of (a) F-relaxation and (b) C-relaxation with a coarsening factor of four

much cheaper, $\Phi_\triangle$. When $\Phi_\triangle$ is used to approximate Φ^m in $\mathbf{A}_\triangle$, we get a new operator on the coarse-grid, $\mathbf{B}_\triangle \approx \mathbf{A}_\triangle$

With the partition of F- and C-points as depicted in Figure 2.1, there are two fundamental types of relaxation: F- and C-relaxation. The F-relaxation updates the F-point values based on the C-point values, i.e., one F-sweep is

$$\mathbf{u}_i = \Phi \mathbf{u}_{i-1} + \mathbf{g}_i \quad \forall i \in \{\text{F-points}\} \tag{2.6}$$

On the other hand, the C-relaxation updates each C-point value based on the preceding F-point value, i.e., the set of F-points is replaced by the set of C-points in equation (2.6). All F-points in each interval (T_i, T_{i+1}) for $i = 0, ..., N_T - 1$ can be updated simultaneously, and each C-point also can be updated simultaneously in parallel. Figure 2.2 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 is equivalent to Φ^m applied to $\mathbf{u}_{(k-1)m}$ for $k = 1, ..., N_T$. This FC-sweep corresponds to a block Jacobi on the coarse-grid. Using the coarse-

grid operator $\mathbf{A}_{\triangle}$, this block Jacobi relaxation can be written as

$$\begin{aligned}\mathbf{u}_{\triangle}^{(k+1)} &= (I - D^{-1}\mathbf{A}_{\triangle})\mathbf{u}_{\triangle}^{(k)} + D^{-1}\mathbf{g}_{\triangle} \\ &= \begin{bmatrix} 0 & & & \\ \Phi^m & 0 & & \\ & \ddots & \ddots & \\ & & \Phi^m & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_0^{(k)} \\ \mathbf{u}_m^{(k)} \\ \vdots \\ \mathbf{u}_{N_T m}^{(k)} \end{bmatrix} + \begin{bmatrix} \mathbf{g}_0 \\ \tilde{\mathbf{g}}_m \\ \vdots \\ \tilde{\mathbf{g}}_{N_T m} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \Phi^m \mathbf{u}_0^{(k)} + \tilde{\mathbf{g}}_m \\ \vdots \\ \Phi^m \mathbf{u}_{(N_T-1)m}^{(k)} + \tilde{\mathbf{g}}_{N_T m} \end{bmatrix}\end{aligned}$$

The MGRIT algorithm performs either an F-relaxation or an FCF-relaxation. An FCF-relaxation consists of the initial F-relaxation, a C-relaxation, and a second F-relaxation.

2.2 Weighted-Jacobi on FCF-relaxation

The weighted-Jacobi iterative method with $\omega > 0$ has the form

$$\begin{aligned}\mathbf{u}^{(k+1)} &= \omega\{(I - D^{-1}A)\mathbf{u}^{(k)} + D^{-1}\mathbf{g}\} + (1-\omega)\mathbf{u}^{(k)}, \quad k = 0, 1, 2, ... \\ &= \omega G_{Jac} + (1-\omega)\mathbf{u}^{(k)}\end{aligned} \tag{2.7}$$

where G_{Jac} is the Jacobi iteration matrix. Since F- and C-relaxations correspond to the block Jacobi methods, we will examine replacing the existing G_{Jac} applications in MGRIT with weighted relaxation applications of the form (2.7). The form of the G_{Jac} for F- and C-relaxations will appear in the next section. The original python MGRIT code was modified to apply the weighted-Jacobi method of (2.7) after each relaxation.

Chapter 3

Convergence Analysis of MGRIT

3.1 Convergence estimates for MGRIT

3.1.1 MGRIT error propagator for unweighted FCF-relaxation

Let the fine-grid operator $\mathbf{A}$ in (2.3) be reordered so that F-points appear first and C-points second. Then by using the subscripts c and f to indicate the two sets of points, we have the Schur complement decomposition

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

where A_{ff} and A_{cc} are the identity submatrices. This decomposition implies an ideal interpolation operator which formulates a Schur complement.[1] Then, the injection restriction operator R_I, ideal interpolation operator P, and map S are defined by

$$R_I = \begin{bmatrix} 0 & I_c \end{bmatrix}, \quad P = \begin{bmatrix} -A_{ff}^{-1}Afc \\ I_c \end{bmatrix}, \quad S = \begin{bmatrix} I_f \\ 0 \end{bmatrix}$$

[1] This is one reason why P is called *ideal*. The other related reason is that if an exact solution is available at C-points, then multiplication by P will yield the exact solution at all C- and F-points.

Using these operators, the Schur complement is given by $R_I AP = A_{cc} - A_{cf}A_{ff}^{-1}A_{cf}$ which is then equal to the coarse-grid operator $\mathbf{A}_\triangle$. Then, together with $S^T AS = A_{ff}$, we have

$$A^{-1} = P(R_I AP)^{-1}R_I + S(S^T AS)^{-1}S^T$$

This gives the error propagator of the exact two-level multigrid method for F-relaxation

$$0 = I - A^{-1}A = (I - P(R_I AP)^{-1}RA)(I - S(S^T AS)^{-1}S^T A) \tag{3.1}$$

where equivalence occurs since $RAS = 0$. The error propagator is a matrix that when given the error relative to some initial guess, it determines how the error changes from iteration to iteration. The first term of (3.1) corresponds to the error propagator of coarse-grid correction, and the second term corresponds to the error propagator of F-relaxation. It is the error propagator of F-relaxation because the operator S^T is a map to F-points, hence $S(S^T AS)^{-1}S^T$ carries out a block inverse over the F-points. It is equivalent to PR_I since R_I is a map to C-points, and the application of P sets the residual equal to zero at all F-points (F-relaxation). Since $\mathbf{A}_\triangle = R_I AP$ and $\mathbf{B}_\triangle \approx \mathbf{A}_\triangle$ in an iterative multigrid reduction method, the error propagator (3.1) becomes

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

The two-level error propagator for FCF-relaxation can be derived by replacing the error propagator of F-relaxation in (3.1) with the error propagator of FCF-relaxation. The error propagator for C-relaxation is expressed as $(I - R_I^T(R_I AR_I^T)^{-1}R_I A)$ since R_I is a map to C-points, and $R_I^T(R_I AR_I^T)^{-1}R_I$ carries out a block inverse over the C-points. Then, the two-level error propagator for FCF-relaxation is

$$\begin{aligned}
&(I - S(S^T AS)^{-1}S^T A)(I - R_I^T(R_I AR_I^T)^{-1}R_I A)(I - S(S^T AS)^{-1}S^T A) \\
&= PR_I(I - R_I^T(R_I AR_I^T)^{-1}R_I A)PR_I \\
&= P(I_c - (R_I AR_I^T)^{-1}R_I AP)R_I \\
&= P(I - A_\triangle)R_I
\end{aligned}$$

This yields the two-level error propagator for FCF-relaxation defined as

$$(I - PB_{\triangle}^{-1}R_I A)P(I - A_{\triangle})R_I = P(I - B_{\triangle}^{-1}A_{\triangle})(I - A_{\triangle})R_I \tag{3.2}$$

3.1.2 Error propagators for stand-alone weighted F- and FCF-relaxation

The weighted-Jacobi for F-relaxation using (2.7) and its error propagator can be written as

$$\begin{aligned} \mathbf{u}^{k+1} &= \omega\{(I - S(S^T AS)^{-1}S^T A)\mathbf{u}^k + D^{-1}\mathbf{g}\} + (1-\omega)\mathbf{u}^k \\ &= (I - \omega S(S^T AS)^{-1}S^T A)\mathbf{u}^k + D^{-1}\mathbf{g} \end{aligned} \tag{3.3}$$

Similarly, the weighted-Jacobi for C-relaxation using its error propagator can be written as

$$\mathbf{u}^{k+1} = (I - \omega R_I^T(R_I A R_I^T)^{-1}R_I A)\mathbf{u}^k + D^{-1}\mathbf{g} \tag{3.4}$$

Hence, the error propagator of FCF-relaxation with the weighted-Jacobi is given by

$$(I - \omega_3 S(S^T AS)^{-1}S^T A)(I - \omega_2 R_I^T(R_I A R_I^T)^{-1}R_I A)(I - \omega_1 S(S^T AS)^{-1}S^T A)$$

where $\omega_1, \omega_2, \omega_3 > 0$. We will only consider $\omega_3 = 1.0$ because if $\omega_3 \neq 1$, then MGRIT would no longer be a reduction method and experiments and intuition both indicate that performance suffers. With this simplification, the error propagator becomes

$$\begin{aligned} &(I - S(S^T AS)^{-1}S^T A)(I - \omega_2 R_I^T(R_I A R_I^T)^{-1}R_I A)(I - \omega_1 S(S^T AS)^{-1}S^T A) \\ &= \left(I - \begin{bmatrix} I_f & A_{ff}^{-1}A_{fc} \\ 0 & 0 \end{bmatrix}\right)\left(I - \omega_2 \begin{bmatrix} 0 & 0 \\ A_{cc}^{-1}A_{cf} & I_c \end{bmatrix}\right)\left(I - \omega_1 \begin{bmatrix} I_f & A_{ff}^{-1}A_{fc} \\ 0 & 0 \end{bmatrix}\right) \\ &= \begin{bmatrix} -\omega_2(\omega_1 - 1)A_{ff}^{-1}A_{fc}A_{cc}^{-1}A_{cf} & -(1-\omega_2)A_{ff}^{-1}A_{fc} - \omega_2\omega_1 A_{ff}^{-1}A_{fc}A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \\ \omega_2(\omega_1 - 1)A_{cc}^{-1}A_{cf} & (1-\omega_2)I_c + \omega_2\omega_1 A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \end{bmatrix} \end{aligned} \tag{3.5}$$

3.1.3 MGRIT error propagator when $\omega_1 = 1.0$

If $\omega_1 = 1.0$, the matrix of the form (3.5) becomes

$$\begin{aligned}
&= \begin{bmatrix} 0 & -A_{ff}^{-1}A_{fc} + \omega_2 A_{ff}^{-1}A_{fc} - \omega_2 A_{ff}^{-1}A_{fc}A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \\ 0 & I_c - \omega_2 I_c + \omega_2 A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \end{bmatrix} \\
&= \begin{bmatrix} 0 & -A_{ff}^{-1}A_{fc}\{I_c - \omega_2 A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc})\} \\ 0 & I_c - \omega_2 A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}) \end{bmatrix} \\
&= \begin{bmatrix} -A_{ff}^{-1}A_{fc} \\ I_c \end{bmatrix} \Big[I_c - \omega_2 A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}) \Big] \begin{bmatrix} 0 & I_c \end{bmatrix} \\
&= P(I - \omega_2 \mathbf{A}_\triangle) R_I
\end{aligned}$$

Based on the two-level MGRIT error propagator with FCF-relaxation defined in (3.2), the two-level error propagator for FCF-relaxation with the weighted-Jacobi results in

$$(I - PB_\triangle^{-1}R_I A)P(I - \omega_2 \mathbf{A}_\triangle)R_I = P(I - \mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)(I - \omega_2 \mathbf{A}_\triangle)R_I$$

When the coarse-grid equations are satisfied, i.e., the coarse-grid solution is exact, one application of P yields the exact solution at F-points leaving the residual equal to zero at these points. Thus, the two-level error propagator can be analyzed on the coarse-grid and it becomes

$$\begin{aligned}
E_{\triangle,\omega_1=1}^{FCF} &= (I - \mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)(I - \omega_2 \mathbf{A}_\triangle) \\
&= (\Phi^m - \Phi_\triangle) \left\{ (1-\omega_2) \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ \Phi_\triangle & 1 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_\triangle^{N_T-1} & \Phi_\triangle^{N_T-2} & \cdots & 1 & 0 \end{bmatrix} + \omega_2 \Phi^m \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_\triangle^{N_T-2} & \Phi_\triangle^{N_T-3} & \cdots & 0 & 0 \end{bmatrix} \right\}
\end{aligned} \tag{3.6}$$

Let $\mathbf{e}_{km}$ be the error at a C-point for $k = 0, 1, ..., N_T$ and define the time-space error as $\bar{\mathbf{e}} = [\mathbf{e}_0^T, \mathbf{e}_m^T, \mathbf{e}_{2m}^T, ..., \mathbf{e}_{N_T m}^T]$. The two-level error propagation formula is then given by

$$(E_{\triangle,\omega_1=1}^{FCF}\bar{\mathbf{e}})_k = \begin{cases} 0 & k = 0 \\ (\Phi^m - \Phi_\triangle)\{(1-\omega_2)\sum_{q=0}^{k-1}\Phi_\triangle^{k-1-q} + \omega_2 \sum_{q=0}^{k-2}\Phi_\triangle^{k-2-q}\Phi^m\}\bar{\mathbf{e}}_q & k = 1, 2, ..., N_T \end{cases} \tag{3.7}$$

where $\bar{\mathbf{e}}_k = \bar{\mathbf{e}}_{km}$.

Let λ_γ be the eigenvalues of Φ and μ_γ be the eigenvalues of $\Phi_\triangle$, corresponding to the same set of eigenvectors $\{v_\gamma\}$. That is, Φ and $\Phi_\triangle$ are diagonalized by the same vectors, which is the case for our problems. For instance if backward Euler is used,

$$\lambda_\gamma = (1 - h_t\kappa_\gamma)^{-1}, \quad \mu_\gamma = (1 - mh_t\kappa_\gamma)^{-1} \quad \text{for } \gamma = 1, 2, ..., N_x \tag{3.8}$$

where $\kappa_\gamma \geq 0$ is an eigenvalue of the linear operator G in (2.1). Using these eigenvalues, the norm of the two-level error propagation formula (3.7) can be expressed as

$$\begin{aligned} &||E^{FCF}_{\triangle,\,\omega_1=1}\bar{\mathbf{e}}||_2 \\ &\leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma| \left\{ |1-\omega_2| \Big(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|} \Big) + |\omega_2||\lambda_\gamma|^m \Big(\frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \Big) \right\} ||\bar{\mathbf{e}}||_2 \end{aligned} \tag{3.9}$$

For the derivation of this theoretical convergence estimate, see the appendix A.

3.1.4 MGRIT error propagator when $\omega_1 \neq 1.0$

If $\omega_1 \neq 1.0$, the matrix of the form (3.5) becomes

$$\begin{aligned} &= \begin{bmatrix} 0 & -\omega_2\omega_1 A_{ff}^{-1}A_{fc}A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \\ 0 & \omega_2\omega_1 A_{cc}^{-1}A_{cf}A_{ff}^{-1}A_{fc} \end{bmatrix} + \begin{bmatrix} -\omega_2(\omega_1-1)A_{ff}^{-1}A_{fc}A_{cc}^{-1}A_{cf} & -(1-\omega_2)A_{ff}^{-1}A_{fc} \\ \omega_2(\omega_1-1)A_{cc}^{-1}A_{cf} & (1-\omega_2)I_c \end{bmatrix} \\ &= \omega_2\omega_1 \begin{bmatrix} 0 & -A_{ff}^{-1}A_{fc}\{I_c - A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc})\} \\ 0 & I_c - A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}) \end{bmatrix} \\ &\qquad + (1-\omega_2)\begin{bmatrix} 0 & -A_{ff}^{-1}A_{fc} \\ 0 & I_c \end{bmatrix} + \omega_2(\omega_1-1)\begin{bmatrix} -A_{ff}A_{fc}A_{cc}^{-1}A_{cf} & 0 \\ A_{cc}^{-1}A_{cf} & 0 \end{bmatrix} \\ &= \omega_2\omega_1 \begin{bmatrix} -A_{ff}^{-1}A_{fc} \\ I_c \end{bmatrix} \Big[I_c - A_{cc}^{-1}(A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}) \Big] \begin{bmatrix} 0 & I_c \end{bmatrix} \\ &\qquad + (1-\omega_2)\begin{bmatrix} -A_{ff}^{-1}A_{fc} \\ I_c \end{bmatrix} \begin{bmatrix} 0 & I_c \end{bmatrix} + \omega_2(1-\omega_1)\begin{bmatrix} -A_{ff}^{-1}A_{fc} \\ I_c \end{bmatrix} \Big[A_{cc}^{-1}A_{cf} \Big] \begin{bmatrix} I_f & 0 \end{bmatrix} \\ &= \omega_2\omega_1 P(I - \mathbf{A}_\triangle)R_I + (1-\omega_2)PR_I + \omega_2(1-\omega_1)PA_{cf}S^T \end{aligned}$$

Hence, the error propagator is

$$\begin{aligned} &(I - P\mathbf{B}_\triangle^{-1}RA)\{\omega_2\omega_1 P(I-\mathbf{A}_\triangle)R_I + (1-\omega_2)PR_I + \omega_2(1-\omega_1)PA_{cf}S^T\} \\ &= \omega_2\omega_1 P(I - \mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)(I-\mathbf{A}_\triangle)R_I + (1-\omega_2)P(I-\mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)R_I + \omega_2(1-\omega_1)P(I-\mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)A_{cf}S^T \\ &= P\{\omega_2\omega_1(I-\mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)(I-\mathbf{A}_\triangle) + (1-\omega_2)(I-\mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)\}R_I + \omega_2(1-\omega_1)P(I-\mathbf{B}_\triangle^{-1}\mathbf{A}_\triangle)A_{cf}S^T \end{aligned}$$

The error propagator is then

$$E^{FCF}_{\triangle,\omega_1\neq 1} = \omega_2\omega_1(I-B^{-1}_{\triangle}\mathbf{A}_{\triangle})(I-\mathbf{A}_{\triangle})+(1-\omega_2)(I-\mathbf{B}^{-1}_{\triangle}\mathbf{A}_{\triangle})+\omega_2(1-\omega_1)(I-\mathbf{B}^{-1}_{\triangle}\mathbf{A}_{\triangle})A_{cf}$$

The first term of error propagator in matrix form is expressed as

$$\omega_1\omega_2\begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ \Phi^m(\Phi^m-\Phi_{\triangle}) & 0 & 0 & & \\ \vdots & \ddots & \ddots & 0 & \\ \Phi^m\Phi^{N_T-2}_{\triangle}(\Phi^m-\Phi_{\triangle}) & \cdots & \Phi^m(\Phi^m-\Phi_{\triangle}) & 0 & 0 \end{bmatrix}$$

The third term of $E^{FCF}_{\triangle,\,\omega_1\neq 1}$ depends on a coarsening factor m due to a submatrix A_{cf}, and is expressed as

$$\begin{cases} 0 & k=0,1 \\ -\omega_2(1-\omega_1)(\Phi^m-\Phi_{\triangle})\sum_{q=0}^{k-2}\Phi^{k-2-q}_{\triangle}\Phi\,\bar{\mathbf{e}}_{(m-1)q+(m-2)} & k=2,...,N_T \end{cases}$$

The middle term is already known, thus by combining the all terms together, the error propagation formula is given by

$$(E^{FCF}_{\triangle,\omega_1\neq 1}\bar{\mathbf{e}})_k = \begin{cases} 0 & k=0 \\ (\Phi^m-\Phi_{\triangle})\{\omega_1\omega_2\sum_{q=0}^{k-2}\Phi^{k-2-q}_{\triangle}\Phi^m+(1-\omega_2)\sum_{q=0}^{k-1}\Phi^{k-1-q}_{\triangle}\}\bar{\mathbf{e}}_q & \\ \qquad -\omega_2(1-\omega_1)(\Phi^m-\Phi_{\triangle})\sum_{q=0}^{k-2}\Phi^{k-2-q}_{\triangle}\Phi\,\bar{\mathbf{e}}_{(m-1)q+(m-2)} & k=1,2,...,N_T \end{cases} \tag{3.10}$$

Using the eigenvalues defined in the form of (3.8),

$$\begin{aligned} &||E^{FCF}_{\triangle,\,\omega_1\neq 1}\bar{\mathbf{e}}||_2 \\ &\leq \max_{\gamma}\,|\lambda^m_{\gamma}-\mu_{\gamma}|\{\Big(|\omega_1\omega_2||\lambda_{\gamma}|^m-|\omega_2||1-\omega_1||\lambda_{\gamma}|\Big)\Big(\frac{1-|\mu_{\gamma}|^{N_T-1}}{1-|\mu_{\gamma}|}\Big) \\ &\qquad\qquad +|1-\omega_2|\Big(\frac{1-|\mu_{\gamma}|^{N_T}}{1-|\mu_{\gamma}|}\Big)\}||\bar{\mathbf{e}}||_2 \end{aligned} \tag{3.11}$$

For the derivation of this theoretical convergence estimate, see the Appendix A. If $\omega_1 = 1$, the error propagation formula above is equivalent to the error propagation formula for $\omega_1 = 1$ of the form (3.7). Hence, $E^{FCF}_{\triangle,\omega_1=1} = E^{FCF}_{\triangle,\omega_1\neq 1}$ for FCF-relaxation. Moreover, if $\omega_1 = \omega_2 = 1$, then these expressions are equivalent to the original analysis in [13].

3.2 Modified theoretical convergence estimates for MGRIT

We now propose modifications to the above estimate of (3.9), with the intention of providing a sharper convergence estimate. The first and second modified theoretical convergence estimates below are derived from the matrix form of the two-level error propagator for $\omega_1 = 1$ in the form of (3.7).

3.2.1 First modified convergence estimate

We begin defining the matrices M_1, M_2 and Q as follows

$$M_1 = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ \Phi_{\triangle} & 1 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_{\triangle}^{N_T-1} & \Phi_{\triangle}^{N_T-2} & \cdots & 1 & 0 \end{bmatrix}, M_2 = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_{\triangle}^{N_T-2} & \Phi_{\triangle}^{N_T-3} & \cdots & 0 & 0 \end{bmatrix}, Q = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 0 & 1 & \ddots & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

where Q is a shift matrix, and M_2 can be written as the product of M_1Q. Then, the norm of the two-level error propagator can be also derived in the following way using the eigenvalues of (3.8).

$$\begin{aligned} ||E^{FCF}_{\triangle,\,\omega_1=1}\bar{\mathbf{e}}||_\infty &\leq \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\ ||(1-\omega_2)M_1 + \lambda_\gamma^m\omega_2 M_2||_\infty||\bar{\mathbf{e}}||_\infty \\ &= \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\ ||(1-\omega_2)M_1 + \lambda_\gamma^m\omega_2 M_1Q||_\infty||\bar{\mathbf{e}}||_\infty \\ &= \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\ ||\{(1-\omega_2)I + \lambda_\gamma^m\omega_2 Q\}M_1||_\infty||\bar{\mathbf{e}}||_\infty \\ &\leq \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\ ||\{(1-\omega_2)I + \lambda_\gamma^m\omega_2 Q\}||_\infty||M_1||_\infty||\bar{\mathbf{e}}||_\infty \\ &= \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\{|1-\omega_2| + |\omega_2\lambda_\gamma^m|\}\Big(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|}\Big)||\bar{\mathbf{e}}||_\infty \end{aligned}$$

Then, the fist modified theoretical convergence estimate is given by

$$||E^{FCF}_{\triangle,\,\omega_1=1}\bar{\mathbf{e}}||_2 \leq \max_\gamma\ |\lambda_\gamma^m - \mu_\gamma|\{|1-\omega_2| + |\omega_2\lambda_\gamma^m|\}\Big(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|}\Big)||\bar{\mathbf{e}}||_2 \qquad (3.12)$$

It turned out that this modified theoretical convergence estimate does not always produce a sharper estimate than the original theoretical convergence estimate (3.9). Therefor, this modified theoretical convergence estimate is not used in numerical experiment.

3.2.2 Second modified convergence estimate

Let the matrices from (3.6) be combined into

$$
\begin{aligned}
&E^{FCF}_{\triangle,\,\omega_1=1} \\
&= (\Phi^m - \Phi_\triangle)\begin{bmatrix} 0 & & & & \\ (1-\omega_2) & 0 & & & \\ (1-\omega_2)\Phi_\triangle + \omega_2\Phi^m & (1-\omega_2) & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ (1-\omega_2)\Phi_\triangle^{N_T-1} + \omega_2\Phi^m\Phi_\triangle^{N_T-2} & (1-\omega_2)\Phi_\triangle^{N_T-2} + \omega_2\Phi^m\Phi_\triangle^{N_T-3} & \cdots & (1-\omega_2) & 0 \end{bmatrix} \\
&= (\Phi^m - \Phi_\triangle)\left\{ \begin{bmatrix} 0 & & & & \\ (1-\omega_2) & 0 & & & \\ 0 & (1-\omega_2) & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ 0 & 0 & \cdots & (1-\omega_2) & 0 \end{bmatrix} + [(1-\omega_2)\Phi_\triangle + \omega_2\Phi^m] \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_\triangle^{N_T-2} & \Phi_\triangle^{N_T-3} & \cdots & 0 & 0 \end{bmatrix} \right\}
\end{aligned}
$$

Using the eigenvalues of (3.8), the two-level error propagator becomes

$$
(\lambda_\gamma^m - \mu_\gamma)\left\{ \begin{bmatrix} 0 & & & & \\ (1-\omega_2) & 0 & & & \\ 0 & (1-\omega_2) & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ 0 & 0 & \cdots & (1-\omega_2) & 0 \end{bmatrix} + [(1-\omega_2)\mu_\gamma + \omega_2\lambda_\gamma^m] \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \mu_\gamma^{N_T-2} & \mu_\gamma^{N_T-3} & \cdots & 0 & 0 \end{bmatrix} \right\}
$$

Then, taking the infinity norm of the above equation for each γ, the second modified theoretical bound can be written as

$$
\begin{aligned}
&||E^{FCF}_{\triangle,\,\omega_1=1}\bar{\mathbf{e}}||_2 \\
&\leq \max_\gamma\, |\lambda_\gamma^m - \mu_\gamma| \left\{ |1-\omega_2| + |(1-\omega_2)\mu_\gamma + \omega_2\lambda_\gamma^m| \Big(\frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \Big) \right\} ||\bar{\mathbf{e}}||_2
\end{aligned}
\tag{3.13}
$$

This modified theoretical convergence estimate resulted in producing a sharper estimate than the original theoretical convergence estimate (3.9). as will be shown in the following numerical results section.

Chapter 4

Numerical results

In these numerical experiments, we consider various modified weights $\tilde{\omega}_2 \neq 1 = \omega_2$ for C-relaxation, while keeping $\tilde{\omega}_1 = 1 = \omega_1$ for F-relaxation fixed. For each model problem, the optimal weight of real-, imaginary-, and complex-valued $\tilde{\omega}_2$ are explored. Results of modified weights $\tilde{\omega}$ are compared against results of the unitary weights ω (defined as the combination of $\omega_1 = 1$ and $\omega_2 = 1$). The experimentally measured asymptotic convergence rate is taken to be the average convergence rate over the last five iterations, and it is compared with the original and second modified theoretical convergence estimates (3.9, 3.13). In this section, "modified convergence estimate" refers to the second modified convergence estimate from Section 3.2.2

4.1 One-dimensional heat equation

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

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial^2 u}{\partial x^2} &= f(x,t), \quad \alpha > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T] \\ u(x,0) &= u_0(x), \quad x \in \Omega \\ u(x,t) &= 0, \quad x \in \partial\Omega, \quad t \in [0, T] \end{aligned} \tag{4.1}$$

We transform the model problem to a system of ODEs of the form (2.1) by using second-order central differencing for discretizing the spatial derivative and then a standard one-time step method of the form (2.2) for discretizing the time derivation, e.g., backward Euler (*Backward Time, Central Space* or BTCS scheme). We obtain

$$\mathbf{u}_j = (I - \delta_t G)^{-1} \mathbf{u}_{j-1} + (I - \delta_t G)^{-1} \delta_t \mathbf{f}_j, \quad j = 1, 2, ..., N_t \tag{4.2}$$

where the linear operator G in (2.1) is the three-point stencil $\frac{\alpha}{h_x^2}[1, -2, 1]$. In the form of (2.2), $\Phi = (I - \delta_t G)^{-1}$ 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, $\kappa_\gamma = -\frac{4}{h_x^2} \sin^2\left(\frac{\gamma\pi}{2(N_x+1)}\right)$ for $\gamma = 1, 2, ..., N_x$, and the formulas (3.8) in order to compute the theoretical convergence estimates.

4.1.1 Problem statement

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

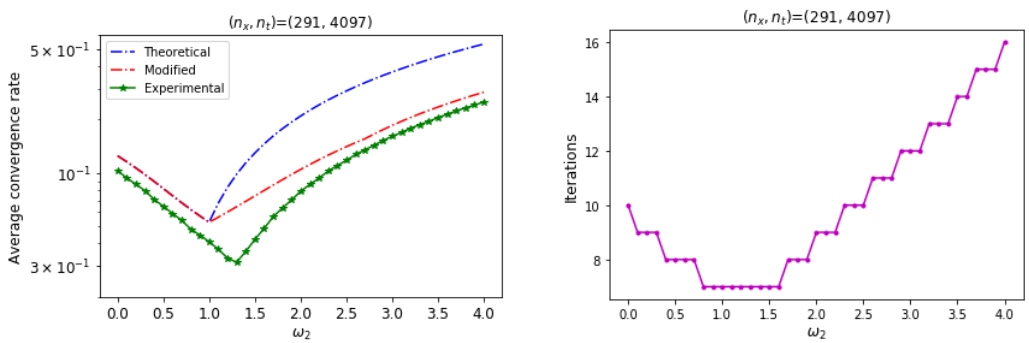
$$\begin{aligned} u(x,t) &= \sin(\pi x)\cos(t) \\ f(x,t) &= \sin(\pi x)[\sin(t) - \pi^2 \cos(t)] \\ \alpha &= 1, \quad x \in [0,1], \quad t \in [0, 0.625] \end{aligned}$$

The residual norm halting tolerance for MGRIT is set to $\frac{1.0e^{-10}}{\sqrt{h_x \delta_t}}$ throughout the experiments. The combination of grid points in space $n_x = N_x$ and time $n_t = N_t + 1$

are chosen so that a Courant-Friedrichs-Lewys condition (CFL) is $\frac{\delta_t}{h_x^2} = 12.8$ or $\frac{\delta_t}{h_x^2} = 1.0$.

4.1.2 Real valued weight

We start by considering non-unitary weights of $\tilde{\omega}_2 \in \mathbb{R}$ for $\frac{\delta_t}{h_x^2} = 12.8$. Figure 4.1a captures the original and modified theoretical convergence estimates (3.9, 3.13) from the sections 3.1.3, 3.2.2, and the experimentally measured asymptotic convergence rate of the problem size $n_x \times n_t = 291 \times 4097$ when $m = 2$. It depicts that the sharp prediction of a convergence rate from the theoretical estimates when $\tilde{\omega}_2 \leq 1.0$. When $\tilde{\omega}_2 > 1.0$, however, the original theoretical estimate is not sharp while the modified theoretical estimate produces a sharper estimate. Unfortunately, the theoretical estimates do not predict the optimal weight of $\tilde{\omega}_2$. Together with iterations from Figure 4.1b, the experimentally optimal weight is $\tilde{\omega}_{op} = 1.3$ under the given conditions.



(a) Asymptotic convergence rates and theoretical convergence estimates

(b) Iterations

Figure 4.1: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{R}$, considering $m = 2$

Table 4.1 compares the observed convergence rates and iterations with various coarsening factors m and problem sizes for the two-level and multi-level methods. The top table shows the results of unitary ω as a reference. The smaller convergence

rates and non-increased iterations of non-unitary $\tilde{\omega}$ than ω for all m and problem sizes suggest that $\tilde{\omega}_2 = 1.3$ is a good candidate for a universal optimal value. In fact, Table 4.2 depicts that this $\tilde{\omega}_{op} = 1.3$ is fairly stable for various m and problem sizes. With the fixed problem size, as m increases, the two-level results for $\tilde{\omega}$ deteriorate. The convergence rates do not depend on the problem sizes with the fixed m. The iterations are reduced by about $10 - 14\%$ for the both methods when the modified weight is used.

It is interesting to note that the observed asymptotic convergence rate can be a misleading indicator for the performance of MGRIT. For instance, Table 4.3 shows the iterations, residual norms, and asymptotic convergence rates for V-cycles with the problem size $n_x \times n_t = 411 \times 8193$. The asymptotic convergence rate indicates a better performance with $m = 4$, however, $m = 32$ converges in one less iteration. Thus, we always report convergence rates and iterations to give the full picture.

For the other CFL number under consideration, $\frac{\delta_t}{h_x^2} = 1.0$, similar tables and plots exist, and yield a nearly identical experimentally optimal weight. These results are omitted for brevity.

Unitary weights ω

m / Size		291×4097	411×8193	581×16385	821×32769	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	0.049 (7)	0.048 (7)	0.039 (7)	0.039 (7)	0.053
	4	0.077 (8)	0.077 (8)	0.058 (8)	0.058 (8)	0.081
	32	0.097 (8)	0.104 (8)	0.078 (8)	0.078 (8)	0.108
V-cycle	2	0.118 (9)	0.121 (9)	0.093 (9)	0.094 (9)	-
	4	0.101 (9)	0.102 (9)	0.080 (9)	0.080 (9)	-
	32	0.097 (8)	0.104 (8)	0.078 (8)	0.076 (8)	-

Non-unitary weights $\tilde{\omega}_2 = 1.3$

m / Size		291×4097	411×8193	581×16385	821×32769	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	0.036 (7)	0.036 (7)	0.034 (6)	0.034 (6)	0.064
	4	0.056 (7)	0.055 (7)	0.055 (7)	0.054 (7)	0.098
	32	0.081 (8)	0.077 (8)	0.077 (8)	0.078 (8)	0.128
V-cycle	2	0.092 (8)	0.095 (8)	0.096 (8)	0.096 (8)	-
	4	0.072 (8)	0.073 (8)	0.074 (8)	0.074 (8)	-
	32	0.082 (8)	0.077 (8)	0.077 (8)	0.074 (7)	-

Table 4.1: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Asymptotic convergence rates and iterations for MGRIT. The modified theoretical estimate appears in the final column.

	m / Size	291×4097	411×8193	581×16385	821×32769
Two-level	2	1.3	1.2	1.3	1.2
	4	1.3	1.3	1.3	1.2
	32	1.3	1.3	1.3	1.2
V-cycle	2	1.5	1.5	1.5	1.5
	4	1.4	1.4	1.4	1.4
	32	1.3	1.3	1.3	1.3

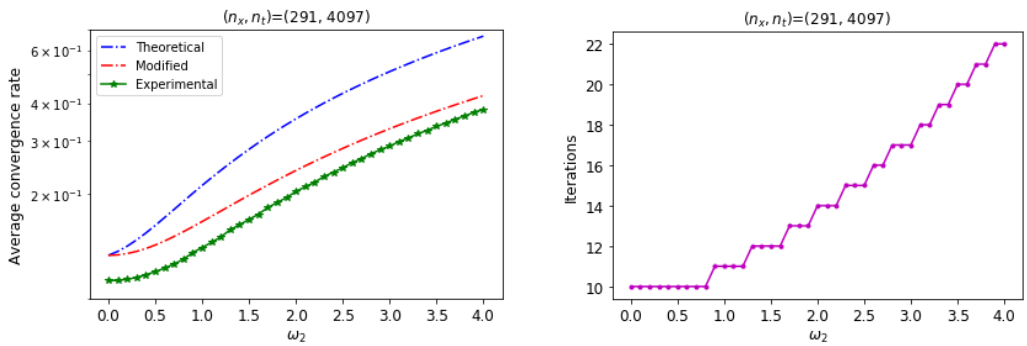
Table 4.2: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Experimentally optimal weights $\tilde{\omega}_2 \in \mathbb{R}$ for MGRIT.

	$m = 4$		$m = 32$	
iteration	$\|\|r\|\|$	convergence rate	$\|\|r\|\|$	convergence rate
0	$5.51e^{2}$	-	$5.51e^{2}$	-
1	4.32	$0.78e^{-2}$	2.12	$0.385e^{-2}$
2	$3.58e^{-1}$	$0.829e^{-1}$	$1.90e^{-1}$	$0.897e^{-1}$
3	$3.44e^{-2}$	$0.960e^{-1}$	$1.85e^{-2}$	$0.973e^{-1}$
4	$3.42e^{-3}$	$0.993e^{-1}$	$1.87e^{-3}$	0.100
5	$3.45e^{-4}$	0.101	$1.93e^{-4}$	0.103
6	$3.52e^{-5}$	0.102	$2.01e^{-5}$	0.104
7	$3.61e^{-6}$	0.102	$2.12e^{-6}$	0.105
8	$3.71e^{-7}$	0.103	$2.24e^{-7}$	0.106
9	$3.82e^{-8}$	0.103	-	-
Average rate	0.102		0.104	

Table 4.3: Residual, asymptotic convergence rates, and iterations for one-dimensional heat equation when using V-cycles and the problem size $n_x \times n_t = 411 \times 8193$

4.1.3 Imaginary valued weight

We consider non-unitary weights of $\tilde{\omega}_2 \in \mathbb{I}$ for $\frac{\delta_t}{h_x^2} = 12.8$. Figure 4.2 concludes that the experimentally optimal weight for the two-level method with $m = 2$ is $\tilde{\omega}_{op} = 0$, which is the F-relaxation. The modified theoretical estimate is sharp and correctly predicts the optimal weight of $\tilde{\omega}_2$. Interestingly, the first modified theoretical convergence estimate (3.12), derived the same way as the original and second modified theoretical estimates, yields larger estimate for some $\tilde{\omega}_2$.



(a) Asymptotic convergence rates and theoretical convergence estimates

(b) Iterations

Figure 4.2: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{I}$, considering $m = 2$

Each table entry of Table 4.4 is formatted as *experimentally optimal weight : convergence rate (iteration count)*. For example, the first entry of *0.6i: 0.434 (22)* indicates that an $\tilde{\omega}_2 = 0.6i$ was found to be close to optimal experimentally, and lead to a convergence rate of 0.434 with 22 iterations. The table considers coarsening factors of $m = 2, 4, 32$.

The experimentally optimal weight for the two-level method turned out to be $\tilde{\omega_{op}} = 0$ for various problem sizes and coarsening factors. On the other hand, the Table 4.4 depicts that the experimentally optimal weight for the multi-level varies. The iterations for the both methods are larger than these for the unitary weights,

concluding that a imaginary valued weight does not contribute to a better performance.

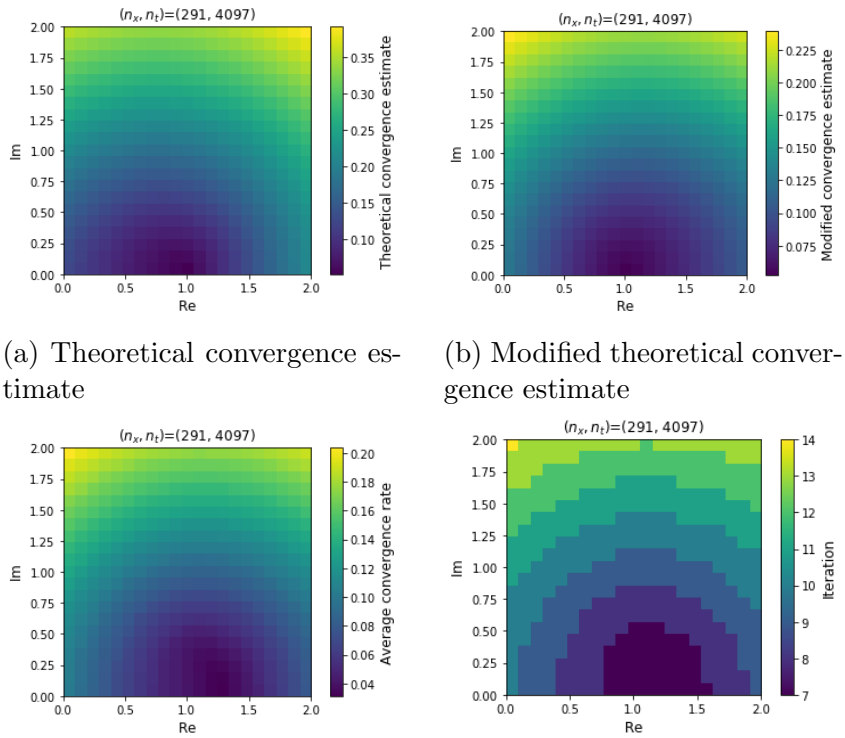
non-unitary weights $\tilde{\omega}_2 \in \mathbb{I}$			
m / Size		291×4097	411×8193
V-cycle	2	$0.6i$: 0.434 (22)	$0.4i$: 0.443 (22)
	4	$0.5i$: 0.368 (19)	$0.5i$: 0.367 (19)
	32	0 : 0.273 (15)	0 : 0.279 (15)

Table 4.4: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Experimentally optimal weights, asymptotic convergence rates and iterations for MGRIT with $\tilde{\omega}_2 \in \mathbb{I}$.

When $\frac{\delta_t}{h_x^2} = 1.0$, the results are again similar to the case of $\frac{\delta_t}{h_x^2} = 12.8$ and are thus omitted for brevity.

4.1.4 Complex valued weight

Lastly, we study non-unitary weights of $\tilde{\omega}_2 \in \mathbb{C}$ for $\frac{\delta_t}{h_x^2} = 12.8$. From Figure 4.3c and 4.3d, the smallest convergence rate and iterations for the two-level method with $m = 2$ are attained when $\tilde{\omega}_{op} = 1.3$, which is the combination of the real valued- and imaginary valued-experimentally optimal weights found in the previous two sections. The original theoretical estimate captured in Figure 4.3a appear to be the same shape as the observed convergence rate, but not sharp. The experimentally optimal weight for the multi-level method with $m = 2$ is $\tilde{\omega}_{op} = 1.3 + 0.6i$, which is also the combination of the experimentally optimal weights found in Table 4.2 and Table 4.4. This produces the smallest convergence rate among the three weight types. When $m = 4$, however, the experimentally optimal weight is not the combination but consists of just the real-valued experimentally optimal weight. For the case of $\frac{\delta_t}{h_x^2} = 1.0$, the experimentally optimal weights are the real-valued experimentally optimal weights when $m = 2$ and 4.



(a) Theoretical convergence estimate

(b) Modified theoretical convergence estimate

(c) Asymptotic convergence rate

(d) Iterations

Figure 4.3: One-dimensional heat equation for $\frac{\delta_t}{h_x^2} = 12.8$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{C}$, considering $m = 2$

4.1.5 Scalability

We now study scalability of MGRIT with the weighted-Jacobi and consider strong- and the weak-scaling.

Strong scaling: Strong scaling aims to minimize the time-to-solution when the problem size is fixed and the number of processors is increased. Amdahl's law states that if p is the parallel fraction of a system, and $(1-p)$ is the serial fraction, then the maximum speedup that can be achieved using N processors is given by

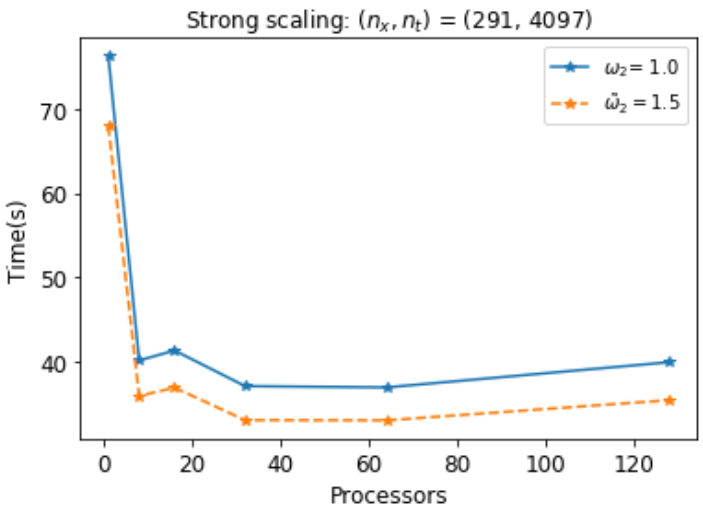
$$S_N = \frac{1}{(1-p) + \frac{p}{N}} \quad \text{and} \quad \lim_{N \to \infty} S_N = \frac{1}{(1-p)}$$

Hence, the speedup is limited by the serial fraction of the system. Our observed speedups are much less than this due to machine limitations.

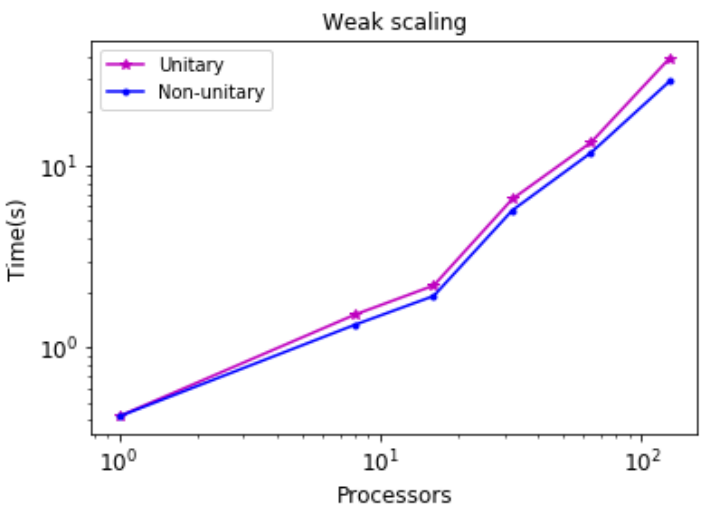
Weak scaling: Weak scaling aims to achieve a constant time-to-solution when the problem size and the number of processors are increased commensurately, e.g., a doubling in problem size is accompanied by a doubling in processors.

Scaling studies: For the strong scaling study, the numbers of processors, $N = 1, 8, 16, 32, 64, 128$, with the fixed problem size of $n_x \times n_t = 291 \times 4097$ and $m = 2$. The number of processors in space is 1, as the code does not support spatial parallelism. Figure 4.4a depicts the comparison in speedup between unitary and non-unitary weights when using V-cycles and the experimentally optimal weight $\tilde{\omega}_2$ for the non-unitary case. A speedup gain of approximately 20% is achieved near the minimum at 32 processors, roughly reflecting the gain in iteration counts.

For the weak scaling study, the number of temporal grid points are increased by a factor of 2 while the spatial grid points are chosen so that CFL number is held fixed at $\frac{h_t}{h_x^2} = 1.0$. The number of processors in space is again 1, and for this reason we expect to see a growth in time to solution as the spatial problem increases, i.e., as we double the number of processors and the number of time points, these factors balance each other out, but the increase in the spatial problem size is not balanced by more processors and will thus lead to a longer time-to-solution. Figure 4.4b depicts a comparison in time-to-solution between unitary and non-unitary weights when using V-cycles and the experimentally optimal weight $\tilde{\omega}_2$ for the non-unitary case. We see a growth in time to solution, and we see a similar speedup when using non-unitary weights that corresponds to the gain in iteration counts.



(a) Strong scaling



(b) Weak scaling

Figure 4.4: Strong and weak scaling studies for the one-dimensional heat equation using $\frac{\delta_t}{h_x^2} = 12.8$ and MGRIT V-cycles with $m = 2$.

4.2 One-dimensional linear advection equation

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

$$\begin{aligned} \frac{\partial u}{\partial t} - \alpha \frac{\partial u}{\partial x} = 0, \quad \alpha > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T] \\ u(x, 0) = u_0(x), \quad x \in \Omega \\ u(0, t) = u(L, t), \quad t \in [0, T] \end{aligned} \tag{4.4}$$

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 in (2.1) 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, $\kappa_\gamma = \frac{i}{h_x} \sin(\frac{2\pi\gamma}{N_x})$ for $\gamma = 1, 2, ..., N_x$, and the formulas (3.8).

4.2.1 Problem statement

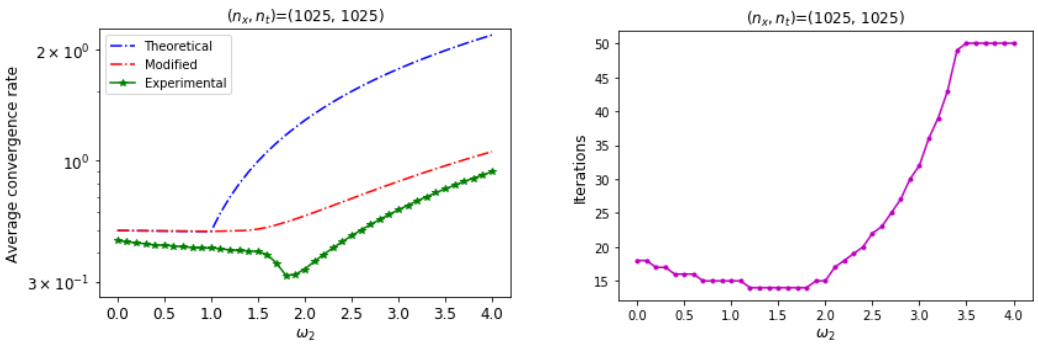
The following function with the given domain is used for numerical experiments. The function is chosen to satisfy the periodic boundary conditions, and is a standard test problem.

$$u(x, t) = e^{-25((x-t)-0.5)^2}$$
$$\alpha = 1, \quad x \in [0, 1], \quad t \in [0, 1]$$

The MGRIT residual norm halting tolerance is set to $\frac{1.0e^{-8}}{\sqrt{h_x \delta_t}}$ and the maximum iteration is set to 70 throughout the experiments. The combination of grid points in space n_x and time n_t are chosen so that a CFL is $\frac{\delta_t}{2h_x} = 0.5$ or 1.0.

4.2.2 Real valued weight

We start by varying $\tilde{\omega}_2 \in \mathbb{R}$ for $\frac{\delta_t}{2h_x} = 0.5$. Figure 4.5a and 4.5b show that the experimentally optimal weight is achieved at $\tilde{\omega}_{op} = 1.8$ for the problem size $n_x \times n_t = 1025 \times 1025$ with $m = 2$. Although the modified theoretical estimate is much sharper than the original, it still does not predict the optimal weight.



(a) Asymptotic convergence rates and theoretical convergence estimates

(b) Iterations

Figure 4.5: One-dimensional linear advection equation for $\frac{\delta_t}{2h_x} = 0.5$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{R}$, considering $m = 2$

Each entry of Table 4.5 is formatted as *experimentally optimal weight : convergence rate (iteration count)*. The different values of $\tilde{\omega}_{op}$ illustrate the difficulty of having a universal optimal weight which produces the best performance for all problem sizes and coarsening factors. Unlike the one-dimensional heat equation, $\tilde{\omega}_{op}$ for a specific m does not work well for all coarsening factor m. In addition, $\tilde{\omega}_{op}$ found in the two-level method does not necessary work well for the multi-level. The value of $\tilde{\omega}_{op}$ for the multi-level are smaller than the values for the two-level with the same m and problem size. As the problem size increases, $\tilde{\omega}_{op}$ does not change. By comparing iterations for ω and $\tilde{\omega}$, MGRIT with $\tilde{\omega}$ reduces iterations for the two- and multi-level methods. The percentage reduction is about $6-9\%$ for the two-level and about $15-24\%$ for the multi-level.

Unitary weights ω

m / Size		1025×1025	2049×2049	3073×3073	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	0.420 (15)	0.425 (15)	0.425 (15)	0.499
	4	0.694 (34)	0.714 (35)	0.716 (35)	0.750
V-cycle	2	0.602 (44)	0.692 (66)	(> 70)	-
	4	0.647 (42)	0.692 (60)	(> 70)	-

Non-unitary weights $\tilde{\omega}_2$

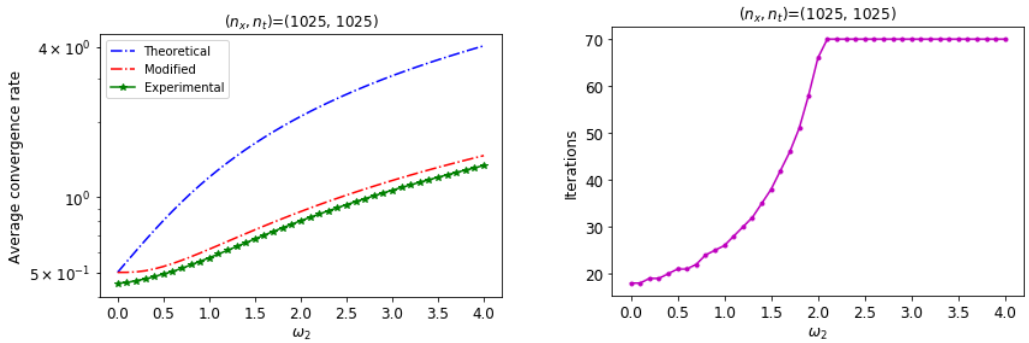
m / Size		1025×1025	2049×2049	3073×3073	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	1.8: 0.323 (14)	1.8: 0.326 (14)	1.8: 0.328 (14)	0.544
	4	1.5: 0.634 (31)	1.5: 0.664 (32)	1.5: 0.676 (33)	0.929
V-cycle	2	1.5: 0.556 (34)	1.6: 0.625 (50)	1.6: 0.685 (65)	-
	4	1.4: 0.597 (35)	1.4: 0.663 (49)	1.4: 0.692 (64)	-

Table 4.5: One-dimensional linear advection equation for $\frac{\delta_t}{2h_x} = 0.5$. Experimentally optimal weights, asymptotic convergence rates and iterations for MGRIT. The modified theoretical estimate appears in the final column.

When $\frac{\delta_t}{2h_x} = 1.0$, the results are similar to the current case of $\frac{\delta_t}{2h_x} = 0.5$. The iterations are reduced by about 6 - 9 % for the two-level method and about 17 - 25 % for the multi-level method. For this table of values, see Appendix C.

4.2.3 Imaginary valued weight

For $\tilde{\omega}_2 \in \mathbb{I}$, the experimentally optimal weights for the two- and multi-level methods are $\tilde{\omega}_{op} = 0$ for various problem sizes, coarsening factors, and $\frac{\delta_t}{2h_x} = 0.5, 1.0$. Figure 4.6b shows that the smallest iterations for the two-level method are 18 iterations which is larger than the iterations for ω. The iterations for the multi-level method are also larger than the iterations for ω. Thus, these results conclude that imaginary valued weights do not contribute to an increased performance.



(a) Asymptotic convergence rates and theoretical convergence estimates

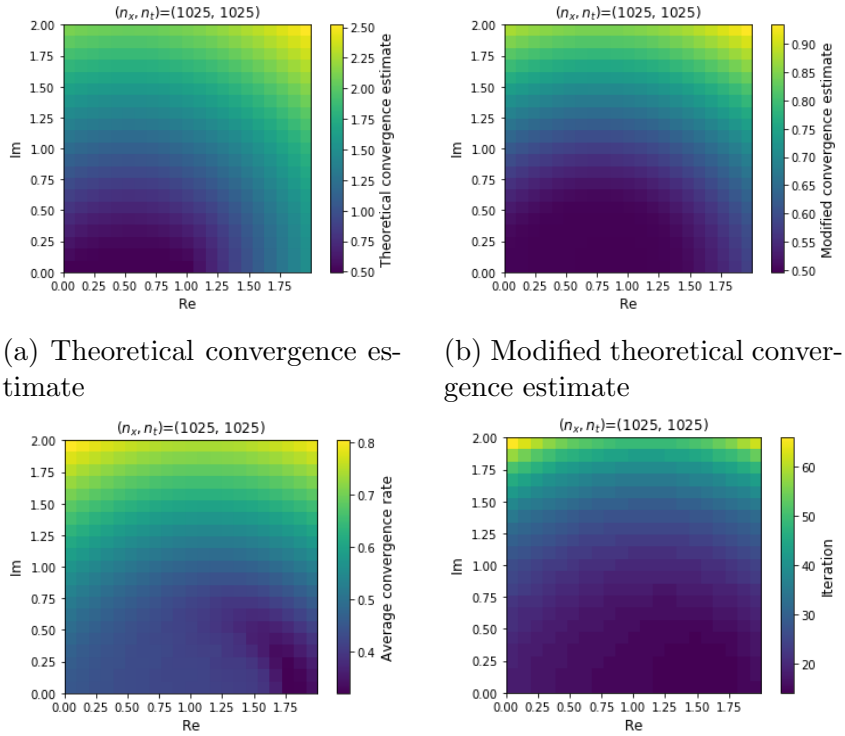
(b) Iterations

Figure 4.6: One-dimensional linear advection equation for $\frac{\delta_t}{2h_x} = 0.5$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{I}$, considering $m = 2$

4.2.4 Complex valued weight

Lastly, we study non-unitary weights of $\tilde{\omega}_2 \in \mathbb{C}$. From Figure 4.7c and 4.7d, the smallest convergence rate and iterations are attained when $\tilde{\omega}_{op} = 1.8$ for $\frac{\delta_t}{2h_x} = 0.5$, which is the combination of the real valued- and imaginary valued- experimentally optimal weights found in the previous two sections. The original theoretical estimate captured in Figure 4.7a does not have the same pattern as the observed convergence rate and is not sharp. For the multi-level method with the same problem size, the

optimal weights are $\tilde{\omega}_{op} = 1.5$ when $m = 2$ and $\tilde{\omega}_{op} = 1.4$ when $m = 4$. It appears that a complex valued-weight does not contribute to better performance.



(a) Theoretical convergence estimate

(b) Modified theoretical convergence estimate

(c) Asymptotic convergence rate

(d) Iterations

Figure 4.7: One-dimensional linear advection equation for $\frac{\delta_t}{2h_x} = 0.5$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{C}$, considering $m = 2$

4.3 One-dimensional advection equation with grid-dependent dissipation

Lastly, we consider the one-dimensional advection equation with grid-dependent dissipation subject to an initial condition and periodic boundary conditions,

$$\begin{aligned} &\frac{\partial u}{\partial t} - \alpha \frac{\partial u}{\partial x} - \epsilon h_x \frac{\partial^2 u}{\partial x^2} = 0 \\ &\alpha > 0, \quad \epsilon > 0, \quad x \in \Omega = [0, L], \quad t \in [0, T] \\ &u(x, 0) = u_0(x), \quad x \in \Omega \\ &u(0, t) = u(L, t), \quad t \in [0, T] \end{aligned} \tag{4.5}$$

By applying the central difference for discretizing the spatial derivatives, we obtain the upwind difference scheme with $\epsilon = 0.5$. The backward difference for discretizing the temporal derivative results in

$$\mathbf{u}_j = (I - \delta_t G)^{-1} \mathbf{u}_{j-1}, \quad j = 1, 2, ..., N_t \tag{4.6}$$

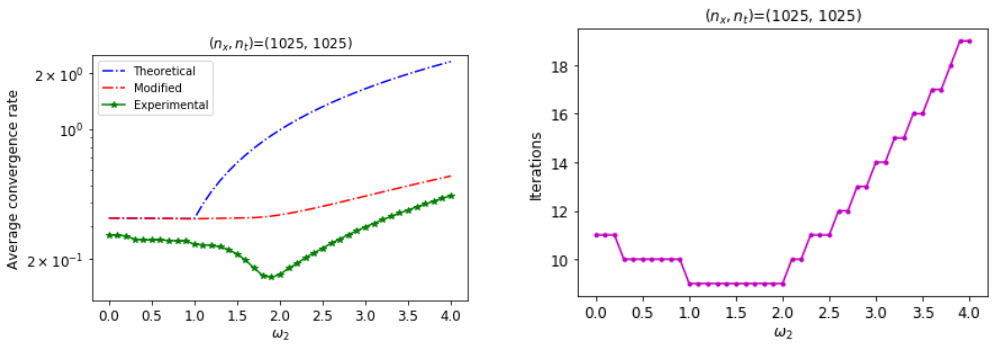
where the linear operator G in (2.1) is the two-point stencil $\frac{\alpha}{h_x}[-1, 1, 0]$. The eigenvalues of G can be computed from the combination of the eigenvalues of a heat equation and a linear advection equation, $\kappa_\gamma = \frac{i}{h_x} \sin(\frac{2\pi\gamma}{N_x}) - \frac{4\epsilon}{h_x} \sin^2\left(\frac{\gamma\pi}{2(N_x+1)}\right)$ for $\gamma = 1, 2, ..., N_x$.

4.3.1 Problem statement

The same function, domains, MGRIT residual norm tolerance, and maximum iterations as the linear advection equation are used. The combination of grid points in space n_x and time n_t are chosen so that a CFL is $\frac{\delta_t}{h_x} = 1.0$.

4.3.2 Real valued weight

We start by varying non-unitary weights of $\tilde{\omega}_2 \in \mathbb{R}$. Figure 4.8b shows that the experimentally optimal weight is $\tilde{\omega}_{op} = 1.9$. Similar to the linear advection equation, the modified theoretical estimate is much sharper but does not predict the optimal weight of $\tilde{\omega_2}$.



(a) Asymptotic convergence rates and theoretical convergence estimates for

(b) Iterations

Figure 4.8: One-dimensional advection equation with dissipation $\frac{\delta_t}{h_x} = 1.0$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{R}$, considering $m = 2$

Each entry of Table 4.6 is formatted as *experimentally optimal weight : convergence rate (iteration count)*. We note that $\tilde{\omega}_{op}$ for the multi-level are again smaller than those for the two-level with the same m and problem size. As the problem size increases, $\tilde{\omega}_{op}$ are fairly stable. The iterations are reduced by about 6% for the two-level and about $10 - 20\%$ for the multi-level.

4.3.3 Imaginary valued weight

For non-unitary weights of $\tilde{\omega}_2 \in \mathbb{I}$, the experimentally optimal weights for the two- and multi-level methods are $\tilde{\omega}_{op} = 0$ for various problem sizes and coarsening factors. Since the iterations for the both methods are larger than those for ω, imaginary

Unitary weights ω

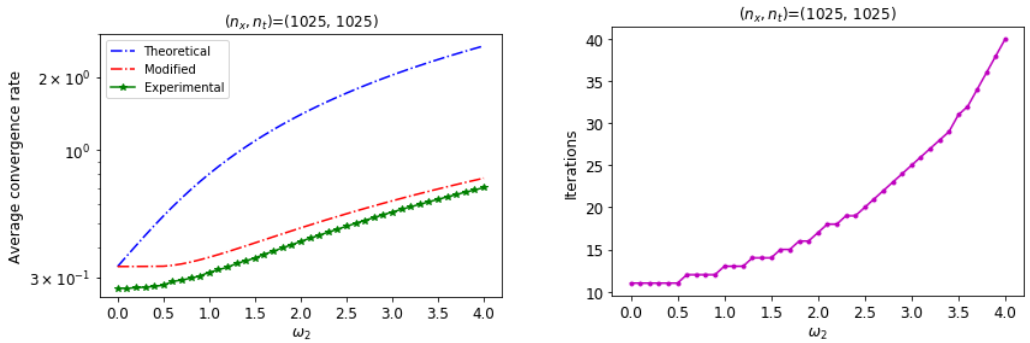
m / Size		1025×1025	2049×2049	3073×3073	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	0.239 (9)	0.240 (9)	0.242 (9)	0.332
	4	0.516 (18)	0.528 (18)	0.534 (18)	0.599
V-cycle	2	0.521 (30)	0.671 (44)	0.671 (57)	-
	4	0.523 (28)	0.646 (41)	0.707 (51)	-

Non-unitary weights $\tilde{\omega}$

m / Size		1025×1025	2049×2049	3073×3073	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	1.9: 0.164 (9)	1.9: 0.162 (9)	1.9: 0.160 (9)	0.340
	4	1.7: 0.469 (17)	1.7: 0.479 (17)	1.7: 0.464 (17)	0.688
V-cycle	2	1.6: 0.445 (25)	1.7: 0.563 (35)	1.5: 0.618 (45)	-
	4	1.4: 0.451 (25)	1.4: 0.604 (35)	1.5: 0.659 (40)	-

Table 4.6: One-dimensional advection equation with dissipation for $\frac{\delta_t}{h_x} = 1.0$. Experimentally optimal weights, asymptotic convergence rates and iterations for MGRIT. The modified theoretical estimates appears in the final column.

weights do not improve performance.



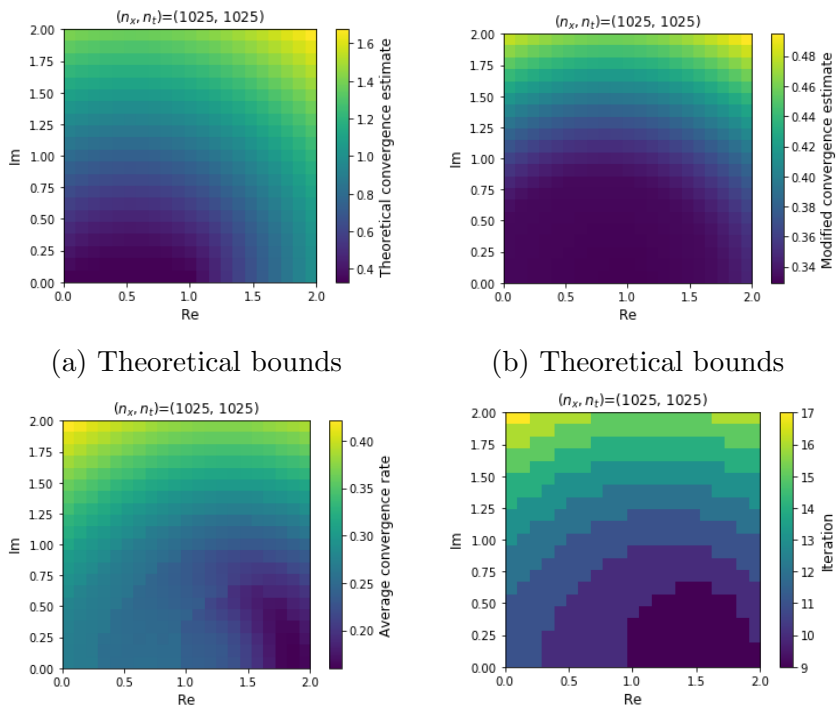
(a) Asymptotic convergence rates and theoretical convergence estimates

(b) Iterations

Figure 4.9: One-dimensional advection equation with dissipation for $\frac{\delta_t}{h_x} = 1.0$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{I}$, considering $m = 2$

4.3.4 Complex valued weight

For modified weights of $\tilde{\omega}_2 \in \mathbb{C}$, Figure 4.7c and 4.7d depict that the smallest convergence rate and iterations are attained when $\tilde{\omega}_2 = 1.9$, which is the combination of the real valued- and imaginary valued- experimentally optimal weights found in the previous two sections. Similar to the linear advection equation, the original theoretical convergence estimate captured in Figure 4.10a does not have the same pattern as the observed convergence rate and is not sharp. For the multi-level method, the experimentally optimal weights are also the real valued experimentally optimal weights. Therefore, complex valued weights do not seem to contribute to better performance.



(a) Theoretical bounds

(b) Theoretical bounds

(c) Asymptotic convergence rate

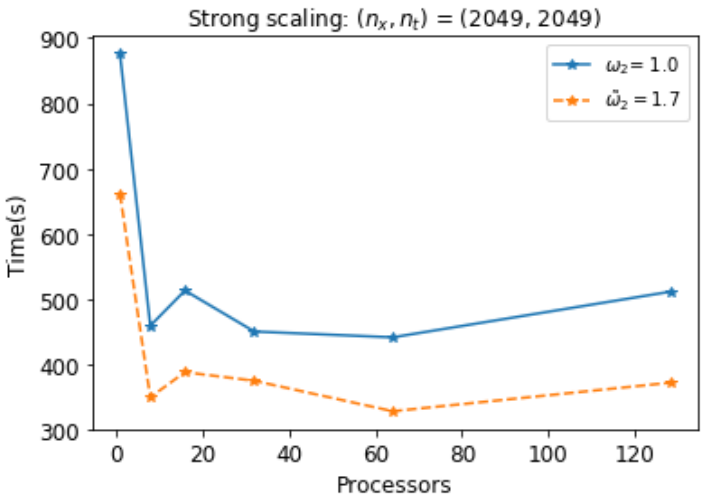
(d) Iterations

Figure 4.10: One-dimensional advection equation with dissipation for $\frac{\delta_t}{h_x} = 1.0$. Two-level MGRIT with $\tilde{\omega}_2 \in \mathbb{C}$, considering $m = 2$

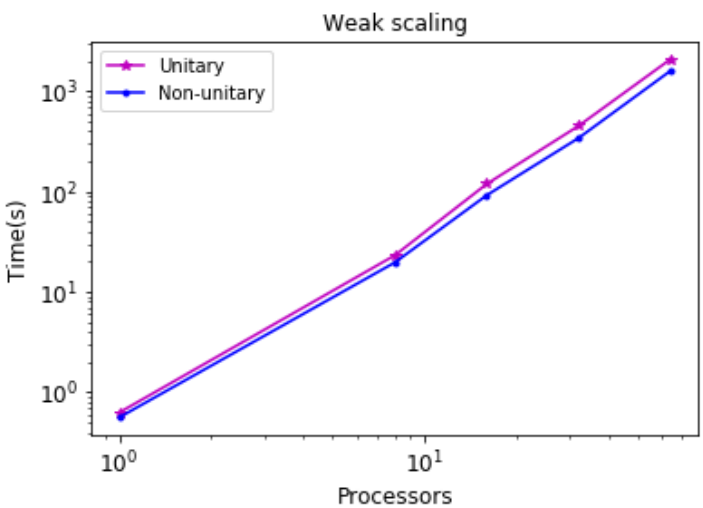
4.3.5 Scalability

A scaling study similar to Section 4.1.5 is now carried out. For the strong scaling study, the numbers of processors, $N = 1, 8, 16, 31, 64, 128$, with the fixed problem size $n_x \times n_t = 2049 \times 2049$ and $m = 2$ are tested. The number of processors in space is again 1, as the code does not support spatial parallelism. Figure 4.11a depicts the comparison in speedup between unitary and non-unitary weights when using V-cycles and the experimentally optimal weight $\tilde{\omega}_2$ for the non-unitary case. A speedup gain of approximately 25% is achieved near the minimum at 64 processors, roughly reflecting the gain in iteration counts.

For the weak scaling study, the problem size is increased by a factor of 2 in both dimensions so that the CFL number is held fixed at $\frac{h_t}{h_x} = 1.0$. The number of processors in space is again 1, and for this reason we expect to see a growth in time to solution as the spatial problem increases, i.e., as we double the number of processors, the problem size increases by a factor of 4, and only the increase in cost in the time dimension is balanced out by more processors. The larger spatial problem size is expected to increase the run-time because there is no spatial parallelism available. Figure 4.11b depicts a comparison in time-to-solution between unitary and non-unitary weights when using V-cycles and the experimentally optimal weight $\tilde{\omega}_2$ for the non-unitary case. We see an expected growth in time to solution, and we see a similar speedup when using non-unitary weights that corresponds to the gain in iteration counts.



(a) Strong scaling



(b) Weak scaling

Figure 4.11: Strong and weak scaling studies for the one-dimensional advection equation with upwinding using $\frac{h_t}{h_x} = 1.0$ and MGRIT V-cycles with $m = 2$.

Chapter 5

Conclusion

Enhancing MGRIT with weighted-Jacobi and the experimentally optimal weight can improve the convergence rate and save iterations. To our knowledge, these are the first results exploring such weighted relaxation with MGRIT. For the one-dimensional heat equation, the optimal weight found for one specific problem size and a coarsening factor of two worked well for various problem sizes and coarsening factors, saving $10 - 15\%$ of the iterations. On the other hand, finding a universal (near) optimal weight for the one-dimensional advection equation is not possible with the tools derived here. While we had hoped that the theoretical convergence estimates would help predict good experimental weight values, this did not turn out to be the case. Even when relying on experiments, finding a good weight can be difficult for advection. For example, if the experimentally optimal weight for the two-level method is used for multi-level, this can yield worse results than the unitary weight. The experimentally optimal weight needs to be varied depending on the problem size and the coarsening factor m. Although there is not a simple way to find the experimentally optimal weight for advection, if a good weight choice is found, it can save $6 - 24\%$ of the iterations. Among the three types of weight values, a real valued weight worked best for advection. For the heat equation, a complex valued

weight can produce the smallest convergence rate, although a real valued weight was almost as effective, and for practical purposes performed as well.

Chapter 6

Future Work

1. Improve the theoretical convergence estimate for powers of the error propagation matrix. See Appendix B.

2. Look at other discretizations such as SDIRK-2 and SDIRK-3 and higher-order in space.

3. Consider F-cycles and modified weights in XBraid.

References

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

[2] A. Brandt, *Multi–level adaptive solutions to boundary–value problems*, Math. Comp., 31 (1977), pp. 333–390.

[3] H. De Sterck, S. Friedhoff, A. J. Howse, and S. P. MacLachlan, *Convergence analysis for parallel-in-time solution of hyperbolic systems*, arXiv preprint arXiv:1903.08928, (2019).

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

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

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

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

[8] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM Journal on Scientific Computing, 29 (2007), pp. 556–578.

[9] A. Hessenthaler, B. S. Southworth, D. Nordsletten, O. Röhrle, R. D. Falgout, and J. B. Schroder, *Multilevel convergence analysis of multigrid-reduction-in-time*, SIAM Journal on Scientific Computing, (2018). submitted, arXiv preprint arXiv:1812.11508.

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

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

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

[13] N. P. V. Dobrev, Tz. Kolev and J. Schroder, *Two-level convergence theory for multigrid reduction in time (mgrit)*, SIAM J. Sci. Comput., 39 (2017), pp. S501–S527.

Appendix A

Theoretical convergence estimate derivations

From the section 3.1.3, the two-level error propagator for $\omega_1 = 1$ is given by

$$
\begin{aligned}
E^{FCF}_{\triangle,\,\omega_1=1} &= (I - \mathbf{B}_{\triangle}^{-1}\mathbf{A}_{\triangle})(I - \omega_2 \mathbf{A}_{\triangle}) \\
&= (\Phi^m - \Phi_{\triangle}) \left\{ (1-\omega_2) \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ \Phi_{\triangle} & 1 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_{\triangle}^{N_T-1} & \Phi_{\triangle}^{N_T-2} & \cdots & 1 & 0 \end{bmatrix} + \omega_2 \Phi^m \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \Phi_{\triangle}^{N_T-2} & \Phi_{\triangle}^{N_T-3} & \cdots & 0 & 0 \end{bmatrix} \right\}
\end{aligned}
$$

Let λ_γ be the eigenvalues of Φ and μ_γ be the eigenvalues of $\Phi_{\triangle}$, corresponding to the same set of eigenvectors $\{v_\gamma\}$. That is, Φ and $\Phi_{\triangle}$ are diagonalized by the same vectors, which is the case for our problems. We can then diagonalize $E^{FCF}_{\triangle,\omega_1=1}$ with $\{v_\lambda\}$ and consider each γ separately. Then, the error propagator for each γ becomes

$$
E^{FCF}_{\triangle,\,\omega_1=1,\,\gamma} = (\lambda_\gamma^m - \mu_\gamma) \left\{ (1-\omega_2) \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ \mu_\gamma & 1 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \mu_\gamma^{N_T-1} & \mu_\gamma^{N_T-2} & \cdots & 1 & 0 \end{bmatrix} + \omega_2 \lambda_\gamma \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ \vdots & \vdots & \ddots & 0 & \\ \mu_\gamma^{N_T-2} & \mu_\gamma^{N_T-3} & \cdots & 0 & 0 \end{bmatrix} \right\}
$$

The infinity norm of the two-level error propagator for each γ can be written as

$$
||E^{FCF}_{\triangle,\,\omega_1=1,\gamma}||_\infty = |\lambda_\gamma^m - \mu_\gamma| \left\{ |1-\omega_2| \sum_j^{N_T-1} |\mu_\gamma|^j + |\omega_2||\lambda_\gamma|^m \sum_j^{N_T-2} |\mu_\gamma|^j \right\} \tag{A.1}
$$

Then, each summation term can be expressed using the geometric series

$$\sum_{j}^{N_T-1} |\mu_\gamma|^j = \frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|}, \quad \text{and} \quad \sum_{j}^{N_T-2} |\mu_\gamma|^j = \frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \tag{A.2}$$

From (A.1) and (A.2), we have

$$\begin{aligned} &||E^{FCF}_{\triangle,\,\omega_1=1,\gamma}||_\infty \\ &= |\lambda_\gamma^m - \mu_\gamma| \left\{ |1-\omega_2| \left(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|} \right) + |\omega_2||\lambda_\gamma|^m \left(\frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \right) \right\} \end{aligned} \tag{A.3}$$

From the construction of the two-level error propagator matrices,

$$||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_1 = ||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_\infty$$

Using this equality and the inequality of matrix norms, it is also true that

$$||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_2 \leq \sqrt{||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_1 ||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_\infty} = ||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_1 = ||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_\infty \tag{A.4}$$

Hence,

$$||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_2 \leq ||E^{FCF}_{\triangle,\omega_1=1,\gamma}||_\infty \tag{A.5}$$

Then, the formula (A.4), the inequality (A.5), and taking the max over all γ implies that

$$\begin{aligned} &||E^{FCF}_{\triangle,\,\omega_1=1}\bar{\mathbf{e}}||_2 \\ &\leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma| \left\{ |1-\omega_2| \Big(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|} \Big) + |\omega_2||\lambda_\gamma|^m \Big(\frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \Big) \right\} ||\bar{\mathbf{e}}||_2 \end{aligned} \tag{A.6}$$

The two-level error propagator from Section 3.1.4 for $\omega_1 \neq 1$ can be derived in the similar way, which results in

$$\begin{aligned} &||E^{FCF}_{\triangle,\,\omega_1\neq 1}\bar{\mathbf{e}}||_2 \\ &\leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma| \{ \Big(|\omega_1\omega_2||\lambda_\gamma|^m - |\omega_2||1-\omega_1||\lambda_\gamma| \Big) \Big(\frac{1-|\mu_\gamma|^{N_T-1}}{1-|\mu_\gamma|} \Big) \\ &\qquad + |1-\omega_2| \Big(\frac{1-|\mu_\gamma|^{N_T}}{1-|\mu_\gamma|} \Big) \} ||\bar{\mathbf{e}}||_2 \end{aligned} \tag{A.7}$$

Appendix B

Theoretical convergence estimate for matrix powers

Based on the results in the paper *Convergence analysis for parallel-in-time solution of hyperbolic systems* [3], we consider the theoretical convergence estimate from the view of powers of the error propagation matrix. The intention is to provide a sharper convergence estimate that predicts the behavior of MGRIT with weighted-Jacobi. From the paper, the powers of the matrices $E^F_{\triangle,\gamma}$ and $E^{FCF}_{\triangle,\gamma}$ with $k \geq 2$ are given by

$$\left(E^F_{\triangle,\gamma}\right)^k = (\lambda_\gamma^m - \mu_\gamma)^k \begin{bmatrix} 0 & & & & & & & \\ \vdots & & & & & & & \\ 1 & 0 & 0 & 0 & & & & \\ \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & & & \\ \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & & \\ \binom{k+2}{3}\mu_\gamma^3 & \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \binom{N_T-1}{N_T-k}\mu_\gamma^{N_T-k} & \cdots & & \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & \cdots & 0 \end{bmatrix}$$

Appendix B. Theoretical convergence estimate for matrix powers

$$\left(E_{\triangle,\gamma}^{FCF}\right)^k = (\lambda_\gamma^m - \mu_\gamma)^k (\lambda_\gamma^m)^k \begin{bmatrix} 0 & & & & & & & & \\ \vdots & & & & & & & & \\ 1 & 0 & 0 & 0 & & & & & \\ \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & & & & \\ \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & & & \\ \binom{k+2}{3}\mu_\gamma^3 & \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & 0 & 0 & & \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ \binom{N_T-1-k}{N_T-2k}\mu_\gamma^{N_T-2k} & \cdots & & \binom{k+1}{2}\mu_\gamma^2 & \binom{k}{1}\mu_\gamma & 1 & 0 & \cdots & 0 \end{bmatrix}$$

The theoretical convergence estimates for F- and FCF-relaxations are derived as

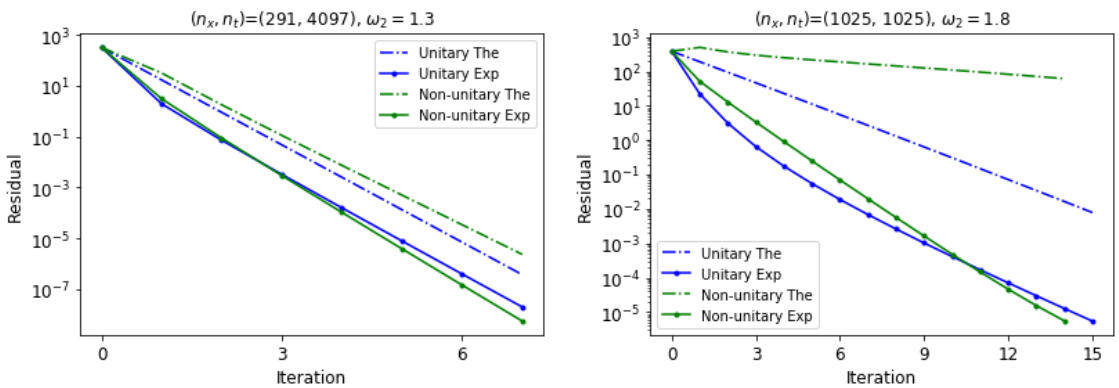
$$||E_{\triangle}^{F}\bar{\mathbf{e}}||_2 \leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma|^k \frac{1}{(k-1)!} \left[\sum_{j=0}^{N_T-k} \left(\prod_{i=1}^{k-1}(j+i)\right) |\mu_\gamma|^j\right] ||\bar{\mathbf{e}}||_2$$

$$||E_{\triangle}^{FCF}\bar{\mathbf{e}}||_2 \leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma|^k |\mu_\gamma^m|^k \frac{1}{(k-1)!} \left[\sum_{j=0}^{N_T-2k} \left(\prod_{i=1}^{k-1}(j+i)\right) |\mu_\gamma|^j\right] ||\bar{\mathbf{e}}||_2$$

Using the above convergence estimates, the theoretical convergence estimate for non-unitary weights with $\omega_1 = 1$ and $\omega_2 \neq 1$ can be written as

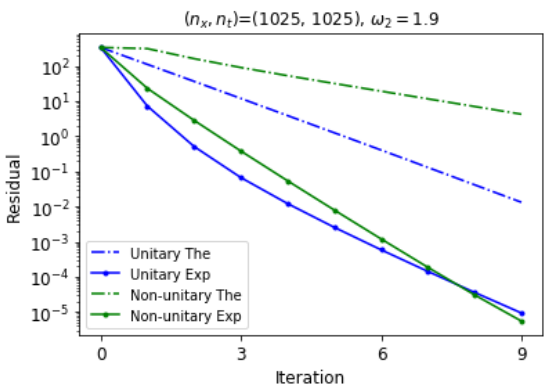
$$||E_{\triangle,\,\omega_1\neq 1}^{FCF}\bar{\mathbf{e}}||_2 \leq \max_\gamma |\lambda_\gamma^m - \mu_\gamma|^k \frac{1}{(k-1)!} \left\{|1-\omega_2|^k \left[\sum_{j=0}^{N_T-k} \left(\prod_{i=1}^{k-1}(j+i)\right) |\mu_\gamma|^j\right] + |\omega_2|^k |\mu_\gamma^m|^k \left[\sum_{j=0}^{N_T-2k} \left(\prod_{i=1}^{k-1}(j+i)\right) |\mu_\gamma|^j\right]\right\} ||\bar{\mathbf{e}}||_2 \tag{B.1}$$

Figure B.1 shows the theoretical convergence estimates (B.1) and experimental convergence rates of unitary and non-unitary weights with respect to iterations. The theoretical estimate is not sharp, especially for advection equation. It also does not predict the behavior of MGRIT with non-unitary weights. It could possibly produce a sharper estimate and a better prediction by combining the terms $(1-\omega_2)$ and ω_2 in (B.1) as it was done in the modified theoretical estimate in the section 3.2.2. However, a path forward to acommplish this was not clear.



(a) Heat equation, $\omega_2 = 1.3$

(b) Linear advection equation, $\omega_2 = 1.8$



(c) Advection equation with dissipation, $\omega_2 = 1.9$

Figure B.1: Theoretical convergence estimates using powers of the error propagation matrix compared with experimental convergence rates for the three model problems. "Unitary The" refers to the theoretical estimate based on matrix powers when using unitary relaxation weights and "Unitary Exp" refers to the experimental data when using unitary relaxation weights. Terms "Non-unitary The" and "Non-unitary Exp" are defined analogously.

Appendix C

Extra Tables

Unitary weights ω

m / Size		1025×513	2049×1025	4097×2049	$\|\|E^{FCF}_{\triangle,\,\omega_1=1}\|\|_2 \leq$
Two-level	2	0.410 (14)	0.416 (14)	0.420 (14)	0.498
	4	0.602 (29)	0.692 (33)	0.712 (34)	0.750
V-cycle	2	0.520 (30)	0.635 (43)	0.713 (67)	-
	4	0.571 (30)	0.652 (41)	0.689 (61)	-

Non-unitary weights $\tilde{\omega}_2$

m / Size		1025×513	2049×1025	4097×2049
Two-level	2	1.8: 0.299 (13)	1.8: 0.308 (13)	1.7: 0.352 (13)
	4	1.4: 0.573 (27)	1.4: 0.663 (30)	1.5: 0.663 (31)
V-cycle	2	1.5: 0.456 (24)	1.6: 0.590 (34)	1.6: 0.615 (50)
	4	1.4: 0.536 (27)	1.4: 0.582 (34)	1.4: 0.638 (50)

Table C.1: One-dimensional advection equation with the central difference scheme for $\frac{\delta_t}{2h_x} = 1.0$. Asymptotic convergence rates and iterations for MGRIT with ω and $\tilde{\omega}$. The theoretical convergence estimate appears in the final column.