Adaptive Algebraic Multigrid for Lattice QCD Computations

Dissertation

zur Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich C - Mathematik und Naturwissenschaften -
der Bergischen Universität Wuppertal vorgelegt von

Dipl.-math. Karsten Kahl

Promotionsausschuß:

Gutachter/Prüfer:	Prof. Dr. Andreas Frommer
Gutachter/Prüfer:	Prof. Dr. James Brannick
Prüfer:	Prof. Dr. Bruno Lang
Prüfer:	Prof. Dr. Roland Pulch

Dissertation eingereicht am:	20. November 2009
Tag der Disputation:	21. Dezember 2009

Acknowledgements

This thesis would not have been written without the support of many people that I would like to thank.

First, I would like to thank my supervisor Prof. Andreas Frommer who sparked my interest for numerical linear algebra in his unparalleled lectures and who made the overall experience of my graduate studies very enjoyable by providing a nice work environment in our work group, advice and inspirational comments. He encouraged me to develop my own path of research and supported me in all my plans to collaborate with a lot of researchers outside the University of Wuppertal. I would also like to thank Prof. James Brannick for his tremendous effort in developing the contents of this thesis. Whitout his tireless help and advice, his devotion to advance linear solvers and the discussions we had this thesis would not be what it became. Beside his influence on my research interests he also introduced me to the Algebraic Multigrid community and enabled a great deal of interesting discussions with researchers in the field.

Next, I would like to thank Rob Falgout who made it possible for me to spend two very productive and interesting summers at the Lawrence Livermore National Laboratory in Livermore, California. The opportunity to get together with leading experts of the field on a daily basis was an unforgettable experience that fuelled my research.

Herein, I especially want to mention the countless discussions with Ira Livshits, and Achi Brandt that inspired many of the developments made in the scope of this thesis.

Last but not least, I would like to thank my family and friends, who gave me the much needed emotional backing with their unwavering support to overcome the hardships of my studies and the writing of this thesis. Herein, a special thank is due to my girlfriend who stood always by my side.

Wuppertal, Karsten Kahl

Contents

1 Introduction **1**

2 Algebraic Multigrid (AMG) Review **5**
2.1 Geometric Multigrid 5
2.2 Generic AMG Components 11
2.3 The AMG$\not{e}$ Framework 24
2.4 Classical AMG 33
2.5 Reduction-based AMG 34
2.6 Aggregation-based AMG 37
2.7 Compatible Relaxation – Coarsening 42

3 Applications **49**
3.1 Lattice Gauge Theory 49
3.2 Markov Chains 62

4 General Framework for AMG **65**
4.1 Adaptive Reduction-based AMG Interpolation 67
4.2 Least Squares AMG Interpolation 78
4.3 Bootstrap AMG Techniques 101

5 Numerical Results **113**
5.1 Scalar Elliptic PDEs 114
5.2 Non-symmetric Eigensolver – Markov Chains 136
5.3 Systems of PDEs – Wilson-Schwinger 146

6 Conclusion and Outlook **161**

Bibliography **165**

List of Figures

2.1 Coupling of variables on equidistant grid for FD Laplace . . . 7
2.2 Eigenvector representation for FD Laplace 9
2.3 ω-Jacobi and Gauss-Seidel for EV of FD Laplace 10
2.4 Two-grid eigenvector representation for FD Laplace 11
2.5 Geometric multigrid vs. algebraic multigrid 12
2.6 Finite element neighborhood 26
2.7 Graph neighborhood in the AMG$\not e$ framework 31
2.8 Aggregation-based coarsening 38
2.9 Piece-wise constant interpolation for linear error 39
2.10 Unsmoothed and smoothed local basis vectors 40

3.1 Naming convention in Lattice Gauge Theory 51
3.2 Spectra of $\Sigma_3 S_W$ and S_W (free and gauged case) 59
3.3 Eigenvector representation of $S_W, \mathcal{U}$ at $\beta = 5$ 60
3.4 Eigenvector representation of $A(\mathcal{U}), \mathcal{U}$ at $\beta = 5$ 61
3.5 Algebraically smooth error for $A(\mathcal{U}), \mathcal{U}$ at $\beta = 5$ 61

4.1 Odd-even reduction of 5-pt operators 74
4.2 Spectral equivalence of $S_{ww}(A^2)$ and A for FE Laplace 100
4.3 Bootstrap AMG setup W-cycle. 106

5.1 Quadrilateral stiffness matrices for Laplace Δ 115
5.2 Coarsening pattern of full-coarsening 115
5.3 Interpolation relations for full-coarsening 116
5.4 Multigrid eigenvector representation for FE Laplace 120
5.5 Bootstrap AMG setup V^2-cycle 120
5.6 Finest grid eigenvector representations for FE Laplace 122
5.7 Computational complexity of W- and V^2-BootAMG setup . . 123
5.8 Stretched quadrilateral stiffness matrices for ani. Laplace . . . 124
5.9 Algebraic distance $(k = 16, \eta = 4)$ for FE ani. Laplace 126
5.10 Algebraic distance $(k = 16(EV), \eta = 4)$ for FE ani. Laplace . . 127

5.11 Eigenvector representation of $A(\mathcal{U}), \mathcal{U}$ at $\beta = 5$ 133
5.12 Multigrid eigenvector representation for $A(\mathcal{U}), \mathcal{U}$ at $\beta = 5$ . . . 135
5.13 EWs of B and $\mathbf{F}(B)$ for TQN with and w/o preconditioning . 140
5.14 Uniform network model 141
5.15 Tandem-Queuing network model 143
5.16 Eigenvector approximations for the TQN problem 144
5.17 Unstructured planar graph network model 145
5.18 Coarsening of UPG by compatible relaxation 145
5.19 EWs of S_W and $\mathbf{F}(S_W)$ for S_W at $\beta = 5$ 150
5.20 Multigrid spectra and $\mathbf{F}(A_l, T_l)$ for S_W with $P_{A(\mathcal{U})}, \mathcal{U}$ at $\beta = 5$ 153
5.21 Multigrid EW configurations of divergence 159

List of Tables

4.1 αAMGr for $A(\mathcal{U}_0)$ 75
4.2 αAMGr for $A(\mathcal{U})$ 77
4.3 αAMGr as preconditioner for CG for $A(\mathcal{U})$ 77

5.1 CR for FE Laplace, full-coarsening 116
5.2 Two-grid LS interpolation for FE Laplace (k vs. η) 117
5.3 Two-grid LS interpolation ($k = 8, \eta = 4$) for FE Laplace . . . 118
5.4 Two-grid LS interpolation ($k = 7 + \mathbf{1}, \eta = 4$) for FE Laplace . 119
5.5 Multigrid LS interpolation ($k = 8, \eta = 4$) for FE Laplace . . . 119
5.6 V^2-BootAMG ($k = 8, \eta = 4$) for FE Laplace 121
5.7 EW approximation of V^2-BootAMG for FE Laplace 121
5.8 W-BootAMG ($k = 8, \eta = 4$) for FE Laplace 122
5.9 V^3-BootAMG ($k = 8, \eta = 4$) for $A(\mathcal{U}_0)$ 129
5.10 V^3-BootAMG ($k = 8, \eta = 4$) for $A(\mathcal{U}_\pi)$ 132
5.11 V^3-BootAMG ($k = 8, \eta = 4$) for $A(\mathcal{U}_{\frac{\pi}{7}})$ 132
5.12 CR for $A(\mathcal{U})$, full-coarsening 133
5.13 Multigrid BootAMG ($k = 8, |\mathcal{V}| = 16, \eta = 4$) for $A(\mathcal{U})$ 134
5.14 BootAMG ($k = 6, \eta = 2$) MLE for uniform network problem . 142
5.15 BootAMG ($k = 6, \eta = 2$) MLE for TQN problem 143
5.16 Accuracy of EV approximations for the TQN problem 144
5.17 BootAMG ($k = 6, \eta = 2$) MLE for UPG problem 146
5.18 BootAMG for $S_W, m_{\lambda_{\min}(A(\mathcal{U})+mI)}, P_{A(\mathcal{U})}$, Kaczmarz 152
5.19 BootAMG for $S_W, m_{\lambda_{\min}(A(\mathcal{U})+mI)}, P_{S_W}$, ω-Richardson, $N = 64$ 155
5.20 BootAMG for $S_W, m_{\lambda_{\min}(A(\mathcal{U})+mI)}, P_{S_W}$, Kaczmarz, $N = 64$ . . 155
5.21 BootAMG for $S_W, m_{\lambda_{\min}(A(\mathcal{U})+mI)}, P_{S_W}$, ω-Richardson, $\beta = 5$. 156
5.22 BootAMG for $S_W, m_{\lambda_{\min}(A(\mathcal{U})+mI)}, P_{S_W}$, Kaczmarz, $\beta = 5$ . . . 157
5.23 BootAMG for $S_W, m_{\min(\Re(\sigma(S_W+mI)))}, P_{S_W}$, ω-Richardson . . . 158
5.24 BootAMG for $S_W, m_{\min(\Re(\sigma(S_W+mI)))}, P_{S_W}$, Kaczmarz 159

Chapter 1

Introduction

The main aim of this thesis is the development and analysis of adaptive algebraic multigrid methods for sparse linear systems of equations. We want to develop adaptive techniques in the algebraic multigrid framework that extend the applicability of algebraic multigrid to a broader class of problems. Our main applications in mind are the linear systems arising in Lattice Gauge Theory, a field that not only is one of the most important and interesting fields in modern physics, but also features a lot of interesting challenges for the application of (algebraic) multigrid methods. Although these problems guide the development of our adaptive algebraic multigrid techniques and methods in the following by making high demands on the properties of such a multigrid method, we emphasize on the general applicability of the adaptive techniques and methods developed and analyzed in this thesis.

The Quantum Dynamics is a theoretical model that describes the interaction of particles (fermions) and their interacting counterpart (bosons), e.g., electrons and photons in Quantum Electrodynamics (QED), quarks and gluons in Quantum Chromodynamics (QCD). These interactions are modeled by the Dirac operator,

$$\mathcal{D} = \frac{1}{2} \sum_{\mu} \gamma_\mu \otimes \left(\partial_\mu + i A_\mu \right),$$

in presence of a background gauge-field A_μ. Formally, the Dirac operator without a background gauge-field is a square-root of the Laplace operator, Δ, and can be seen as its relativistic counterpart. In order to simulate the interactions of particles and compute observables of interest, this coupled system of partial differential operators is discretized on a finite lattice and thus becomes Lattice Gauge Theory. The systems of Lattice Gauge Theory present numerous challenges for the development of linear solvers, especially

multigrid solvers. Due to the highly disordered coupling in the system, introduced by the background gauge-field A_μ, the eigenvectors associated with the small eigenvalues become oscillatory and are locally supported. The systems are ill-conditioned with a large number of eigenvectors associated with small eigenvalues depending on the formulation of the discretization. Further, the systems arising in Lattice Gauge Theory are complex-valued and non-hermitian, adding to the challenges already mentioned. The algorithmic techniques developed in this thesis, although motivated by these applications, lead to new adaptive algebraic multigrid techniques and methods that are generally applicable to a broad range of problems.

This thesis is organized as follows. In chapter 2 we give an introduction into state-of-the-art algebraic multigrid techniques and theory, covering a broad range of approaches to define algebraic multigrid methods and stating the main theoretical results found in the algebraic multigrid literature. In chapter 3 we present the applications we have in mind when developing the adaptive algebraic multigrid techniques. Besides a more detailed description of the systems arising in Lattice Gauge Theory, we also introduce the static Markov chain model as an interesting problem which nicely motivates certain parts of the adaptive multigrid framework that we discuss in great detail in chapter 4. We introduce several adaptive algebraic multigrid techniques and present theoretical analysis for these approaches. Last, we present numerical experiments for the adaptive algebraic multigrid framework in chapter 5, before concluding the thesis with remarks on open questions, ideas of improvement and future research in chapter 6.

In the opening chapter 2, we give a general introduction into algebraic multigrid methods and theory. Here, the main focus is on the formulation of existing approaches in algebraic multigrid in a unifying framework based on the principles of element-free AMG$\not e$, presented and analyzed in section 2.3. In this way, we provide insight into the relationship between the reviewed methods and pave the road for new developments made in chapter 4 that are formulated in the unified framework as well.

Chapter 3 contains a review of the two main applications we consider in this thesis. First, in section 3.1, we introduce the systems arising in Lattice Gauge Theory, including a discussion of important properties of these operators that motivate certain choices in the development of an adaptive algebraic multigrid method. Furthermore, we outline milestones on the way to solving the Wilson-Dirac operator of Lattice Quantum Chromodynamics and stress out the main challenges posed by the problems of Lattice Gauge Theory. The other main applications are time-independent Markov chain models, encountered in the description of discrete stochastic processes. In section 3.2, we introduce the notation and the main features of these problems. The

challenges encountered in these two applications range from complex-valued systems, over non-hermiticity or indefiniteness to singular or nearly singular systems.

The main contribution of this thesis is contained in chapter 4, namely the development of adaptive algebraic multigrid algorithms and theory. Herein, we considered two main approaches. First, in section 4.1 we introduce a modified and generalized version of the adaptive reduction-based algebraic multigrid approach. We develop a generalized theory that allows us to handle complex-valued systems, and develop a new solver for such systems in the hermitian positive definite setting. Although the work on adaptive reduction-based algebraic multigrid leads to several new results the proposed method has limited success in treating all challenges of the systems arising in Lattice Gauge Theory. Hence, we consider developing a much more generally applicable method in a bootstrap algebraic multigrid framework. In sections 4.2 and 4.3 we develop a first practical algorithm based on Brandt's bootstrap algebraic multigrid framework and significantly build on the work started by Brandt and Livne in [BL04]. Our contributions include a thorough study of least squares based interpolation, a multilevel eigensolver that is used in the bootstrap setup and a handful of additional bootstrap techniques. We provide an analysis of these formulations as rigorous and complete possible to date.

Finally, in chapter 5 we apply the bootstrap algebraic multigrid framework developed in chapter 4 to the applications we described in chapter 3. More specifically, we start by testing the method for discretized scalar partial differential equations in section 5.1. Herein, we analyze the components of the algorithm one at a time, increasing the difficulties involved in the test problems and finally successfully testing the method on Gauge Laplace systems with physical background gauge configurations. This part is limited to the discussion of hermitian positive definite systems and demonstrates the power of the new adaptive formulation in a setting where much knowledge on the behavior of algebraic multigrid methods is available, though extending the applicability to the randomly-coupled Gauge Laplace systems. Following this introductory section, we present in section 5.2 an algorithm for the systems arising in Markov chain processes. The proposed method which combines the multilevel eigensolver with least squares interpolation and Krylov subspace methods yields very promising results, suggesting that this approach – or some of the techniques we used – might significantly improve solvers for this setting. Our method, besides the use of the bootstrap framework, can be regarded as a merge of the work by Virnik in [Vir07], where she introduced the use of reduction-based algebraic multigrid as a preconditioner of GMRES for the solution of Markov chain problems and the idea of an alge-

braic multigrid eigensolver, introduced in the aggregation-based framework (cf. [Sch91]) and further developed in this framework in [DSMM$^+$09]. In the concluding section of the last chapter, section 5.3, we present preliminary results for the application of the bootstrap algebraic multigrid framework for the Wilson-Schwinger operator of Lattice Gauge Theory. The initial results for a multigrid solver are promising, though we point out a few obstacles that yet have to be overcome in order to get a very robust multigrid method. Though, in contrast to the difficulties we encounter defining a true multigrid solver for these non-hermitian systems, the proposed algorithm turns out to be a very efficient preconditioner for Krylov subspace methods (e.g., GMRES) in this context. Motivated by these observations, we outline paths of future development of the proposed techniques that lead to an efficient solver for systems arising in Lattice Gauge Theory and that is able to significantly cut down the costs of existing adaptive algebraic multigrid methods which are mainly formulated in a (smoothed) aggregation-based framework (cf. [BBC$^+$08]).

Chapter 2

Algebraic Multigrid (AMG) Review

In this opening chapter we review the principles of algebraic multigrid, beginning with a motivating excursion into geometric multigrid in section 2.1. Thereafter, we introduce the necessary notations and language of algebraic multigrid in section 2.2, including theoretical results for algebraic multigrid methods. In the following sections 2.3 to 2.6 we present several state-of-the-art variants of interpolation in algebraic multigrid. Our focus here is to formulate the known approaches in the framework of element-free algebraic multigrid (AMG$\not e$) that allows us to compare and discuss the approaches in a unified setting. The chapter is concluded by the introduction of compatible relaxation in section 2.7, a tool to generate and measure coarse-variable sets using the multigrid relaxation.

2.1 Geometric Multigrid

The development of multigrid methods was originally motivated by a natural geometric interpretation of iterative methods for simple boundary value problems. In this section, we present this motivating interpretation, following the discussion in the introductory chapters of [BHM00]. Similar descriptions of geometric multigrid can be found in [Hac91].

For the sake of simplicity, we consider the two-dimensional boundary value problem

$$\begin{aligned} -u_{xx} - u_{yy} &= f(x,y), \quad (x,y) \in \Omega = (0,1)^2 \\ u(x,y) &= 0, \quad (x,y) \in \delta\Omega. \end{aligned}$$

Discretization by second-order finite differences on an equidistant grid with grid-spacing $h = h_x = h_y = \frac{1}{N}$ yields a set of $m = (N-1)^2$ linear

equations

$$\begin{gathered}\frac{-u_{i-1,j}+2u_{ij}-u_{i+1,j}}{h^2}+\frac{-u_{i,j-1}+2u_{ij}-u_{i,j+1}}{h^2}=f_{ij},\\ u_{i0}=u_{im}=u_{0j}=u_{mj}=0,\end{gathered} \tag{2.1}$$

where $f_{ij}=f\left(ih_x,jh_y\right)$, and $i,j=1,\ldots,N-1$. Introducing lexicographically ordered vectors

$$u=\begin{pmatrix}u_{11} & u_{21} & \ldots & u_{N-1,1} & u_{1,2} & \ldots & u_{N-1,N-1}\end{pmatrix}^T$$

and

$$f=\begin{pmatrix}f_{11} & f_{21} & \ldots & f_{N-1,1} & f_{1,2} & \ldots & f_{N-1,N-1}\end{pmatrix}^T$$

this system of linear equations reads

$$Au=h^2f, \tag{2.2}$$

with A given given by

$$A=\begin{pmatrix}B & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B & -I \\ & & & -I & B\end{pmatrix},$$

where each block B is given by

$$B=\begin{pmatrix}4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4\end{pmatrix}.$$

As illustrated in Figure 2.1 each variable is coupled only to its direct neighbors. Thus, it is common practice to denote the linear system (2.1) in stencil notation

$$A=\frac{1}{h^2}\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix}.$$

We use this notation throughout the thesis for discretizations of PDEs on equidistant grids.

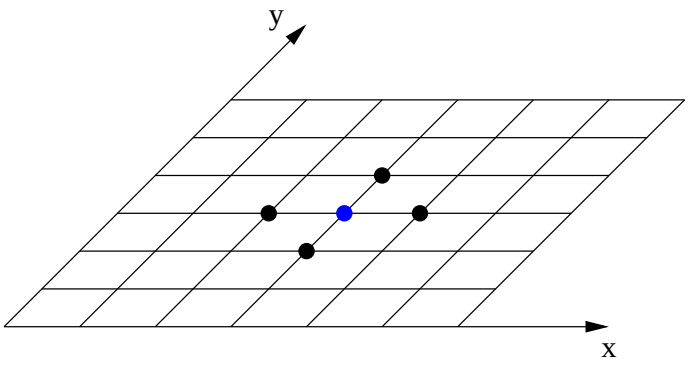


Figure 2.1: The coupling of a typical variable in the discrete equations (2.1) on the two-dimensional equidistant grid.

Solving such sparse linear systems has been a topic of intense research for many years. One possible approach for solving such sparse systems is to apply a direct solution technique like Gaussian elimination. The computational complexity of the best direct methods for two-dimensional elliptic equations which are based on fast Fourier transforms and cyclic reduction schemes, is $\mathcal{O}(N^2 \log(N))$, which is close to optimal complexity $\mathcal{O}(N^2)$. Such approaches are however limited to two dimensional problems with constant coefficients.

Alternative methods for solving sparse linear systems of equations are given by iterative methods. Interesting enough, the inefficiency of stationary iterative methods for discretizations of PDEs motivates the multigrid solution of such systems as we explain below.

2.1.1 Basic iterative methods

Assuming that the sparse linear system of equations (2.2) has a unique solution u^* we can define the error e of an approximate solution u by

$$e = u^* - u.$$

We are then interested in iterative schemes that start with an arbitrary initial approximation and converge to the solution, i.e., eliminate the error. Unfortunately, the error is in general inaccessible, such that we introduce a measure for the quality of the current iterate that is available during the iteration and that gives some information about the convergence process. This is the residual, r, defined by

$$r = f - Au.$$

The residual measures how well the current solution fulfills the system of equations. It is well known that the residual can be misleading, i.e., a small residual does not imply that the corresponding error is small as well.

Indeed, $e = A^{-1}r$ and thus $\|e\|_2 \leq \|A^{-1}\|_2\|r\|_2$. In the case that $\|A^{-1}\|_2$ is large, $\|r\|_2$ can be small, while $\|e\|_2$ can be large.

Recalling that $Au^* = f$, we see that the error fulfills the equation

$$Ae = r. \tag{2.3}$$

Hence, we can use this equation in order to improve a given approximation u^k of $Au = f$ by solving (2.3) for e^k and defining a new approximation by

$$u^{k+1} = u^k + e^k.$$

The idea of residual correction is the basis of almost all iterative schemes to solve systems of linear equation. Though, in general the actual application has to be done more carefully than indicated here.

Given the system matrix A, we consider the splitting,

$$A = D - L - U, \tag{2.4}$$

into a diagonal part D and a strictly lower, L, and upper triangular part, U, respectively.

A simple iterative scheme that can be defined using this splitting is the so-called *Jacobi* iteration. Writing (2.4) as

$$Du = (L + U)u + f$$

we solve by inverting the diagonal part D and get the new iterate as

$$u^{k+1} = D^{-1}\left(L + U\right)u^k + D^{-1}f = \left(I - D^{-1}A\right)u^k + D^{-1}f = u^k + D^{-1}r^k.$$

By defining an intermediate iterate and the next iterate as a weighted average we get the ω-*Jacobi* method as

$$u^{k+1} = \left(I - \omega D^{-1}A\right)u^k + \omega D^{-1}f = u^k + \omega D^{-1}r^k.$$

Another possible way to define an iterative process exploiting the splitting (2.4) is the so-called *Gauss-Seidel* iteration defined by

$$u^{k+1} = \left(I - (D - L)^{-1}A\right)u^k + (D - L)^{-1}f = u^k + (D - L)^{-1}r^k. \tag{2.5}$$

It is easy to see that Gauss-Seidel implies a certain ordering of the variables, which might imply sequential processing. Exploiting the sparsity of A it is

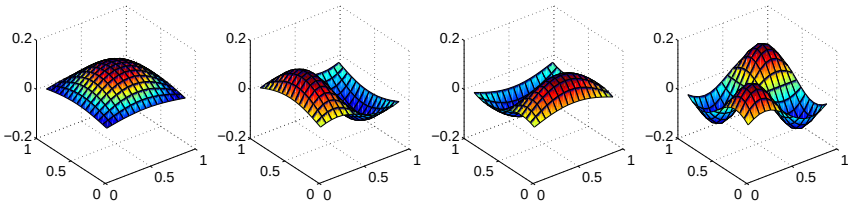


Figure 2.2: Eigenvectors $z^{(k,l)}$ corresponding to $\lambda_{1,1}, \lambda_{1,2}, \lambda_{2,1}$ and $\lambda_{2,2}$ of (2.1) on a 15×15 grid.

possible to get back to a partly simultaneous iteration using an appropriate coloring scheme, e.g., red-black and re-ordering the variables according to this coloring scheme.

Stationary iterative methods that use the same formulation in each step such as Jacobi, ω-Jacobi and Gauss-Seidel can always be written as

$$u^{k+1} = u^k + Br^k,$$

where B can be seen as an approximation to A^{-1}, with $B = D^{-1}$, $B = \omega D^{-1}$ and $B = (D - L)^{-1}$ for Jacobi, ω-Jacobi and Gauss-Seidel, respectively. The choice $B = A^{-1}$ yields the solution u^* in just one iteration.

Next, we discuss the performance of stationary iterative methods when applied to the discretized Poisson problem (2.1) and how this leads to a natural motivation of multigrid.

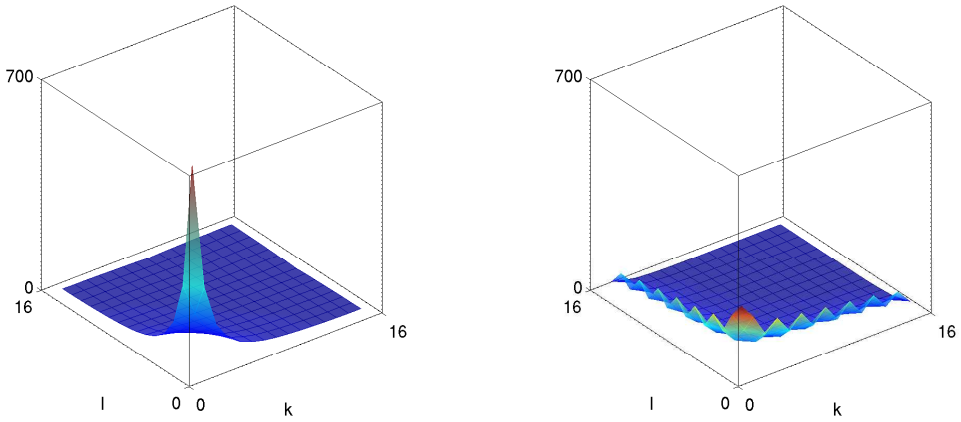
The eigenvalues of the discrete Poisson problem (2.1) on a $(N-1)\times(N-1)$ grid are given by

$$\lambda_{k,l} = 4 - 2\left(\cos\left(\frac{k\pi}{N}\right) + \cos\left(\frac{l\pi}{N}\right)\right), \quad k, l = 1, \ldots, N-1,$$

with corresponding eigenvectors $z^{(k,l)}$ given by

$$z_{i,j}^{(k,l)} = \sin\left(\frac{k\pi i}{N}\right)\sin\left(\frac{l\pi j}{N}\right), \quad i, j = 1, \ldots, N-1.$$

In Figure we present the eigenvectors associated to the eigenvalues with wave-numbers $(1,1), (1,2), (2,1)$ and $(2,2)$, i.e., the eigenvectors corresponding to the smallest 4 eigenvalues of (2.1) on a 15×15 grid. To analyze the behavior of stationary iterative methods we consider applying them to (2.1) with initial error e defined as the eigenvector $z^{(k,l)}$. In Figure 2.3 the number of iterations needed to reduce the error by a factor of 10^3 is given. For this



(a) ω-Jacobi (b) Gauss-Seidel

Figure 2.3: Number of iterations needed to reduce the initial error $z^{(k,l)}$ by a factor of 10^3 on a grid of size 15×15. In 2.3(a) using ω-Jacobi with $\omega = \frac{2}{3}$ and in 2.3(b) using Gauss-Seidel.

system, it occurs that the stationary iterations are able to reduce error corresponding to oscillatory modes with large "wave-number" k, l much better than the smooth modes corresponding to small "wave-numbers" k, l.

This "smoothing property" – reduction of oscillations of an arbitrary error – amounted in the naming convention for such sort of iterative methods as *smoother* or *relaxation*. Throughout this thesis, these names will be used interchangeably when talking about a stationary iterative method that fulfills a "smoothing property". In section 2.2 we generalize this property to an algebraic framework, where "smoothness" no longer implies geometric smoothness.

As shown in Figure 2.4, smooth error can be well approximated on a grid that consists of every other grid point and interpolating linearly in between. Furthermore, the coarse version of the smooth mode becomes more oscillatory with respect to the coarse grid, hence it can be treated more efficiently by relaxation on this grid.

The idea of geometric multigrid is to combine relaxation with a coarse-grid correction that exploits the strength of relaxation for oscillatory errors and the natural representation of smooth error on coarse grids using geometric (linear, cubic, ...) interpolation.

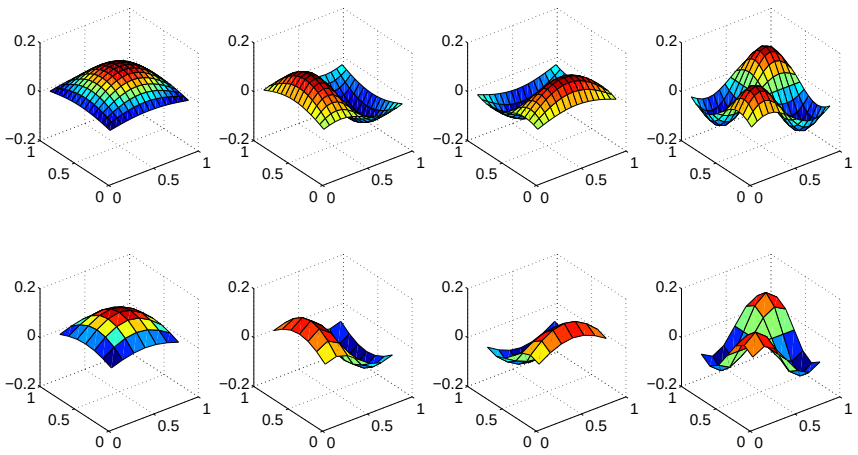


Figure 2.4: Eigenvectors $z^{(k,l)}$ corresponding to $\lambda_{1,1}, \lambda_{1,2}, \lambda_{2,1}$ and $\lambda_{2,2}$ of (2.1) on a 15×15 and the corresponding coarse 7×7 grid.

2.2 Generic AMG Components

Introduced in the early 1980s in [BMR84, Bra86] algebraic multigrid (AMG) is an approach for solving the sparse linear system of equations

$$Au = f \quad \text{or} \quad \sum_{j=1}^{m} a_{ij} u_j = f_i, \;\; i = 1, \ldots, m, \tag{2.6}$$

without the assumption that the problem originates from a discretization of a known partial differential equation (PDE), as in geometric multigrid.

AMG can formally be described in a similar way as geometric multigrid by replacing the terms *grids*, *sub-grids* and *grid points* with *sets of variables*, *subsets of variables* and *single variables*, respectively. Similar to coarse grids in geometric multigrid, successive systems of equations of reduced dimension are used in algebraic multigrid to eliminate components of the error in the coarse-grid subspace. In most situations the coarse-grid subspaces correspond to the subspace spanned by the lower end of the spectrum, i.e., eigenvectors to small eigenvalues, as they are not treated well by most stationary iterative methods used as smoothers. In contrast to geometric multigrid, where the grid hierarchy is known and appropriate smoothers have to be defined to achieve multigrid optimality, the goal of classical AMG, as stated in [RS86], is to maintain simple smoothers and find suitable coarse-grid problems that yield an efficient interplay between smoother and coarse-grid correction. In

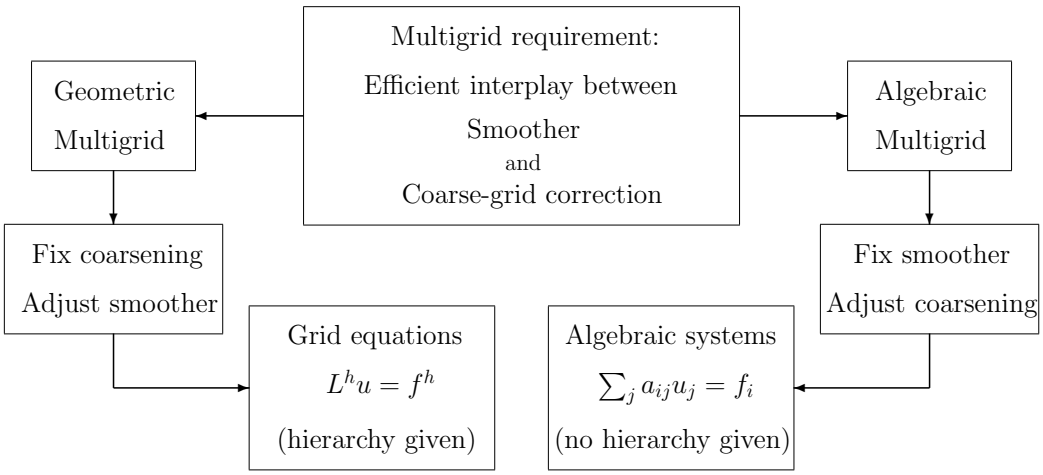


Figure 2.5: Mindsets of geometric vs. algebraic multigrid (cf. [Stü01, p. 416])

Figure 2.5 we illustrate the differences of the approaches in geometric and algebraic multigrid.

In this section, we introduce the generic notation for algebraic multigrid and present some fundamental results on convergence theory for algebraic multigrid in the hermitian positive definite case. These results guide the discussion throughout the remainder of this thesis.

2.2.1 Notations

In order to define an algebraic multigrid method for the linear system of equations (2.6), we must define successively smaller linear systems

$$A^l u^l = f^l \quad \text{or} \quad \sum_{j=1}^{m_l} a_{ij}^l u_j^l = f_i^l, \; i = 1, \ldots, m_l, \tag{2.7}$$

where $l = 0, \ldots, L$ denotes the number of the grid in the hierarchy (i.e., its level), with $A^0 = A$ and $m = m_0 > m_1 > \ldots > m_L$. Herein

$$A^l \in \mathbb{C}^{m_l \times m_l}, u^l \in \mathbb{C}^{m_l}, f^l \in \mathbb{C}^{m_l},$$

and we often refer to A^{l+1} as the coarse-grid system with respect to the fine-grid system A^l, for $l = 0, \ldots, L-1$.

In order to define an analogue of the geometric interpretation of the relation between variables on successive levels, we first define an artificial "grid", given by the sparsity of the system of linear equations.

Definition 2.1. *Let the graph $G_A = (V_A, E_A)$ with vertices V_A and edges E_A corresponding to $A = (a_{ij})_{i,j=1,\dots,m}$ be defined as*

$$V_A = \{1, \dots, m\} \quad \textit{and} \quad E_A = \{(i,j) \in \{1, \dots, m\}^2 : a_{ij} \neq 0\}.$$

Although this artificial geometry does not necessarily resemble any physical geometry, it does allow us to talk about distances and locality in an algebraic framework. As we assume the original system of linear equations (2.6) to be sparse, it inhibits a locality in terms of Definition 2.1. The aim is to preserve this kind of locality and with it the sparsity of the operators in the multigrid hierarchy.

As a first component of algebraic multigrid, in analogy to geometric multigrid, we introduce an iterative method that serves us as a smoother.

Definition 2.2. *The* smoother *or* relaxation $\mathcal{S}_l$ *is a linear operator*

$$\mathcal{S}_l : \mathbb{C}^{m_l} \times \mathbb{C}^{m_l} \longrightarrow \mathbb{C}^{m_l},$$

such that $\mathcal{S}_l\left(u^l, f^l\right) = u_l$ iff $A^l u^l = f^l$.

Remark 2.3. *We denote the error propagator of the smoother $\mathcal{S}_l$, by S_l. In general, we assume that the smoother is convergent, i.e.,*

$$\lim_{\nu \to \infty} S_l^\nu e = 0, e \in \mathbb{C}^{m_l},$$

even if this is no necessary condition for multigrid convergence.

Analogously to geometric multigrid, we assume that the given smoother $\mathcal{S}_l$ reduces the complexity of (2.7) on level l by reducing certain error components very efficiently. Hence, after application of just a few iterations of the smoother we can accurately represent the system on the next coarser level, level $l+1$, using fewer variables.

With this notion of an artificial geometry, we can further describe the relation between variables.

Remark 2.4. *According to [Bra00], there exists a linear mapping Q between fine-grid variables and coarse-grid variables, i.e.,*

$$u^{l+1} = Q u^l.$$

Thus, for each coarse-grid variable $u_i^{l+1},\ i = 1, \dots, m_{l+1}$ we have

$$u_i^{l+1} = \sum_{j=1}^{m_l} q_{ij} u_j^l,\ i = 1, \dots, m_{l+1},\ l = 0, \dots, L-1.$$

Hence, coarse-grid variables can be interpreted as re-combinations of fine-grid variables.

Remark 2.5. *In most cases we assume that coarse-grid variables are given as local re-combinations of fine-grid variables. Hence, $q_{ij} \neq 0$ only for j close to i on the grid defined by the graph G_A.*

The special case of $q_{ij} = \delta_{ij}$ corresponds to a partitioning of the variables $\Omega_l = \{1, \ldots, m_l\}$ into a set of variables that appear on the next coarser grid $\mathcal{C}^l$, and its complement set $\mathcal{F}^l$, i.e., $\Omega_l = \mathcal{F}^l \cup \mathcal{C}^l$.

Closely connected to the definition of coarse-grid variables is the definition of inter-grid transfer operators in algebraic multigrid.

Definition 2.6. *Given the system of linear equations $A^l u^l = f^l \in \mathbb{C}^{m_l}$ define full-rank interpolation operators P_{l+1}^l and restriction operators R_l^{l+1},*

$$P_{l+1}^l : \mathbb{C}^{m_{l+1}} \longrightarrow \mathbb{C}^{m_l} \quad \text{and} \quad R_l^{l+1} : \mathbb{C}^{m_l} \longrightarrow \mathbb{C}^{m_{l+1}},$$

respectively, and define the coarse-grid system of equations as the Galerkin operator

$$A^{l+1} = R_l^{l+1} A^l P_{l+1}^l \in \mathbb{C}^{m_{l+1} \times m_{l+1}}.$$

Remark 2.7. *Whenever it is clear from the context, e.g., for discussions of a two-level setting, we omit the sub- and superscripts for u^l, f^l, P_{l+1}^l, R_l^{l+1}, $\mathcal{S}_l$, S_l, A^l and write u, f, P, R, $\mathcal{S}$, S, A instead. Furthermore, we denote all corresponding coarse-grid entities u^{l+1}, f^{l+1}, A^{l+1} as u^c, f^c, A^c.*

With the smoother, grid and inter-grid components defined, Algorithm 1 implements a generic algebraic multigrid V(ν_1, ν_2)-cycle, which is analyzed in the next subsection.

Algorithm 1 amg_solve { Generic AMG V(ν_1, ν_2) cycle }

Input: A^l, f^l
Output: u_l
if $l = L$ **then** {Direct Solve}
  $u^l = \left(A^l\right)^{-1} f^l$
else
  $u^l = \mathcal{S}_l^{\nu_1}\left(0, f^l\right)$ {Pre-smoothing}
  $f^{l+1} = R_l^{l+1}\left(f^l - A^l u^l\right)$ {Restriction}
  $A^{l+1} = R_l^{l+1} A^l P_{l+1}^l$
  $v^{l+1} = \text{amg\_solve}\left(A^{l+1}, f^{l+1}\right)$ {Solve coarse-grid system}
  $u^l = u^l + P_{l+1}^l v^{l+1}$ {Coarse-grid correction}
  $u^l = \mathcal{S}_l^{\nu_2}\left(u^l, f^l\right)$ {Post-smoothing}
end if

In order to analyze the error propagation of the AMG V(ν_1, ν_2)-cycle we begin with a two-grid scheme.

Proposition 2.8. *For $L = 1$ the error propagator of coarse-grid correction C in Algorithm 1 is given by*

$$C = \left(I - P\left(A^c\right)^{-1} RA\right). \tag{2.8}$$

Proof. Let $\breve{u}$ be the solution of $Au = f$ and $e = \breve{u} - u$ the error after coarse-grid correction. With $\left(A^c\right)^{-1} f^c = \text{amg\_solve}(A^c, f^c)$ and $u = \tilde{u} - P\left(A^c\right)^{-1} R\left(f - A\tilde{u}\right)$, where $\tilde{u}$ denotes the iterate after pre-smoothing, and the fact that

$$f - A\tilde{u} = A\breve{u} - A\tilde{u},$$

we obtain (2.8). □

In the case that $L > 1$ – a true multigrid situation – we introduce the notation

$$\left(\tilde{A}^{l+1}\right)^{-1} = \text{amg\_solve}\left(A^{l+1}, f^{l+1}\right) \tag{2.9}$$

for the recursive application of amg_solve on level l. This allows us to write down the error propagation of coarse-grid correction on any level in the multigrid hierarchy similar to (2.8).

Proposition 2.9. *The error propagation operator $\tilde{C}^l$ of coarse-grid correction on level l is given by*

$$\tilde{C}^l = \left(I - P_{l+1}^l \left(\tilde{A}^{l+1}\right)^{-1} R_l^{l+1} A^l\right).$$

Proof. Analogously to the proof of Proposition 2.8, but using (2.9) to define the approximate coarse-grid solution. □

Remark 2.10. *In the case the coarse-grid system is inverted exactly – as in the two-grid case – we are going to denote the error propagation operator as*

$$C^l = \left(I - P_{l+1}^l \left(A^{l+1}\right)^{-1} R_l^{l+1} A^l\right),$$

in analogy to Proposition 2.8.

Corollary 2.11. *The error propagation operator $\tilde{E}_{\nu_1,\nu_2}^l$ on level l of the AMG V(ν_1, ν_2)-cycle as given in Algorithm 1 is given by*

$$\tilde{E}_{\nu_1,\nu_2}^l = S_l^{\nu_2} \left(I - P_{l+1}^l \left(\tilde{A}^{l+1}\right)^{-1} R_l^{l+1} A^l\right) S_l^{\nu_1} = S_l^{\nu_2} \tilde{C}^l S_l^{\nu_1},$$

where $\tilde{A}^{l+1}$ is an approximate inverse of A^{l+1} for $l = 0, \ldots, L-1$ given by the multigrid recursion.

Assuming that the coarse-grid system is inverted exactly we correspondingly obtain

$$E^l_{\nu_1,\nu_2} = S_l^{\nu_2}\left(I - P^l_{l+1}\left(A^{l+1}\right)^{-1} R^{l+1}_l A^l\right) S_l^{\nu_1} = S_l^{\nu_2} C^l S_l^{\nu_1},$$

Proof. Follows from Definition 2.2, Proposition 2.9 and Remark 2.10. □

2.2.2 Theory

In this subsection, we present some two- and multigrid theory for AMG in the hermitian positive definite case. It is a review of results found in [McC85, MR82, McC87, RS87, BPWX91, Bra86, FVZ05]. An interesting observation to be made in the review of algebraic multigrid theory is the fact that the necessary assumptions on convergence can be naturally separated into assumptions on the smoother and assumptions on coarse-grid correction, i.e., the interpolation used in the algebraic multigrid method.

In what follows, let $A \in \mathbb{C}^{m\times m}$ be hermitian positive definite, we then define restriction R^{l+1}_l for $l = 0, \ldots, L-1$ as the adjoint of interpolation, i.e.,

$$R^{l+1}_l = \left(P^l_{l+1}\right)^H \quad \text{and} \quad A^{l+1} = \left(P^l_{l+1}\right)^H A^l P^l_{l+1}.$$

With this choice all matrices in the grid hierarchy A^l are hermitian and positive definite, as we assume that all P^l_{l+1} have full-rank. Further assume that A has unit diagonal, i.e., $\operatorname{diag}(A) = I$.

Proposition 2.12. *With $A^l \in \mathbb{C}^{m_l\times m_l}$ and $A^{l+1} = \left(P^l_{l+1}\right)^H A^l P^l_{l+1}$ we have for all $w \in \mathbb{C}^{m_{l+1}}$*

$$\|P^l_{l+1} w\|_{A^l} = \|w\|_{A^{l+1}}.$$

Proof. By definition of the A_l-norm we obtain

$$\begin{aligned}\|P^l_{l+1} w\|^2_{A^l} &= \langle P^l_{l+1} w, P^l_{l+1} w\rangle_{A^l} \\ &= \langle \left(P^l_{l+1}\right)^H A^l P^l_{l+1} w, w\rangle_2 \\ &= \langle w, w\rangle_{A^{l+1}} = \|w\|^2_{A^{l+1}}.\end{aligned}$$

□

Proposition 2.13. *The coarse-grid correction error propagator C^l is an A^l-orthogonal projection onto $\mathcal{R}\left(I - P^l_{l+1}\right)$ and $\mathcal{R}\left(C^l\right) \perp_{A^l} \mathcal{R}\left(P^l_{l+1}\right)$.*

Proof. We have for $x, y \in \mathbb{C}^{m_l}$

$$\begin{aligned}
& \langle \left(I - C^l\right) x, C^l y \rangle_{A^l} \\
= \; & \langle P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l x, \left(I - P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l\right) y \rangle_{A_l} \\
= \; & \langle A^l P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l x, \left(I - P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l\right) y \rangle_2 \\
= \; & \langle x, \Big(A^l P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l \\
& - A^l P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l P_{l+1}^l \left(A^{l+1}\right)^{-1} \left(P_l^{l+1}\right)^H A^l\Big) y \rangle_2 \\
= \; & 0.
\end{aligned}$$

That is, we have $\left(C^l\right)^2 = C^l$ and $\mathcal{R}\left(C^l\right) \perp_{A^l} \mathcal{R}\left(P_{l+1}^l\right)$. □

In order to simplify notation, we denote the A^l orthogonal projection onto $\mathcal{R}\left(P_{l+1}^l\right)$ by $\pi_{A^l}\left(P_{l+1}^l\right)$. Thus, the coarse-grid correction error propagator can be written as

$$C^l = I - \pi_{A^l}\left(P_{l+1}^l\right).$$

With these preliminary observations in place, we state a result that motivates the use of the Galerkin formulation for the coarse-grid system. Namely, that it fulfills a variational principle, minimizing the A-norm of the error after coarse-grid correction.

Theorem 2.14. *For $e^l \in \mathbb{C}^{m_l}$ and $e^{l+1} \in \mathbb{C}^{m_{l+1}}$ we have*

$$\|C^l e^l\|_{A^l} = \min_{e^{l+1}} \|e^l - P_{l+1}^l e^{l+1}\|_{A^l}.$$

Proof. As $\mathcal{R}\left(C^l\right) \perp_{A^l} \mathcal{R}\left(P_{l+1}^l\right)$ we know that $\mathcal{R}\left(P_{l+1}^l\right) = \mathcal{R}\left(I - C^l\right)$. Hence, with the A^l-orthogonality of C^l and the decomposition of e^l into $e^l = C^l e^l + \left(I - C^l\right) e^l$ we have

$$\begin{aligned}
\min_{e^{l+1}} \|e^l - P_{l+1}^l e^{l+1}\|_{A^l}^2 &= \min_{g^l \in \mathcal{R}\left(I - C^l\right)} \|C^l e^l + \left(I - C^l\right) e^l - g^l\|_{A^l}^2 \\
&= \min_{g^l \in \mathcal{R}\left(I - C^l\right)} \|C^l e^l - g^l\|_{A^l}^2 \\
&= \min_{g^l \in \mathcal{R}\left(I - C^l\right)} \left(\|C^l e^l\|_{A^l}^2 + \|g^l\|_{A^l}^2\right) \\
&= \|C^l e^l\|_{A^l}^2.
\end{aligned}$$

□

The early multigrid theory found in [McC85, MR82, McC87, RS87] relies on two fundamental requirements split up into an assumption on interpolation and another pair of assumptions on the smoother.

Definition 2.15. *Let $e^l \in \mathbb{C}^{m_l}$. We say that $\mathcal{S}_l$ fulfills the* smoothing property *if*

$$\begin{aligned} \|S_l e^l\|_{A^l}^2 &\leq \langle A^l e^l, e^l\rangle_2 - \alpha_1 \langle A^l e^l, A^l e^l\rangle_2, \quad \text{or} && (2.10)\\ \|S_l e^l\|_{A^l}^2 &\leq \langle A^l e^l, e^l\rangle_2 - \alpha_2 \langle A^l S_l e^l, A^l e^l\rangle_2, && (2.11) \end{aligned}$$

for $\alpha_1, \alpha_2 \geq 0$.

Definition 2.16. *Let $e^l \in \mathbb{C}^{m_l}$, we say P_{l+1}^l fulfills the* approximation property *with constant $K > 0$ if*

$$\|C^l e^l\|_{A^l}^2 \leq K \langle A^l e^l, A^l e^l\rangle_2 \quad \text{for all } e^l \in \mathbb{C}^{m_l} \tag{2.12}$$

Assuming that the smoother fulfills the smoothing properties (2.10) and (2.11) and that interpolation fulfills the approximation property (2.12) we can show multigrid convergence of the algebraic multigrid V(1, 1)-cycle.

To do so we show separate results for the pre- and post-smoothing case, i.e., V(ν_1, 0)- and V(0, ν_2)-cycle, respectively (cf. [RS87]). First we prove an auxiliary result.

Proposition 2.17. *Let $e^l \in \mathbb{C}^{m_l}$, then the following relation holds between the multigrid coarse-grid correction $\tilde{C}^l$ in the multigrid V-cycle and the two-grid coarse-grid correction C^l*

$$\tilde{C}^l e^l = C^l e^l + P_{l+1}^l \left(v^{l+1} - \tilde{v}^{l+1}\right). \tag{2.13}$$

Proof. By Proposition 2.9 we have

$$\begin{aligned} \tilde{C}^l e^l &= e^l - P_{l+1}^l \tilde{v}^{l+1} \\ &= e^l - P_{l+1}^l v^{l+1} + P_{l+1}^l \left(v^{l+1} - \tilde{v}^{l+1}\right) \\ &= C^l e^l + P_{l+1}^l \left(v^{l+1} - \tilde{v}^{l+1}\right). \end{aligned}$$

□

Proposition 2.18. (Multigrid convergence of the V(0, 1)-cycle) *Assume that S_l fulfills* (2.10) *for $l = 0, \ldots, L-1$ with some $\alpha_1 > 0$ independent of l. Further assume that the interpolation operators P_{l+1}^l fulfill* (2.12) *for $l = 0, \ldots, L-1$ with $K > 0$ independent of l. Then we have for the error propagator of the V(0, 1)-cycle $\tilde{E}_{0,1}^0$ and $e \in \mathbb{C}^m$*

$$\|\tilde{E}_{0,1}^0 e\|_{A^0}^2 \leq \left(1 - \frac{\alpha_1}{K}\right) \|e\|_{A^0}^2.$$

Hence, the algebraic multigrid V(0, 1)-cycle converges and the convergence factor is bounded by $\sqrt{1 - \frac{\alpha_1}{K}}$ in the A-norm.

Proof. First, we show that we have for all $e^l \in \mathbb{C}^{m_l}, l = 0, \ldots, L-1$

$$\|S_l e^l\|_{A^l}^2 \leq \|e^l\|_{A^l}^2 - \frac{\alpha_1}{K}\|C^l e^l\|_{A^l}^2. \tag{2.14}$$

This estimate follows directly from (2.10) and (2.12)

$$\begin{aligned}\|S_l e^l\|_{A^l}^2 &\overset{(2.10)}{\leq} \langle A^l e^l, e^l\rangle_2 - \alpha_1 \langle A^l e^l, A^l e^l\rangle_2, \\ &\overset{(2.12)}{\leq} \langle A^l e^l, e^l\rangle_2 - \frac{\alpha_1}{K}\|C^l e^l\|_{A^l}^2.\end{aligned}$$

The remainder of the proof is done recursively. Hence we consider a two-grid situation between levels l and $l+1$ for $l = 0, \ldots, L-1$ and assume that the V-cycle on the coarse-grid has a convergence factor of $0 \leq \eta_{l+1} < 1$ in the A_l-norm. With this assumption and Proposition 2.12 we have

$$\begin{aligned}\|P_{l+1}^l \left(v^{l+1} - \tilde{v}^{l+1}\right)\|_{A^l} &= \|v^{l+1} - \tilde{v}^{l+1}\|_{A^{l+1}} \\ &\leq \eta_{l+1}\|v^{l+1}\|_{A^{l+1}} = \eta_{l+1}\|P_{l+1}^l v^{l+1}\|_{A^l}.\end{aligned}$$

As $\mathcal{R}\left(C^l\right) \perp_{A^l} \mathcal{R}\left(P_{l+1}^l\right)$, combining the above equality and (2.13) yields for $e^l \in \mathbb{C}^{m_l}$

$$\begin{aligned}\|\tilde{C}^l e^l\|_{A^l}^2 &= \|C^l e^l\|_{A^l}^2 + \|P_{l+1}^l \left(v^{l+1} - \tilde{v}^{l+1}\right)\|_{A^l}^2 \\ &\leq \|C^l e^l\|_{A^l}^2 + \eta_{l+1}^2\|P_{l+1}^l v^{l+1}\|_{A^l}^2 \\ &= \|C^l e^l\|_{A^l}^2 + \eta_{l+1}^2\left(\|e^l\|_{A^l}^2 - \|C^l e^l\|_{A^l}^2\right).\end{aligned} \tag{2.15}$$

Using (2.14) and the facts that $C^l\tilde{C}^l = C^l$ and $\|C^l\|_{A^l} = 1$ the following estimate shows $\eta_l \leq \max\{\eta_{l+1}, 1 - \frac{\alpha_1}{K}\}$

$$\begin{aligned}\|S_l\tilde{C}^l e^l\|_{A^l}^2 &\leq \|\tilde{C}^l e^l\|_{A^l}^2 - \frac{\alpha_1}{K}\|C^l\tilde{C}^l e^l\|_{A^l}^2 \\ &= \|\tilde{C}^l e^l\|_{A^l}^2 - \frac{\alpha_1}{K}\|C^l e^l\|_{A^l}^2 \\ &\leq \left(1 - \frac{\alpha_1}{K} - \eta_{l+1}^2\right)\|C^l e^l\|_{A^l}^2 + \eta_{l+1}^2\|e^l\|_{A^l}^2 \\ &\leq \max\{\eta_{l+1}, 1 - \frac{\alpha_1}{K}\}\|e^l\|_{A^l}^2.\end{aligned}$$

Applying this result recursively with $\eta_L = 0$ yields the desired bound on $\tilde{E}_{0,1}^0$ in the A-norm. □

Assumption (2.10) on the smoother is a natural condition for multigrid-convergence for the V$(0,1)$-case, whereas assumption (2.11) can be used for the proof of convergence in the V$(1,0)$-case.

Proposition 2.19. (Multigrid convergence of the V$(1,0)$-cycle) *Assume that S_l fulfills (2.11) for $l = 0, \dots, L-1$ with some $\alpha_2 > 0$ independent of l. Further assume that the interpolation operators P_{l+1}^l fulfill (2.12) for $l = 0, \dots, L-1$ with $K > 0$ independent of l. Then we have for the multigrid error propagator $\tilde{E}_{1,0}^0$ of the $V(1,0)$-cycle and $e \in \mathbb{C}^m$*

$$\|\tilde{E}_{1,0}^0 e\|_{A^0}^2 \leq \frac{1}{1+\frac{\alpha_2}{K}} \|e\|_{A^L}^2. \tag{2.16}$$

Hence, the algebraic multigrid $V(1,0)$-cycle converges and the convergence factor is bounded by $\sqrt{\frac{1}{1+\frac{\alpha_2}{K}}}$ in the A-norm.

Proof. The proof is similar to the proof of Proposition 2.18. Again, we first provide an estimate on each level l. For all $e^l \in \mathbb{C}^{m_l}, l = 0, \dots, L-1$ we obtain

$$\|S_l e^l\|_{A^l}^2 \leq \|e^l\|_{A^l}^2 - \frac{\alpha_2}{K} \|C^l S_l e^l\|_{A^l}^2. \tag{2.17}$$

Using (2.11) and (2.12), estimate (2.17) follows directly

$$\begin{aligned} \|S_l e^l\|_{A^l}^2 &\overset{(2.11)}{\leq} \langle A^l e^l, e^l \rangle_2 - \alpha_2 \langle A^l S_l e^l, A^l S_l e^l \rangle_2 \\ &\overset{(2.12)}{\leq} \langle A^l e^l, e^l \rangle_2 - \frac{\alpha_2}{K} \|C^l S_l e^l\|_{A^l}^2. \end{aligned}$$

Replacing e^l by $S_l e^l$ in (2.15) and using (2.17) we obtain the following estimates with $\zeta = \frac{\|C^l S_l e^l\|_{A^l}}{\|S_l e^l\|_{A^l}} \in [0,1]$,

$$\|\tilde{C}^l S_l e^l\|_{A^l}^2 \leq \left(\zeta + \eta_{l+1}^2 (1-\zeta)\right) \|S^l e^l\|_{A^l}^2 \text{ and } \left(1 + \frac{\zeta \alpha_2}{K}\right) \|S^l e^l\|_{A^l}^2 \leq \|e^l\|_{A^l}^2.$$

Combining both estimates yields $\|\tilde{C}^l S_l e^l\|_{A^l} \leq \eta_l \|e^l\|_{A^l}$ with

$$\eta_l^2 = \max_{0 \leq \zeta \leq 1} \frac{\zeta + \eta_{l+1}^2 (1-\zeta)}{1 + \zeta \frac{\alpha_2}{K}} = \max\{\eta_{l+1}^2, \frac{1}{1+\frac{\alpha_2}{K}}\}.$$

Again recursive application of this estimate with $\eta_L = 0$ yields the desired bound for $\tilde{E}_{1,0}^0$. □

It is possible to generalize these results for the V$(\nu_1, 0)$- and V$(0, \nu_2)$-cycle case with $\nu_1, \nu_2 > 1$, but we omit this discussion here as it yields only minor additional insight. In practice both pre- and post-smoothing are used. By combining Propositions 2.18 and 2.19 we can formulate the following estimate on $\tilde{E}_{1,1}^0$, i.e., for the V$(1,1)$-cycle.

Theorem 2.20. (Multigrid convergence of the V(1, 1)-cycle) *Assume that* S_l *fulfills* (2.10) *and* (2.11) *for* $l = 0, \ldots, L-1$ *with* $\alpha_1, \alpha_2 > 0$ *independent of* l*. Further assume that the interpolation operators* P_{l+1}^l *fulfill* (2.12) *for* $l = 0, \ldots, L-1$ *with* $K > 0$ *independent of* l*. Then we have for the finest-grid error-propagator* $\tilde{E}_{1,1}^0$ *of the* $V(1,1)$*-cycle*

$$\|\tilde{E}_{1,1}^0 e\|_{A^0}^2 \leq \frac{1 - \frac{\alpha_1}{K}}{1 + \frac{\alpha_2}{K}} \|e\|_{A^L}^2.$$

Hence, the algebraic multigrid $V(1,1)$*-cycle converges and the convergence factor in the A-norm is bounded by* $\sqrt{\frac{1-\frac{\alpha_1}{K}}{1+\frac{\alpha_2}{K}}}$.

Proof. Follows from Propositions 2.18 and 2.19, by applying (2.16) in a similar way to the proof of Proposition 2.19 before taking the maximum in the proof of Proposition 2.18. □

Quoting the interpretation in [McC87] using the property (2.14) (and similarly (2.17)) that describes the interplay of coarse-grid correction C^l and smoother S_l, we can extract an understanding of this interplay directly in a purely algebraic fashion. Error components e^l that cannot be efficiently reduced by $\tilde{C}^l$, i.e., $\|\tilde{C}^l e^l\|_{A^l} \approx \|e^l\|_{A^l}$ have to be uniformly reducible by S_l. The smoother S_l is allowed to be inefficient for error components that are efficiently reduced by $\tilde{C}^l$. As these error components have to be contained in $\mathcal{R}\left(P_{l+1}^l\right)$ we can state a motivating heuristic for all later developments that concern the construction of interpolation.

Heuristic 2.21. *Error components* e^l *that are inefficiently reduced by* S_l*, i.e.,* $\|S_l e^l\|_{A^l} \approx \|e^l\|_{A^l}$ *must have an accurate approximation in* $\mathcal{R}\left(P_{l+1}^l\right)$.

There is another way to prove multigrid convergence with slightly changed assumptions on the coarse-grid correction and smoother. This theory can, in principle, be found in [BPWX91]. Its adaption to the algebraic multigrid framework, especially the aggregation-based approach can be found in [VBM01]. The main ingredient in this theory is a weakened form of the approximation property (2.12) that is a crucial condition for two-grid convergence.

Definition 2.22. *Given interpolation operators* P_{l+1}^l*, we define the composite interpolation operators*

$$P_l : \mathbb{C}^{m_l} \to \mathbb{C}^m, l = 1, \ldots, L$$
$$P_l = P_1^0 P_2^1 \ldots P_l^{l-1}.$$

Apparently, we have $\mathcal{R}(P_{l+1}) \subset \mathcal{R}(P_l), l = 1, \ldots, L-1$.

Definition 2.23. *We say that the interpolation P_l fulfills a* weak approximation property *if there exists a projection $\pi(P_l)$ onto $\mathcal{R}(P_l)$ and $K > 0$, independent of l, such that*

$$\| (I - \pi(P_l)) e\|_2 \leq K \|e\|_A \tag{2.18}$$

is fulfilled for all $e \in \mathbb{C}^m$.

Lemma 2.24. *With*

$$\tilde{Q}_l = \left((P_l)^H P_l \right)^{-1} (P_l)^H$$

the operator $P_l \tilde{Q}_l$ is an orthogonal projection onto $\mathcal{R}(P_l)$, and we have

$$\| \left(P_l \tilde{Q}_l - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2 \leq \| (I - \pi(P_l)) e\|_2^2.$$

Proof. As $\mathcal{R}\left(P_l \tilde{Q}_l\right) \perp \mathcal{R}\left(I - P_l \tilde{Q}_l\right)$ and $\mathcal{R}(P_{l+1}) \subset \mathcal{R}(P_l)$ we have

$$\begin{aligned} \| \left(I - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2 &= \| \left(I - P_l \tilde{Q}_l + P_l \tilde{Q}_l - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2 \\ &= \| \left(I - P_l \tilde{Q}_l \right) e\|_2^2 + \| \left(P_l \tilde{Q}_l - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2 \\ &\geq \| \left(P_l \tilde{Q}_l - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2. \end{aligned}$$

Now by the minimizing property of the orthogonal projection in $\| \cdot \|_2$ we have

$$\| \left(P_l \tilde{Q}_l - P_{l+1} \tilde{Q}_{l+1} \right) e\|_2^2 \leq \| (I - \pi(P_l)) e\|_2^2.$$

□

Theorem 2.25. *Assume that there are linear mappings $Q_l : \mathbb{C}^m \to \mathcal{R}(P_l)$ and constants c_1, c_2 such that for all $e \in \mathbb{C}^m$ and every level $l = 1, \ldots, L$*

$$\|Q_l e\|_A \leq c_1 \|e\|_A$$

and we have

$$\| (Q_l - Q_{l+1}) e\|_2 \leq \frac{c_2}{\sqrt{\rho(A_l)}} \|e\|_A.$$

Furthermore, assume that for every l the smoother S_l is hermitian positive definite satisfying

$$\lambda_{min}(I - S_l A_l) \geq 0 \quad \textit{and} \quad \lambda_{min}(S_l) \geq \frac{1}{c_R^2 \rho(A_l)} \tag{2.19}$$

with a constant $c_R > 0$ independent of the level. Then we have for the multigrid error propagator

$$\|Ee\|_A \leq \left(1 - \frac{1}{c_0(L)}\right) \|e\|_A$$

for all $e \in \mathbb{C}^m$ and $c_0(L) = (1 + c_1 + c_2 c_R)^2 (L - 1)$.

Proof. For the sake of brevity we do not present the proof of this theorem. A detailed proof can can be found in [BPWX91, Theorem 1]. □

Corollary 2.26. *Assume that the interpolations P_l are chosen in such a way that for all $l = 1, \ldots, L$ we have*

$$\|P_l \left((P_l)^H P_l\right)^{-1} (P_l)^H e\|_A \leq c_1 \|e\|_A \tag{2.20}$$

and further

$$\|(I - \pi(P_l)) e\|_2 \leq \frac{c_2}{\sqrt{\rho(A_l)}} \|e\|_A \tag{2.21}$$

for all $e \in \mathbb{C}^m$ and c_1, c_2 independent of l. Further assume that for every level the smoother S_l fulfills (2.19)*. Then we have for the multigrid error propagation operator*

$$\|Ee\|_A \leq \left(1 - \frac{1}{c_0(L)}\right) \|e\|_A$$

for all $e \in \mathbb{C}^m$ and $c_0(L) = (1 + c_1 + c_2 c_R)^2 (L - 1)$.

Proof. The proof follows directly from Lemma 2.24 and Theorem 2.25. Note, that since $P_l \left((P_l)^H P_l\right)^{-1} (P_l)^H$ are orthogonal projections in $\langle ., . \rangle_2$ and the norms associated with $\langle ., . \rangle_2$ and $\langle ., . \rangle_{A_l}$ are equivalent, i.e., there exist $\alpha, \beta \in \mathbb{R}^+$ with $\alpha \|v\|_2 \leq \|v\|_{A_l} \leq \beta \|v\|_2, v \in \mathbb{C}^{m_l}$. Combining the equivalence of norms with the observation

$$\|P_l \left((P_l)^H P_l\right)^{-1} (P_l)^H\|_2 \leq 1,$$

we can always obtain a constant c_1 independent of l in (2.20). □

A completely different theoretical approach for convergence in an algebraic multigrid framework that does not rely on separate assumptions on smoothing and coarse-grid correction and yields an identity for the two-grid case was proposed in [FVZ05] and analyzed in [FV04].

Given a smoother S in a two-grid setting, we define its symmetrized version $\tilde{S}$ by

$$\tilde{S} = S^H \left(S^H + S - A\right)^{-1} S. \tag{2.22}$$

According to [FV04, Theorem 4.3] we have an identity for the two-grid V$(1,1)$-cycle error propagator with smoother S that we present in the following theorem.

Theorem 2.27. *Let E_{tg} be defined by*

$$E_{tg} = \left(I - S^{-H}A\right)\left(I - \pi_A\left(P\right)\right)\left(I - S^{-1}A\right),$$

with $\|S\|_A < 1$ and interpolation $P : \mathbb{C}^{m_c} \to \mathbb{C}^m$ of full rank. Then we have

$$\|E_{tg}\|_A = \left(1 - \frac{1}{K}\right), \quad \textit{where } K = \sup_e \frac{\|\left(I - \pi_{\tilde{S}}\left(P\right)\right)e\|_{\tilde{S}}}{\|e\|_A}.$$

Proof. For the sake of brevity we do not present the proof of this theorem. A detailed proof can can be found in [FV04, Theorem 4.3]. □

In the following sections, we introduce the specific definition of the components of an AMG algorithm. We start with various ways to define interpolation and conclude with compatible relaxation, an idea to algebraically define coarse variable sets driven by multigrid relaxation.

2.3 The AMGe Framework

In this section, we introduce a framework for the remaining review of existing algebraic multigrid approaches. We point out that all approaches reviewed in the following can be viewed as a special case of the AMGe framework introduced in [HV01]. This fact further motivate certain ideas in chapter 4 and guides most theoretical observations made in this thesis.

2.3.1 Element-based AMG (AMGe)

Element-based algebraic multigrid, as introduced in [BCF+00], aims at defining accurate interpolation operators in an algebraic multigrid framework using finite element stiffness matrices of an underlying finite element discretization. It can be regarded as an algebraic extension of geometric multigrid as it assumes knowledge about the finite element discretization, but builds the multigrid hierarchy by algebraic multigrid means.

In a given finite element framework the linear operator A can be expressed as

$$A = \sum_{\alpha \in \mathcal{T}} A_\alpha$$

where A_α are finite element stiffness matrices and $\mathcal{T}$ is the set of finite elements used to discretize the problem.

As stated in heuristic 2.21 in section 2.2, interpolation in algebraic multigrid should be defined such that it complements the given smoother. For most commonly used smoothers (e.g., Richardson, ω-Jacobi, Gauss-Seidel) this is equivalent to the fact that the range of interpolation has to contain accurate approximations to eigenvectors of small eigenvalues. Indeed, for eigenvectors of small eigenvalues one finds $\|Sx\|_A \approx \|x\|_A$ for these smoothers. This observation was proved to be true for Richardson and Jacobi relaxation for constant diagonal. In case of Gauss-Seidel the situation is more complicated, but known to be true for certain classes of matrices, e.g., symmetric M-matrices (cf. [Stü99]).

In the following we assume that the coarse-grid variables $\mathcal{C}$ are given as a subset of all variables Ω, i.e., $\Omega = \mathcal{C} \cup \mathcal{F}$. Further we denote $m_c = |\mathcal{C}|$, hence interpolation is given as a linear operator $P : \mathbb{C}^{m_c} \longrightarrow \mathbb{C}^m$. By reordering the variables $\mathcal{F}$ first, we can write the linear operator A as

$$A = \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix}.$$

With this permutation and the assumption that interpolation is the identity on the coarse variable set $\mathcal{C}$, we have

$$P = \begin{pmatrix} P_{fc} \\ I \end{pmatrix}.$$

Now we define with these notations approximation measures for the defect of interpolation. These definitions resemble the approximation measures (2.18) and (2.12) introduced in section 2.2.

Definition 2.28. *Given the interpolation P and an arbitrary projection $\pi(P) : \mathbb{C}^m \to \mathbb{C}^m$ onto $\mathcal{R}(P)$, we define a weak approximation measure*

$$M_1(\pi(P), e) = \frac{\langle (I - \pi(P))\, e, (I - \pi(P))\, e \rangle_2}{\langle Ae, e \rangle_2}, \tag{2.23}$$

and an approximation measure

$$M_2(\pi(P), e) = \frac{\langle A(I - \pi(P))\, e, (I - \pi(P))\, e \rangle_2}{\langle Ae, Ae \rangle_2}. \tag{2.24}$$

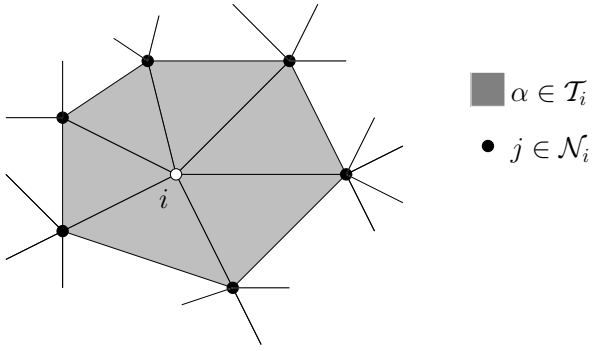


Figure 2.6: Local finite element neighborhood.

Remark 2.29. *Note that the constants K used in Definitions 2.23 and 2.16 are given as*

$$\sup_e M_1\left(\pi\left(P\right), e\right), \quad \textit{and} \quad \sup_e M_2\left(\pi\left(P\right), e\right),$$

respectively.

The main idea of element-based AMG is to localize these two measures using knowledge about the finite-element discretization for the problem at hand. Using localized measures of (2.23) and (2.24) can then be used to construct interpolation and recombined again to obtain a global measure on the defect of interpolation. Recalling that $A = \sum_{\alpha \in \mathcal{T}} A_\alpha$ is the sum of local stiffness matrices we define certain neighborhood sets in this finite-element framework. First we define the point set of an element, i.e., all grid points adjoint to element $\alpha \in \mathcal{T}$ on the given grid.

Definition 2.30. *The point set $\mathcal{J}_\alpha$ is given as*

$$\mathcal{J}_\alpha = \{j \mid e_j^T A_\alpha e_j \neq 0\},$$

where e_j is the canonical basis vector associated with the unknown j. Moreover, we define two neighborhoods of a point i, the set of neighboring elements $\mathcal{T}_i$ and the set of neighboring points $\mathcal{N}_i$ by

$$\begin{aligned} \mathcal{T}_i &= \{\alpha \in \mathcal{T} \mid e_i^T A_\alpha e_i \neq 0\}, \\ \mathcal{N}_i &= \cup_{\alpha \in \mathcal{T}_i} \mathcal{J}_\alpha. \end{aligned} \tag{2.25}$$

In Figure 2.6 we illustrate the relations between i, $\mathcal{T}_i$ and $\mathcal{N}_i$. Using this notation, we can define local versions of the linear operator.

Definition 2.31. *The local operator A_i for a grid point $i \in \Omega$ is given by*

$$A^{(i)} = \sum_{\alpha \in \mathcal{T}_i} A_\alpha.$$

Now, we are able to write down the localized versions of the weak approximation measure (2.23) and approximation measure (2.24) in a straightforward manner. We show later that if the localized measures are bounded, the global measure is bounded as well.

Definition 2.32. *Analogously to the approximation measures* (2.23) *and* (2.24) *we define for each $i \in \Omega$ with e_i denoting the i-th canonical unit vector*

$$M_1^i(\pi(P), e) = \frac{\langle e_i e_i^T (I - \pi(P)) e, e_i e_i^T (I - \pi(P)) e \rangle_2}{\langle A^{(i)} e, e \rangle_2}, \tag{2.26}$$

$$M_2^i(\pi(P), e) = \frac{\langle e_i e_i^T A (I - \pi(P)) e, e_i e_i^T (I - \pi(P)) e \rangle_2}{\langle A^{(i)} e, A^{(i)} e \rangle_2}. \tag{2.27}$$

Remark 2.33. *The definition of the localized approximation measure* (2.27) *does not match the definition in [BCF+00]. In their definition of M_2^i the factor A in the denominator cancels, as they define it with permuted factors $Ae_ie_i^T$ instead of $e_ie_i^TA$. However, in this way it is unclear why it should resemble a localized version of the approximation property. We show that by modifying the definition according to* (2.27) *we are still able to recover a global bound from the local bounds on the measures M_2^i, but we are unable to explicitly state the minimizer.*

In the case, where the coarse-grid variables are given as a subset of Ω we can choose

$$\pi(P) = Q = \begin{pmatrix} 0 & P_{fc} \\ 0 & I \end{pmatrix}.$$

With that choice it is easy to see that $M_1^i = M_2^i = 0$ for $i \in \mathcal{C}$. Further for any choice of Q the measures M_1^i and M_2^i only depend on the ith row of Q, i.e., the i-th row of P_{fc} for any $i \in \mathcal{F}$.

Theorem 2.34. *Assume that the local weak approximation measure* (2.26) *satisfies for any $i \in \mathcal{F}$*

$$M_1^i(Q, e) \leq K_1^i \quad \text{for all} \quad e \in \mathbb{C}^m. \tag{2.28}$$

Then the global weak approximation measure (2.23) *satisfies $M_1(Q, e) \leq K_1$, with*

$$K_1 = \max_{\alpha \in \mathcal{T}} \sum_{i \in \mathcal{J}_\alpha \cap \mathcal{F}} K_1^i.$$

Proof. With $I = \sum_{i\in\Omega} e_i e_i^T$ we have

$$\begin{aligned}\langle (I-Q)\,e, (I-Q)\,e\rangle_2 &= \sum_{i\in\Omega} \langle e_i e_i^T\,(I-Q)\,e, e_i e_i^T\,(I-Q)\,e\rangle_2 \\ &= \sum_{i\in\mathcal{F}} \langle e_i e_i^T\,(I-Q)\,e, e_i e_i^T\,(I-Q)\,e\rangle_2.\end{aligned}$$

Using (2.28) we obtain

$$\langle (I-Q)\,e, (I-Q)\,e\rangle_2 \le \sum_{i\in\mathcal{F}} K_1^i \langle A^{(i)} e, e\rangle_2 = \sum_{i\in\mathcal{F}} K_1^i \sum_{\alpha\in\mathcal{T}_i} \langle A_\alpha e, e\rangle_2.$$

A simple counting argument yields

$$\sum_{i\in\mathcal{F}} K_1^i \sum_{\alpha\in\mathcal{T}_i} \langle A_\alpha e, e\rangle_2 = \sum_{\alpha\in\mathcal{T}} \langle A_\alpha e, e\rangle_2 \sum_{i\in\mathcal{J}_\alpha\cap\mathcal{F}} K_1^i.$$

With this observation we finally get

$$\begin{aligned}\langle (I-Q)\,e, (I-Q)\,e\rangle_2 &\le \sum_{\alpha\in\mathcal{T}} \langle A_\alpha e, e\rangle_2 \sum_{i\in\mathcal{J}_\alpha\cap\mathcal{F}} K_1^i \\ &\le K_1 \sum_{\alpha\in\mathcal{T}} \langle A_\alpha e, e\rangle_2 = K_1 \langle Ae, e\rangle_2.\end{aligned}$$

□

In a similar way, we can bound M_2 by bounding the localized measures M_2^i.

Theorem 2.35. *Assume that the local approximation measure* (2.27) *fulfills for any* $i \in \mathcal{F}$

$$M_2^i\,(Q, e) \le K_2^i \quad \textit{for all} \quad e \in \mathbb{C}^m. \tag{2.29}$$

Then the global approximation measure (2.24) *satisfies* $M_2\,(Q, e) \le K_2$, *with*

$$K_2 = \max_{(\alpha,\beta)\in\mathcal{T}^2} \sum_{i\in\mathcal{J}_\alpha\cap\mathcal{J}_\beta\cap\mathcal{F}} K_2^i.$$

Proof. We have

$$\langle A\,(I-Q)\,e, (I-Q)\,e\rangle_2 = \sum_{i\in\Omega} \langle e_i e_i^T A\,(I-Q)\,e, e_i e_i^T\,(I-Q)\,e\rangle_2.$$

Using (2.29) we obtain

$$\begin{aligned}\langle A\,(I-Q)\,e,(I-Q)\,e\rangle_2 &\leq \sum_{i\in\mathcal{F}} K_2^i\langle A^{(i)}e,A^{(i)}e\rangle_2\\ &= \sum_{i\in\mathcal{F}} K_2^i\langle(\sum_{\alpha\in\mathcal{T}_i}A_\alpha)e,(\sum_{\beta\in\mathcal{T}_i}A_\beta)e\rangle_2\\ &= \sum_{i\in\mathcal{F}} K_2^i\sum_{\alpha\in\mathcal{T}_i}\sum_{\beta\in\mathcal{T}_i}\langle A_\alpha e,A_\beta e\rangle_2.\end{aligned}$$

A reordering of the sums then yields

$$\sum_{i\in\mathcal{F}} K_2^i\sum_{\alpha\in\mathcal{T}_i}\sum_{\beta\in\mathcal{T}_i}\langle A_\alpha e,A_\beta e\rangle_2=\sum_{\alpha\in\mathcal{T}}\sum_{\beta\in\mathcal{T}}\langle A_\alpha e,A_\beta e\rangle_2\sum_{i\in\mathcal{J}_\alpha\cap\mathcal{J}_\beta\cap\mathcal{F}}K_2^i.$$

This observation finally leads to

$$\begin{aligned}\langle A\,(I-Q)\,e,(I-Q)\,e\rangle_2 &\leq \sum_{\alpha\in\mathcal{T}}\sum_{\beta\in\mathcal{T}}\langle A_\alpha e,A_\beta e\rangle_2\sum_{i\in\mathcal{J}_\alpha\cap\mathcal{J}_\beta\cap\mathcal{F}}K_2^i\\ &\leq K_2\sum_{\alpha\in\mathcal{T}}\sum_{\beta\in\mathcal{T}}\langle A_\alpha e,A_\beta e\rangle_2=K_2\langle Ae,Ae\rangle_2.\end{aligned}$$

□

With Theorems 2.34 and 2.35 at hand the aim, when defining interpolation P, is to achieve a small local bounds on (2.26) and (2.27), respectively. However, in order to get an efficient multigrid method, we have to keep the operators in the multigrid hierarchy sparse. This is usually achieved by fixing a sparsity pattern for the interpolation operator P.

In this context, we define $\mathcal{C}_i=\mathcal{C}\cap\mathcal{N}_i$, with $\mathcal{N}_i$ defined according to (2.25). In order to keep P sparse, we assume that for each $i\in\mathcal{F}$ we only consider $j\in\mathcal{C}_i$ as interpolatory points, i.e., writing the ith row of Q as $e_i^TP=q_i$ we take

$$q_i\in\mathcal{Z}_i=\{v\in\mathbb{C}^m : v_j=0 \quad \text{for all } j\notin\mathcal{C}_i\}.$$

Next, consider the problem of finding the best bound on (2.26) and (2.27), respectively as a constrained minimization problem for $p=1,2$

$$K_p^i=\min_{q_i\in\mathcal{Z}_i}\max_{e\notin\mathcal{N}(A^{(i)})}M_p^i\,(q_i,e)\,. \tag{2.30}$$

This min-max problem is explicitly solved for $p=1$ by

$$q_i^*=-A_{cf}^{(i)}\left(A_{ff}^{(i)}\right)^{-1}e_i,$$

for $i \in \mathcal{F}$, where $A_{cf}^{(i)}, A_{ff}^{(i)}$ are blocks in the $\mathcal{F}-\mathcal{C}$ permuted representation of the local operator $A^{(i)}$ (cf. [BCF+00, Theorem 4.3]). Unfortunately, no analytic solution of (2.30) in the case of $p = 2$, in the new formulation of (2.27) is known.

2.3.2 Element-free AMGe (AMG~~e~~)

The element-free AMGe approach of defining interpolation, introduced in [HV01], can be seen as the algebraic counterpart to the semi-algebraic approach of AMGe introduced in section 2.3.1. Moreover the framework of element-free AMGe can be used to describe many, if not all, algebraic approaches to construct interpolation. In this section, we explain the basic idea of element-free AMGe interpolation. We continue to interpret several of the algebraic interpolation approaches as special cases of element-free AMGe.

As in AMGe, we assume for now that the coarse-grid variables $\mathcal{C}$ are given as a subset of the variables Ω, i.e., we have a splitting of variables $\Omega = \mathcal{C} \cup \mathcal{F}$. Recall that by using the finite-element discretization we could formulate localized versions $A^{(i)}$ of the linear operator A for each $i \in \mathcal{F}$. Writing this operator permuted according to variables in $\mathcal{F}$ and $\mathcal{C}$

$$A^{(i)} = \begin{pmatrix} A_{ff}^{(i)} & A_{fc}^{(i)} \\ A_{cf}^{(i)} & A_{cc}^{(i)} \end{pmatrix},$$

the interpolation in AMGe to point $i \in \mathcal{F}$ as the minimizer of (2.26), is then given as $e_i^T \left(-A_{ff}^{(i)}\right)^{-1} A_{fc}^{(i)}$.

However, this approach relies on the knowledge of the local stiffness-matrices. In a more general context, where no further information about the discretization is known, we obviously cannot use this approach, but we can attempt to imitate it. For this purpose, we first introduce some notation for neighborhood-sets with respect to the graph associated to a matrix introduced in Definition 2.1.

Definition 2.36. *For each subset $\mathcal{I} \in \Omega$ we define the* neighborhood $\mathcal{N}_\mathcal{I}$ *of $\mathcal{I}$ as*

$$\mathcal{N}_\mathcal{I} = \{j \in \Omega : a_{ij} \neq 0, i \in \mathcal{I}, j \notin \mathcal{I}\}.$$

Further, we define the fine- and coarse-variable neighborhoods $\chi_\mathcal{I}$ and $\mathcal{C}_\mathcal{I}$, respectively, as

$$\chi_\mathcal{I} = \mathcal{N}_\mathcal{I} \cap \mathcal{F} \quad \textit{and} \quad \mathcal{C}_\mathcal{I} = \mathcal{N}_\mathcal{I} \cap \mathcal{C}.$$

We usually assume that $\mathcal{I}$ is a connected subset of $\Omega \cap \mathcal{F}$ in G_A. One possible situation is illustrated in Figure 2.7.

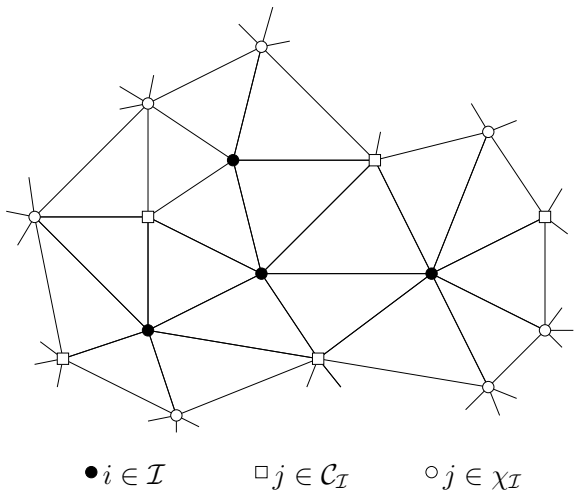


Figure 2.7: Local neighborhood of $\mathcal{I}$ in the AMG$\not{e}$ framework.

With these definitions, we can permute A for each $\mathcal{I} \subset \mathcal{F}$ in the following way

$$A = \begin{pmatrix} A_{ff} & A_{fc} & A_{f\chi} & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \begin{matrix} \}\mathcal{I} \\ \}\mathcal{C}_\mathcal{I} \\ \}\chi_\mathcal{I} \\ \}\text{everything else} \end{matrix} . \tag{2.31}$$

The aim of element-free AMGe is to construct a neighborhood matrix $\hat{A}_\mathcal{I}$ that only connects $\mathcal{I}$ with its interpolatory points $\mathcal{C}_\mathcal{I}$, similar to the localized versions of the linear operator in AMGe. Interpolation is then defined by $e_i^T \left(-\hat{A}_{ff}^{-1}\hat{A}_{fc}\right)$ for $i \in \mathcal{I}$.

In order to define $\hat{A}$ we first define a linear extension map that prolongates the "interior" of $\mathcal{I} \cup \mathcal{C}_\mathcal{I}$ to the "boundary" $\chi_\mathcal{I}$.

Definition 2.37. *Let $m_f = |\mathcal{I}|, m_c = |\mathcal{C}_\mathcal{I}|$ and $m_\chi = |\chi_\mathcal{I}|$. The extension map E is then defined by*

$$\begin{aligned} E : \mathbb{C}^{n_f+n_c} &\longrightarrow \mathbb{C}^{n_f+n_c+n_\chi} \\ \begin{pmatrix} v_f \\ v_c \end{pmatrix} &\longmapsto \begin{pmatrix} v_f \\ v_c \\ v_\chi \end{pmatrix} = \begin{pmatrix} v_f \\ v_c \\ E_{\chi f}v_f + E_{\chi c}v_c \end{pmatrix}, \end{aligned}$$

with $E_{\chi f} \in \mathbb{C}^{m_\chi \times m_f}, E_{\chi c} \in \mathbb{C}^{m_\chi \times m_c}$. That is, we can write E as

$$E = \begin{pmatrix} I & 0 \\ 0 & I \\ E_{\chi f} & E_{\chi c} \end{pmatrix}.$$

We then construct the neighborhood matrix from the first rows of (2.31) by

$$\begin{pmatrix} \hat{A}_{ff} & \hat{A}_{fc} \end{pmatrix} = \begin{pmatrix} A_{ff} & A_{fc} & A_{f\chi} \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \\ E_{\chi f} & E_{\chi c} \end{pmatrix}. \tag{2.32}$$

Hence, $\hat{A}_{ff} = A_{ff} + A_{f\chi}E_{\chi f}$ and $\hat{A}_{fc} = A_{fc} + A_{f\chi}E_{\chi c}$.

With this we define interpolation to $\mathcal{I}$ as the harmonic extension from $\mathcal{C}_\mathcal{I} \cup \chi_\mathcal{I}$ to $\mathcal{I}$ given by

$$v_f = -A_{ff}^{-1} \left(A_{fc} v_c + A_{f\chi} v_\chi \right), \tag{2.33}$$

which can be reduced using the extension map and definition of v_χ to

$$\begin{aligned} v_f &= -A_{ff}^{-1} \left(A_{fc} v_c + A_{f\chi} v_\chi \right) \\ &= -A_{ff}^{-1} \left(A_{fc} v_c + A_{f\chi} \left(E_{\chi f} v_f + E_{\chi c} v_c \right) \right) \\ &= -\left(A_{ff} + A_{f\chi} E_{\chi f} \right)^{-1} \left(A_{fc} + A_{f\chi} E_{\chi c} \right) v_c \\ &= -\hat{A}_{ff}^{-1} \hat{A}_{fc} v_c. \end{aligned}$$

The question arises whether $\hat{A}_{ff}$ is invertible. Clearly, if $E_{\chi f} = 0$ the matrix $\hat{A}_{ff}$ is invertible, since A_{ff} is invertible as a principle sub-matrix of A. Next, we show that we can reformulate any extension map E such that $E_{\chi f} = 0$.

Proposition 2.38. *Let E be given as in Definition 2.37 with $E_{\chi f} \neq 0$ and define $\hat{A}_{ff}, \hat{A}_{fc}$ according to* (2.32)*. Using $\tilde{E} = \left(0, \tilde{E}_{\chi c}\right)$, with $\tilde{E}_{\chi c} = E_{\chi f}\left(-\hat{A}_{ff}^{-1}\hat{A}_{fc}\right) + E_{\chi c}$ instead of E yields the same interpolation.*

Proof. By the definition of interpolation (2.33) using $\tilde{E}$ we have

$$\begin{aligned} -A_{ff}^{-1}\left(A_{fc} + A_{f\chi}\tilde{E}_{\chi c}\right) &= -A_{ff}^{-1}\left(A_{fc} + A_{f\chi}\left(E_{\chi f}\left(-\hat{A}_{ff}^{-1}\hat{A}_{fc}\right) + E_{\chi c}\right)\right) \\ &= -A_{ff}^{-1}\left(A_{fc} + A_{f\chi}E_{\chi c} - A_{f\chi}E_{\chi f}\hat{A}_{ff}^{-1}\hat{A}_{fc}\right) \\ &= -A_{ff}^{-1}\left(\hat{A}_{ff} - A_{f\chi}E_{\chi f}\right)\hat{A}_{ff}^{-1}\hat{A}_{fc} \\ &= -A_{ff}^{-1}A_{ff}\hat{A}_{ff}^{-1}\hat{A}_{fc} \\ &= -\hat{A}_{ff}^{-1}\hat{A}_{fc}. \end{aligned}$$

□

This observation concludes the introduction of the AMG$\not e$ framework. In the following we show that many approaches in algebraic multigrid can be written and analyzed the AMG$\not e$ framework.

2.4 Classical AMG

As stated before, AMG was first mentioned in the early 1980s in [Bra86, BMR84] and further analyzed and developed in [McC85, RS87]. Classical AMG can be seen as a first approach to define multigrid hierarchies without explicit assumptions on the underlying differential equation and knowledge of the discretization.

The focus of this subsection is to introduce interpolation in classical AMG and show that it can be viewed as a special case of the AMG$\not{e}$ framework. Note that this connection has been first mentioned in the original introduction of AMG$\not{e}$ in [HV01]. Throughout this section we assume that a splitting of variables $\Omega = \mathcal{C} \cup \mathcal{F}$ is given. In classical AMG interpolation is defined for each $i \in \mathcal{F}$ independently. Taking $\mathcal{I} = \{i\}$ in Definition 2.36, we have the neighborhood sets $\mathcal{C}_i \subset \mathcal{C}$ and $\chi_i \subset \mathcal{F}$. The definition of interpolation is then equivalent to finding interpolation weights ω_{ij} with

$$e_i = \sum_{j \in \mathcal{C}_i} \omega_{ij} e_j.$$

This interpolation should be accurate for algebraically smooth error $e \in \mathbb{C}^m$, i.e., error that is slowly reduced by the given smoother.

Under the assumption that we have $Ae \approx 0$ for algebraically smooth error we have

$$0 \approx a_{ii} e_i + \sum_{j \in \mathcal{C}_i} a_{ij} e_j + \sum_{j \in \chi_i} a_{ij} e_j. \tag{2.34}$$

It is clear that in order to define interpolatory weights $\omega_{ij}, j \in \mathcal{C}_i$ we have to eliminate the last term. The idea in classical AMG for this is to distribute the "weight" of the connection between $j \in \chi_i$ and i to the interpolatory points $k \in \mathcal{C}_i$.

If we distribute for each $k \in \chi_i$ its contribution to (2.34) by the formula

$$e_k = \frac{1}{\sum_{l \in \mathcal{C}_i} a_{kl}} \sum_{j \in \mathcal{C}_i} a_{kj} e_j, \tag{2.35}$$

we obtain the interpolation formula for $i \in \mathcal{F}$,

$$e_i = \frac{1}{a_{ii}} \sum_{j \in \mathcal{C}_i} \left(a_{ij} + \sum_{k \in \chi_i} \frac{a_{ik} a_{kj}}{\sum_{l \in \mathcal{C}_i} a_{kl}} \right) e_j = \sum_{j \in \mathcal{C}_i} \omega_{ij} e_j. \tag{2.36}$$

That is, the interpolation weights ω_{ij} are given by

$$\omega_{ij} = \frac{1}{a_{ii}} \left(a_{ij} + \sum_{k \in \chi_i} \frac{a_{ik} a_{kj}}{\sum_{l \in \mathcal{C}_i} a_{kl}} \right).$$

This interpolation is known as "standard interpolation" [RS87].

Clearly (2.35) defines an extension map from $\mathcal{C}_i$ to χ_i. The entries of the extension map E of Definition 2.37 are here given as

$$E_{kj} = \frac{a_{jk}}{\sum_{l \in \mathcal{C}_i} a_{kl}}, \quad k \in \chi_i, j \in \mathcal{C}_i,$$

and we have $E_{\chi f} = 0$ for this interpolation.

A little algebra shows that with $\hat{A}_{ff} = a_{ii}$ and $\hat{A}_{fc} = A_{ic} + A_{i\chi}E_{\chi c}$ we recover the interpolation weights p_i of (2.36) using

$$p_i = e_i^T \left(-\hat{A}_{ff}^{-1} \hat{A}_{fc} \right)$$

from the element-free approach.

To summarize, as interpolation in classical AMG is based on (2.34) and the removal of the last sum in this expression, it is possible to write this approach as an extension from $\{i\} \cup \mathcal{C}_i$ to χ_i in the AMG$\not\in$ sense.

2.5 Reduction-based AMG

We again assume that a splitting of variables $\Omega = \mathcal{C} \cup \mathcal{F}$ is given, and that we have permuted the linear operator A,

$$A = \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix}. \tag{2.37}$$

In the extreme case of $\mathcal{I} = \mathcal{F}$ we have $\chi_{\mathcal{I}} = \emptyset$ and consequently $E = I$ such that the AMG$\not\in$ interpolation is given by

$$P = \begin{pmatrix} -A_{ff}^{-1} A_{fc} \\ I \end{pmatrix}. \tag{2.38}$$

In the reduction-based framework this interpolation is derived in a different way. Based on (2.37) we write

$$A = \begin{pmatrix} I & 0 \\ A_{fc}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & 0 \\ 0 & A_{cc} - A_{cf}A_{ff}^{-1}A_{fc} \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix}.$$

It is well known that this decomposition can be interpreted as a partial Gaussian Elimination.

Definition 2.39. *The* Schur complement $S_{cc}(A)$ *of A with respect to the A_{ff} block that reduces the system of equations to the $\mathcal{C}$-variables is defined by*

$$S_{cc}(A) = A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}. \tag{2.39}$$

A little algebra shows that for the inverse of A we have

$$A^{-1} = \begin{pmatrix} I & -A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{ff}^{-1} & 0 \\ 0 & (S_{cc}(A))^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{fc}A_{ff}^{-1} & I \end{pmatrix}.$$

Exploiting this decomposition, the solution of $Au = f$ can be computed in three successive steps by

1. $r_c = \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} f_f \\ f_c \end{pmatrix}$
2. $u_c = (S_{cc}(A))^{-1} r_c$
3. $u = \begin{pmatrix} A_{ff}^{-1} & 0 \\ 0 & 0 \end{pmatrix} f + \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix} u_c$

This procedure looks very similar to the generic AMG Algorithm 1 in a 2-level setting defining interpolation as in (2.38). Note that with this we have

$$\begin{aligned} P^H AP &= \begin{pmatrix} -A_{fc}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix} \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix} \\ &= \begin{pmatrix} -A_{fc}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} 0 \\ A_{cc} - A_{cf}A_{ff}^{-1}A_{fc} \end{pmatrix} \\ &= A_{cc} - A_{cf}A_{ff}^{-1}A_{fc} = S_{cc}(A). \end{aligned}$$

Hence, the direct solution method described above can be viewed as a 2-level algebraic multigrid method with Galerkin coarse-grid formulation if we solve the $\mathcal{F}$-equation exactly in the smoothing process. As it yields the exact solution to the linear system of equations $Au = f$ the interpolation P of (2.38) is often referred to as the "ideal interpolation".

Using the ideal interpolation is prohibited in general due to the lack of sparsity of A_{ff}^{-1} that leads to an enormous amount of fill-in in the Schur complement $S_{cc}(A)$. Instead, we replace A_{ff}^{-1} by an appropriate approximation that aims for spectral equivalence of the Galerkin coarse-grid operator and the Schur complement, i.e., we want to define an approximation $\tilde{A}_{ff}^{-1}$ to A_{ff}^{-1} such that

$$\gamma P^H AP \leq S_{cc}(A) \leq \Gamma P^H AP, \ \gamma, \Gamma \in \mathbb{R},$$

with

$$P = \begin{pmatrix} -\tilde{A}_{ff}^{-1}A_{fc} \\ I \end{pmatrix}.$$

In the multigrid literature, the various multilevel iterations whose design are based on such an approximation, are typically referred to as reduction-based

AMG methods (AMGr), following [RTW83], because of their close relation to total-reduction approaches. A study of various approaches can be found in [Kah06].

The reduction-based approach that defines interpolation via the spectral equivalence of the coarse-grid system to the Schur complement, i.e., using global information, is incompatible to the AMG$\not e$ mindset, where interpolation is built by exploiting local information only. The reduction-based approach however, can be related to other multilevel decomposition schemes that are based on the approximation of the Schur complement that arises in partial Gaussian Elimination. If we replace the exact solve of the $\mathcal{F}$-equations by an appropriate pre- and post-smoothing iteration given by its error propagation

$$e = \left(I - FM_f^{-1}F^H A\right) e, \quad F = \begin{pmatrix} I \\ 0 \end{pmatrix},$$

we can write down the two-grid error propagator of the V(1, 1)-cycle as

$$E = \left(I - FM_f^{-1}F^H A\right)\left(I - P\left(P^H AP\right)P^H A\right)\left(I - FM_f^{-1}F^H A\right).$$

In [MMM06] a theory is presented for the particular choice of diagonal $\tilde{A}_{ff} = D$, replacing the exact solve of the $\mathcal{F}$-equations by an appropriate pre- and post-smoothing that also relies on the same D used to define interpolation. With this particular choice we can give the following bound on the convergence rate of the two-grid V(0, 1) cycle, i.e., a bound on $\|E\|_A$.

Theorem 2.40. *Given a diagonal $D \in \mathbb{R}^{m\times m}$, define interpolation P by*

$$P_D = \begin{pmatrix} -D^{-1}A_{fc} \\ I \end{pmatrix}$$

and a $\mathcal{F}$-smoother M by

$$M = \omega \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.40}$$

Under the assumption that

$$D \leq A_{ff} \leq (1+\epsilon)D, \tag{2.41}$$

where $\epsilon > 0$ is of moderate size and

$$0 \leq A_D = \begin{pmatrix} D & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix} \tag{2.42}$$

and assuming further that

$$\|I - M_{ff}A_{ff}\|_{A_{ff}} < 1,$$

the two-grid $V(0,1)$-cycle is convergent and the following estimate holds for its error propagator E if ω in (2.40) is chosen as $\omega = \frac{2}{2+\epsilon}$

$$\|E\|_A^2 \leq \frac{\epsilon}{1+\epsilon}\left(1 + \left(\frac{\sqrt{\epsilon}}{2+\epsilon}\right)^2\right).$$

Proof. For the sake of brevity we do not present the proof here and refer to [MMM06]. □

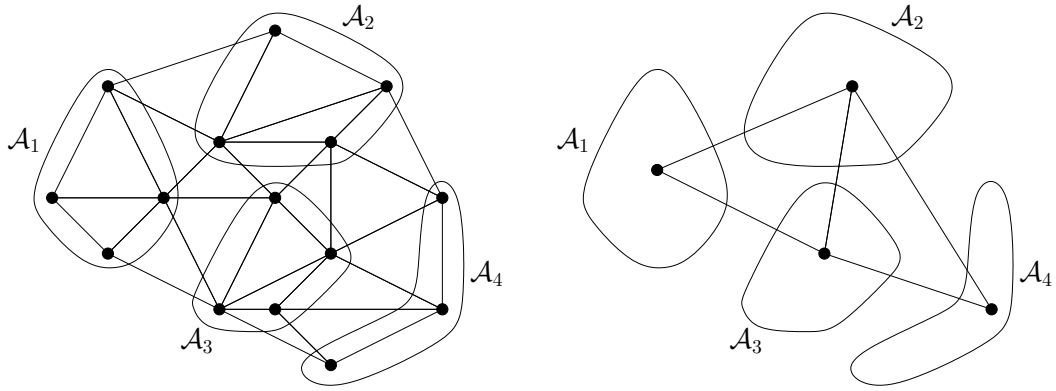
This result also holds with an arbitrary number of pre- or post-relaxation steps. The spectral equivalence relation (2.41) can be viewed as a smoothing property of D with respect to the set of fine variables $\mathcal{F}$. Compatible relaxation, introduced and discussed in [Bra00, Bra05, Liv04, FVZ05, BF09] and described in section 2.7, or the method of greedy partitioning, as presented in [MS07], generate splittings where the set of fine variables, $\mathcal{F}$, yields a well-conditioned block A_{ff}. Hence, it is possible to find such a diagonal approximation D to A. Thus, relation (2.41) states that M in (2.40) fulfills $\|I - M_{ff}A_{ff}\|_{A_{ff}} < 1$. Relation (2.42), on the other hand, can be interpreted as a requirement on the interpolation operator and, hence, the coarse-grid operator.

In section 4.1, we generalize the adaptive approach of reduction-based AMG, as introduced in [MMM06], and state a two-grid analysis that generalizes Theorem 2.40.

2.6 Aggregation-based AMG

In our discussion on classical algebraic multigrid approach to define interpolation and its variants reviewed so far, we assumed a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ to be given. Aggregation-based algebraic multigrid is the canonical approach for construction of interpolation not based on this assumption and is given in order to complete the survey of existing approaches to define interpolation in algebraic multigrid methods.

Again, we start off by interpreting the sparsity structure of the linear operator A as a graph according to Definition 2.1. However, instead of defining a set of coarse variables as a subset of this grid, aggregation-based approaches define coarse variables as overlapping or non-overlapping aggregates of points. Figure 2.8 illustrates this idea.



(a) Fine-grid (b) Coarse-grid

Figure 2.8: Aggregation-based coarsening. Aggregation of fine-grid shown in 2.8(a), associated coarse-grid shown in 2.8(b).

First approaches for aggregation-based multigrid were introduced in the numerical simulation of Markov chains (cf. the survey in [Sch91]) and later generalized by Braess in [Bra95] and Vanek in [Van95]. Further investigations and the introduction of smoothed aggregation (SA) and adaptive smoothed aggregation (αSA) followed (cf. [VMB96, VBM01, BFM$^+$05]). In this section, we review the basic concepts of SA and interpret the approach in the generalized framework we outlined in section 2.3.

Consider the graph G_A corresponding to A according to Definition 2.1.

Definition 2.41. *An* aggregation *of $G_A = (V_A, E_A)$ is a decomposition of V_A, into sets $\mathcal{A}_l,\ l = 1, \ldots, m_c$ with $V_A = \cup_l \mathcal{A}_l$.*

Given an aggregation of G_A we identify each aggregate $\mathcal{A}_l$ with a coarse-grid variable u^c_l. Hence the canonical restriction operator R with $u^c = Ru$ has the block-row structure where row i has entries at $\mathcal{A}_i$. Assuming an aggregation without overlap we can easily choose each row of R to be of norm 1. In a natural way, interpolation can then be defined such that $P = R^H$ and fulfills $P^H P = I$. This property of P is favorable when recalling (2.21) in the regularity-free multigrid convergence theory presented in Theorem 2.25.

Lemma 2.42. *Assuming that aggregation is built by disjoint sets $\mathcal{A}_i$ and $P = R^H$, the columns of P are an orthogonal basis of $\mathcal{R}(P)$.*

Proof. Since $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset,\ i \neq j$ we have $\langle P_{i,\cdot}, P_{j,\cdot} \rangle_2 = 0$. □

This approach is equivalent to the early developments of aggregation-based algebraic multigrid in [Bra95, VMB96]. Assuming that algebraically

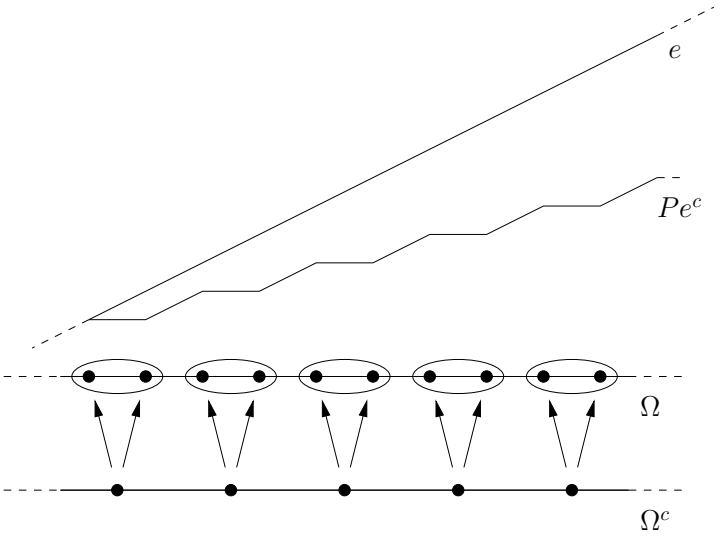


Figure 2.9: Approximation of linear vectors by piece-wise constant interpolation in Galerkin based multigrid.

smooth error, i.e., error that cannot be reduced efficiently by relaxation, is also geometrically smooth, early aggregation-based interpolation operators were simple scaled partitions of the constant vector $\begin{pmatrix} 1, & \ldots, & 1 \end{pmatrix}^T$ over the aggregates. That is, for a non-overlapping aggregation, interpolation is piece-wise constant.

However, as illustrated in Figure 2.9 (cf. [TOS01, Stü99]) the coarse-grid correction $P\left(P^H A P\right)^{-1} P^H A$ that is the optimal correction with respect to the energy norm does not yield a good approximation in euclidean norm. For a linear error the piece-wise constant correction fits the slope rather than the size of the error. This observation led to two completely different ideas to improve the aggregation-based approach. First, in [MV92, Bla88] the idea of over-weighted coarse-grid correction that minimizes

$$\|\tilde{M}\left(I - \omega\pi_A\left(P\right)\right)e\|_A$$

was introduced, including the post-smoothing action of $\tilde{M}$, rather than minimizing

$$\|\left(I - \omega\pi_A\left(P\right)\right)e\|_A.$$

It was observed in Theorem 2.14 that the latter leads to the choice $\omega = 1$. This analysis takes into account that post-smoothing is able to remove interpolation errors introduced by over-weighting the coarse-grid correction. In section 4.3, we present a similar analysis and show that the idea of optimally

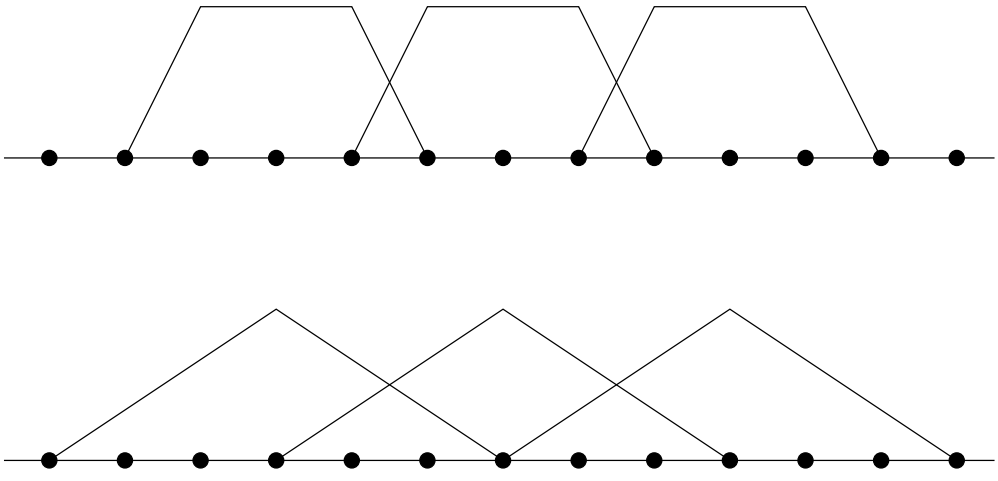


Figure 2.10: Piece-wise constant basis vectors and their smoothed counterparts.

weighted coarse-grid over-correction might be beneficial for any given algebraic multigrid interpolation, not only restricted to the aggregation-based approach. In general, the idea to build P satisfying $P^H P = I$ and having a sparsity pattern that is limited to the aggregation can be interpreted as a description of $\mathcal{R}(P)$ in terms of the columns of P. The columns of P, further on denoted by $\varphi_1, \ldots, \varphi_{m_c}$ are local (i.e., each vector φ_i has limited support on $\mathcal{A}_i$) thus form an orthonormal basis of $\mathcal{R}(P)$. Hence, the aggregation-based framework of defining interpolation can be seen as the definition of appropriate local basis functions that span the subspace of algebraically smooth functions. This interpretation motivated the second idea to overcome the shortcomings of aggregation-based AMG. Introduced in [VMB96], the use of a smoother S to improve the set of basis vectors $\varphi_1, \ldots, \varphi_{m_c}$ by creating a set of smoothed basis functions and introducing overlap of the aggregates in a natural way overcame the problems seen in the aggregation-based approach. Figure 2.10 demonstrates that one step of ω-Jacobi applied to piece-wise constant basis vectors produces basis vectors capable of representing linear functions exactly as well. However, due to this process the nice property of $P^H P = I$ is lost.

Definition 2.43. *Given a disjoint aggregation of* Ω *and a canonical interpolation* P *as well as a smoother* $S : \mathbb{C}^m \to \mathbb{C}^m$ *we define the* smoothed interpolation operator $\tilde{P}$ *by*

$$\tilde{P} = SP.$$

In [VBM01, Theorem 3.5], multigrid theory for smoothed aggregation is given that is based on the fundamental multigrid convergence proof without

regularity assumptions presented in Theorem 2.25.

Unfortunately, aggregation-based approaches do not fit in our general framework in a straight-forward manner, but there are two ways to view the aggregation-based method in this context. As already remarked, the definition of interpolation in the (smoothed) aggregation-based approach can be viewed as the definition of an appropriate set of locally confined basis vectors. In a finite-element framework, this coincides with the definition of interpolation in element-based algebraic multigrid, introduced in section 2.3.1, where one can define interpolation by computing local basis vectors based on aggregated element-stiffness matrices. On the other hand, aggregation-based approaches can be classified in the general framework via the following observation. For any full-rank interpolation operator $P \in \mathbb{C}^{m \times m_c}$ there exists a permutation

$$P = \begin{pmatrix} P_{fc} \\ P_{cc} \end{pmatrix},$$

with non-singular P_{cc}. The main focus in the analysis of interpolation operators in algebraic multigrid is $\mathcal{R}(P)$, so that we can interpret any P as given in a $\mathcal{F}$-$\mathcal{C}$-splitting setting as

$$\mathcal{R}(P) = \text{range}(\begin{pmatrix} P_{fc} \\ P_{cc} \end{pmatrix}) = \text{range}(\begin{pmatrix} P_{fc}P_{cc}^{-1} \\ I \end{pmatrix}).$$

Note that the reverse statement is also true, any interpolation operator defined using a $\mathcal{F}$-$\mathcal{C}$-splitting can be interpreted in the aggregation-based framework, i.e., choosing appropriate basis vectors for certain local aggregates.

2.6.1 Adaptive Smoothed Aggregation

Recently, the idea of smoothed aggregation was generalized and an adaptive setup was integrated [BFM+04], in order to be able to handle more problems in this framework. The basic principle of adaptive Smoothed Aggregation multigrid is the principle "let the current method expose its own weakness".

The adaptive process starts with a canonical method $\mathcal{M}_0$, i.e., the multigrid smoother S. By iterating with this method on the homogeneous system $Ax = 0$ one finds either that the current method is fast to converge and no further work is necessary in order to get a fast solver, or an error component emerges that the current method is unable to resolve. The approach of smoothed aggregation then allows to incorporate this particular error component in the multigrid solver, guaranteeing its treatment in the coarse-grid correction by breaking it up over the aggregates and adding local columns to the aggregation interpolation operator orthogonalized to existing local basis

functions in order to guarantee $P^H P = I$ of the unsmoothed aggregation operator. Thereafter a multigrid hierarchy is generated that becomes the new current method and the adaptive process goes into the next cycle, testing the new current method and improving it in the aforementioned way if necessary. This approach of adaptivity is a very convenient way to guarantee that particular error components are handled in the coarse-grid correction. After finishing the adaptive process the method is guaranteed to converge fast. However, the sequential accumulation of vectors into P may make this method unfeasibly expensive both in setup and operator complexity in general. Heuristics and techniques to get a better control on the complexity were developed and tested in [BFM+05, BBK+07].

2.7 Compatible Relaxation – Coarsening

The main focus of this section is on the algebraic construction and quality control of coarse-grids. So far, we have discussed several variants of interpolation for algebraic multigrid assuming that a sensible definition of coarse-grid variables is available, now we consider the idea of compatible relaxation, a tool that can be used to measure the quality of and construct coarse-grids.

The idea of compatible relaxation was adapted from compatible Monte-Carlo schemes (cf. [BR99, BR01]) and introduced by Brandt in the context of algebraic multigrid methods in [Bra00]. The idea was further analyzed in [Liv04, BF09].

In algebraic multigrid, compatible relaxation is useful to gauge and control the quality of the set of coarse-variables $\mathcal{C}$ prior to building the coarse-level equations. As introduced in section 2.2, coarse-grid variables u_i^c can always be interpreted as linear combinations of the current fine grid variables u_i. That is, we can write

$$u_i^c = \sum_j \mu_{ij} u_j, \;\; i \in \Omega^c, \tag{2.43}$$

according to Definition 2.4. In doing so, we further assume that the coarse-grid variables only depend on fine-grid variables in a local neighborhood based on the graph G_A according to Definition 2.1.

Motivated by the fundamental principle of multigrid, the complementarity of coarse-grid correction and smoothing, compatible relaxation uses the current smoother to gauge the quality of the current set of coarse-grid points.

Definition 2.44. Compatible relaxation *is a relaxation scheme that keeps the coarse-variables invariant, i.e., it keeps the relaxed variables u compatible with their coarse-variable counterparts u^c as defined in* (2.43).

To measure the effectiveness of compatible relaxation, we apply it to the homogeneous system of equations

$$Au = 0. \tag{2.44}$$

The compatibility constraint on the coarse-variables in this particular case means

$$0 = u_i^c = \sum_j \mu_{ij} u_j, \ i \in \Omega^c. \tag{2.45}$$

Based on this observation the introduction to compatible relaxation given in [Liv04] considers two ways to define a compatible relaxation, but one could think of other ways to define a compatible relaxation.

Definition 2.45. *If the compatible relaxation sweep on* (2.44) *simultaneously fulfills* (2.45) *we call it a* concurrent *compatible relaxation.*

We talk about a habituated *compatible relaxation scheme if the compatibility constraint* (2.45) *has to be enforced after each fine grid relaxation sweep on* (2.44)*.*

In the case where the set of coarse-variables is a subset of all variables, i.e., we have a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$, concurrent compatible relaxation has to keep the compatibility constraint fulfilled in each iteration. This amounts to a compatible relaxation method that only changes values of variables in $\mathcal{F}$. Accordingly, such a relaxation method is often referred to as $\mathcal{F}$- relaxation.

More generally using (2.43) we can write (2.45) as

$$u^c = Ru = 0, \quad R \in \mathbb{C}^{m_c \times m},$$

where $R : \mathbb{C}^m \to \mathbb{C}^{m_c}$ is the canonical restriction to the coarse-grid variables and we assume that with interpolation given by $P : \mathbb{C}^{m_c} \to \mathbb{C}^m$ we have $RP = I$.

In line with the analysis of compatible relaxation in [FV04] we further define an operator $F : \mathbb{C}^{m_f} \to \mathbb{C}^m$ satisfying $RF = 0$ and $m_f = m - m_c$. Thus, $\mathcal{R}(F)$ can be interpreted as a "smoother space" complementary to $\mathcal{R}(P)$.

In this case we directly obtain the following result.

Lemma 2.46. *The operators F and R^H define an orthogonal decomposition of $\mathbb{C}^m$ in the sense that $\mathbb{C}^m = \mathcal{R}(F) \oplus \mathcal{R}\left(R^H\right)$ with $\mathcal{R}(F) \perp \mathcal{R}\left(R^H\right)$.*

Exploiting the definitions of R and F we can formulate concurrent compatible relaxation in closed form.

Lemma 2.47. *Let M be a given smoother for the linear operator A with error propagator*

$$e_{k+1} = \left(I - M^{-1}A\right) e_k.$$

Then the relaxation given by the iteration

$$e_{k+1} = \left(I - F\left(F^H M F\right)^{-1} F^H A\right) e_k \tag{2.46}$$

defines a general form of concurrent compatible relaxation.

Proof. Since $RF = 0$ we have $Re_{k+1} = Re_k$, thus (2.46) keeps the coarse-variables invariant. □

Hence, we can reformulate (2.46) by restriction to the space complementary to R as

$$e_{k+1} = E_s e_k = \left(I - \left(F^H M F\right)^{-1}\left(F^H A F\right)\right) e_k.$$

Brandt [Bra00] stated that "a general measure for the quality of the set of coarse variables is the convergence rate of compatible relaxation". In [FV04, Theorem 5.1] this statement is made more rigorous and yields a helpful tool to measure the quality of a given coarse-grid.

Theorem 2.48. *Let $\left(M^H + M - A\right)$ be hermitian and positive definite and $\widetilde{M} = M\left(M^H + M - A\right)^{-1} M^H$. Then there exists $c > 0$ such that*

$$\min_P \max_{e \neq 0} \frac{\langle \widetilde{M}\,(I - PR)\,e, (I - PR)\,e\rangle_2}{\langle Ae, e\rangle_2} \leq \frac{c}{1 - \rho_f}, \quad \rho_f = \|E_s\|_{F^H A F}.$$

Proof. By a transformation similar to [FV04, Lemma 2.3] we get

$$\begin{aligned} &\min_P \max_{e \neq 0} \frac{\langle \widetilde{M}\,(I - PR)\,e, (I - PR)\,e\rangle_2}{\langle Ae, e\rangle_2} \\ &\leq \min_P \max_{e \neq 0} \frac{\langle \sigma(M)\,(I - PR)\,e, (I - PR)\,e\rangle_2}{\langle Ae, e\rangle_2}, \end{aligned} \tag{2.47}$$

with $\sigma(M) = \frac{1}{2}\left(M + M^H\right)$ the hermitian part of M. By [FV04, Theorem 3.1] the minimum of (2.47) is given as

$$\frac{1}{\lambda_{min}\left(\left(F^H \sigma(M) F\right)^{-1}\left(F^H A F\right)\right)},$$

which in turn can be estimated as

$$\min_v \frac{\langle F^H \sigma(M) F v, v\rangle_2}{\langle F^H A F v, v\rangle_2} \leq (1 - \rho)^{-1}. \tag{2.48}$$

Inserting (2.48) in (2.47) yields the desired bound. □

Lemma 2.49. *Assume that*

$$\min_{P} \max_{e \neq 0} \frac{\langle \widetilde{M} \left(I - PR\right) e, \left(I - PR\right) e \rangle_2}{\langle Ae, e \rangle_2} \leq K, \quad \textit{for all } e \in \mathbb{C}^m.$$

That is, we obtain the following estimate for the convergence of the $V(0,1)$ two-level method

$$\| \left(I - M^{-1}A\right) \left(I - \pi_A \left(P\right)\right) e \|_A^2 \leq \left(1 - \frac{1}{K}\right) \|e\|_A.$$

Proof. For the sake of brevity we refer to [FV04, Theorem 2.2]. □

Thus, the convergence rate of compatible relaxation can be used to bound the convergence rate of a two-level Galerkin multigrid method. This makes compatible relaxation useful as a tool in algebraic multigrid to measure the quality of a given set of coarse variables with respect to the multigrid smoother.

2.7.1 Coarsening by Compatible Relaxation

As suggested in [Bra00] and further investigated in [Liv04, Bra05, BF09], compatible relaxation can be used as a tool to construct sets of coarse variables.

This is best understood in the case, where we assume that coarse variables are a subset of all variables, hence we are interested in defining a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$.

The basic idea of coarsening by compatible relaxation is quite simple. Starting with an initial set of coarse variables $\mathcal{C}_0$, we use compatible relaxation on the homogeneous equation

$$Au = 0$$

with an appropriate non-zero initial guess for u in order to measure the convergence rate ρ_f of compatible relaxation for this set of variables. Note that $\mathcal{C}_0$ can be an appropriate choice of geometric coarsening that is naturally imposed on the underlying grid of the discretization or it can be empty if no such canonical coarsening is available for the problem.

If the measured convergence rate of compatible relaxation, ρ_f, is above an a priori defined threshold γ, the current set of coarse variables is deemed insufficient to yield a good coarse-grid description of the operator and has to be enhanced by adding further variables to it. In other words, given a splitting $\Omega = \mathcal{F} \cup \mathcal{C}$ we want to find a subset of $\mathcal{F}$ that we add to $\mathcal{C}$ in

order to achieve better convergence of compatible relaxation. Apparently the choice $\mathcal{C} = \Omega$ yields $\rho_f = 0$, thus this process is well defined.

As we are interested in keeping $\mathcal{C}$ as small as possible while getting $\rho_f \leq \gamma$ we pick the subset of $\mathcal{F}$ that we add to $\mathcal{C}$ using two rules:

1. Determine the set $\mathcal{N}$ of slow-to-converge variables in $\mathcal{F}$, i.e., those variables that did not converge quickly to zero after a few compatible relaxation sweeps on the homogeneous problem (2.44).

2. In order to keep $\mathcal{C}$ as independent as possible in a graph-sense, we pick a maximally independent subset of $\mathcal{N}$ with respect to G_A to be added to $\mathcal{C}$.

Algorithm 2 cr_coarsening {Coarsening by compatible relaxation}

Input: $\mathcal{F}_0, \mathcal{C}_0, \gamma \in (0,1)$
Output: $\mathcal{F}, \mathcal{C}, \rho_f$
Set $\mathcal{F} = \mathcal{F}_0, \mathcal{C} = \mathcal{C}_0$
repeat
  Initialize $u \in \mathbb{C}^m$
  Perform compatible relaxation sweeps on $Au = 0$
  Measure convergence speed ρ_f of compatible relaxation
  if $\rho_f > \gamma$ **then**
    Find $\mathcal{N} \subset \mathcal{F}$ of slow-to-converge variables
    Find maximally independent subset $\mathcal{C}_\mathcal{N} \subset \mathcal{N}$
    Set $\mathcal{F} = \mathcal{F} \setminus \mathcal{C}_\mathcal{N}$ and $\mathcal{C} = \mathcal{C} \cup \mathcal{C}_\mathcal{N}$
  end if
until $\rho_f \leq \gamma$

Algorithm 2 illustrates a rudimentary coarsening algorithm that incorporates compatible relaxation. A more detailed and sophisticated implementation of coarsening by compatible relaxation can be found in [Liv04, BF09]. Usual choices for γ vary between .5 and .8. Note that as we are only considering the case of variable splittings here, the construction of compatible relaxation is very simple, nevertheless the generalization to arbitrary coarse variables $u^c = Ru, \quad R : \mathbb{C}^m \to \mathbb{C}^{m_c}$ is fairly straight-forward. As we can always transform an arbitrary full-rank interpolation operator by choosing a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$, such that

$$P : \mathbb{C}^{m_c} \to \mathbb{C}^m, \quad P = \begin{pmatrix} P_{fc} \\ P_{cc} \end{pmatrix}.$$

Here we assume P_{cc} to be non-singular, without changing $\mathcal{R}(P)$ we obtain

$$P = \begin{pmatrix} P_{fc}P_{cc}^{-1} \\ I \end{pmatrix} = \begin{pmatrix} W \\ I \end{pmatrix}.$$

With this observation it is possible to extend the idea of compatible relaxation in its simple formulation based on a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ to a setting where the coarse-grid variables are defined differently, e.g., aggregation-based approaches, by carefully construction of an associated variable splitting.

This practical method for algebraically constructing the set of coarse variables in a general setting concludes our review of state-of-the-art approaches in algebraic multigrid.

Chapter 3

Applications

3.1 Lattice Gauge Theory

The Dirac operator is used in quantum field theory to model the interaction of particles (fermions) and their interacting counterparts (bosons). The Dirac equation without background gauge field can be seen as a relativistic quantum mechanical wave equation that is consistent with the principles of special relativity and quantum mechanics. We do not go into details of the derivation of the equation, but rather focus on its mathematical representation and properties.

In the continuum, the Dirac operator $\mathcal{D}$ with a background gauge field A is given in d-dimensional euclidean space as

$$\mathcal{D} = \sum_{\mu=1}^{d} \gamma_\mu \otimes (\partial_\mu + iA_\mu), \tag{3.1}$$

where $\gamma_\mu \in \mathbb{C}^{d\times d}$ are generators of a Clifford or Dirac-Algebra, i.e., they satisfy

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 0 \quad \text{and} \quad \gamma_\mu^2 = I \tag{3.2}$$

and $\otimes$ denotes the tensor product. $A_\mu(x) \in \mathbb{C}^{n\times n}$ is a function that satisfies at any point $x \in \mathbb{C}^d$, $A_\mu^*(x) = A_\mu(x)$ and $\operatorname{trace}(A_\mu(x)) = 0$, i.e., $A_\mu(x) \in su(n)$. $\partial_\mu = \frac{\partial}{\partial x_\mu}$ denotes the partial derivative in direction μ.

3.1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD), a quantum field theory, models the strong interaction between quarks, the subatomic constituents of matter, and gluons, their interacting counterparts.

The formal Dirac operator $\mathcal{D}$ of QCD is usually understood to act on a $d = 4$-dimensional euclidean space with 3 space- and 1 time-dimension, where the space is transformed from a Minkowski-space to an euclidean space-time system. In what follows, we choose the generators of the Clifford-Algebra that fulfill (3.2), by

$$\gamma_1 = \begin{pmatrix} & & & i \\ & & i & \\ & -i & & \\ -i & & & \end{pmatrix} \quad \gamma_2 = \begin{pmatrix} & & & -1 \\ & & 1 & \\ & 1 & & \\ -1 & & & \end{pmatrix}$$
$$\gamma_3 = \begin{pmatrix} & & i & \\ & & & -i \\ -i & & & \\ & i & & \end{pmatrix} \quad \gamma_4 = \begin{pmatrix} & & 1 & \\ & & & 1 \\ 1 & & & \\ & 1 & & \end{pmatrix}. \tag{3.3}$$

The gauge field A of QCD is given as complex 3×3 traceless hermitian matrices, i.e., $A_\mu \in \mathbb{C}^{3\times 3}$ with $A_\mu^* = A_\mu$ and trace $(A_\mu) = 0$. Thus the Dirac operator of QCD is a 12×12 coupled system of first order partial differential operators.

To analyze this model, Wilson proposed in [Wil74] to simulate these interactions using a discretized formulation of the Dirac operator within the context of Lattice Gauge Theory (LGT). For this purpose, a lattice with $N \times N$ lattice points is introduced with periodic boundary conditions. The lattice spacing is often denoted by a, but for the sake of convenience one typically parametrizes the formulation such that $a = 1$.

In order to discretize the Dirac equation (3.1) one then defines covariant finite differences that preserve the property of gauge invariance of the given differential operator. This is accomplished by including a discretized gauge field in the definition.

Definition 3.1. *Given the lattice field $A_\mu(x)$ at lattice point x we define the discretized gauge field on a $N \times N$ lattice and lattice-spacing a by*

$$U_\mu^x = e^{-\frac{i}{a} A_\mu(x + \frac{a}{2} e_\mu)} \approx e^{-i \int_x^{x+ae_\mu} A_\mu(s) ds}$$

A collection $\mathcal{U} = \{U_\mu^x\} \subset SU(n)$ is called a gauge configuration*, living on the edges of the lattice.*

The gauge configuration can be interpreted as an approximation to the action of the gauge field along a lattice edge by approximating the integral over the gauge filed along the edge. In Figure 3.1 we illustrate the naming conventions on the lattice.

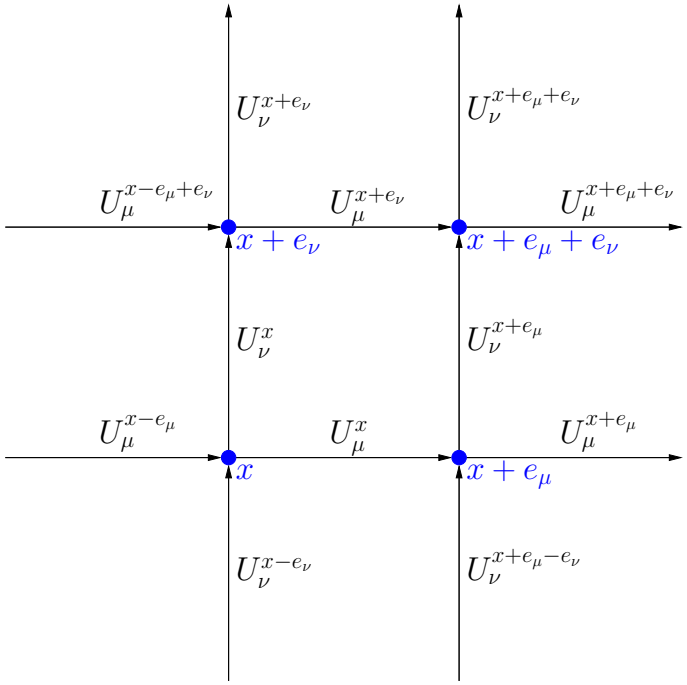


Figure 3.1: The typically used naming convention in Lattice Gauge Theory with gauge configuration $\mathcal{U} = \{U_\mu^x\}$.

Remark 3.2. *Associated with a gauge configuration $\mathcal{U}$ is a temperature parameter β that controls the randomness of the underlying gauge field.*

With this discrete description of the gauge field we can now proceed to define the finite covariant differences.

Definition 3.3. *The forward finite covariant difference of the spinor ψ at lattice site x in direction μ is defined as*

$$\partial^\mu \psi_x = U_\mu^x \psi_{x+e_\mu} - \psi_x,$$

where $x+e_\mu$ is the neighboring lattice site in spatial direction μ. Analogously, one defines a backward finite covariant difference

$$\partial_\mu \psi_x = \psi_x - \left(U_\mu^{x-e_\mu}\right)^H \psi_{x-e_\mu}.$$

Using the forward and backward covariant finite difference operators we can define a central covariant finite difference operator by

$$\hat{\partial}_\mu \psi_x = \frac{1}{2}\left(\partial_\mu + \partial^\mu\right)\psi_x.$$

Hence, the straight-forward formulation of the Dirac operator on the lattice using central discretization for the first order derivative is given by

$$D_L = \frac{1}{2}\sum_{\mu} \gamma_\mu \otimes (\partial_\mu + \partial^\mu) = \frac{1}{2}\sum_{\mu} \gamma_\mu \otimes \hat{\partial}_\mu.$$

However, this formulation suffers from the effect of species doubling encountered when using first order centralized finite difference discretizations, i.e., unphysical low energy modes of the discretized operator (cf. [DD06] for more details). In the context of centralized finite difference approximations of first order partial differential operators this effect is also known as "red-black" instability.

To avoid doubling, Wilson introduced in [Wil74] a stabilization term $\sum_\mu \hat{\partial}_\mu \hat{\partial}^\mu$ that can be seen as equivalent to the Laplace operator, Δ, using covariant finite differences with a background gauge configuration $\mathcal{U} \subset SU(n)$. By addition of this stabilization term, the problem of species doubling is removed.

Definition 3.4. *The Wilson-Dirac operator D_W is defined by*

$$D_W = \frac{1}{2}\left(\sum_{\mu=1}^{4} \gamma_\mu \otimes (\partial_\mu + \partial^\mu) - \hat{\partial}_\mu \hat{\partial}^\mu\right).$$

Usually one adds an additional mass term m to the operator in order to reflect specific properties of the underlying physics, e.g., quark masses. Hence, the problem to solve numerically is given by

$$(D_W + mI)\,\psi = \varphi.$$

Spectral properties of the Dirac-Wilson operator

We first take a look at an important symmetry that the Dirac and Dirac-Wilson operator fulfill. The generators of the Clifford algebra $\gamma_1, \ldots, \gamma_4$ define an operator γ_5 by

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$$

that satisfies

$$\gamma_5\gamma_\mu + \gamma_\mu\gamma_5 = 0 \quad \text{and} \quad \gamma_5^2 = I.$$

Furthermore, γ_5 is hermitian, i.e., $\gamma_5^* = \gamma_5$, and it has only eigenvalues ± 1 as $\gamma_5^2 = I$. Hence, we can define two projections

$$\frac{I - \gamma_5}{2} \quad \text{and} \quad \frac{I + \gamma_5}{2} \tag{3.4}$$

that project any spinor ψ onto its left- or right-handed components, respectively. In the case of the basis of γ-matrices chosen in (3.3), γ_5 is given by

$$\gamma_5 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

and the projections (3.4) have the simple form

$$\frac{I-\gamma_5}{2} = \begin{pmatrix} & & & \\ & 1 & & \\ & & & \\ & & & 1 \end{pmatrix} \quad \text{and} \quad \frac{I+\gamma_5}{2} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & & \\ & & & \end{pmatrix}.$$

An important property of the Dirac-Wilson operator is the γ_5-hermiticity. That is, the continuum operator satisfies

$$\gamma_5 \mathcal{D}^* = \mathcal{D}\gamma_5.$$

This property is preserved in the Wilson formulation of the Lattice Dirac operator. In order to show this, we introduce a lattice equivalent of the γ_5 operator, the Γ_5 operator, defined by

$$\Gamma_5 = \gamma_5 \otimes I_{N^2} \otimes I_3 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \in \mathbb{C}^{12N^2 \times 12N^2}.$$

With this definition we obtain

$$\Gamma_5 D_W^H = D_W \Gamma_5.$$

Due to this symmetry, we can derive many properties of the eigenvalues and eigenvectors of the Wilson-Dirac operator D_W that lead to important insights for the development and implementation of an adaptive algebraic multigrid method for this system.

Definition 3.5. *Given the non-hermitian operator D_W, we denote the right eigenvector of D_W with eigenvalue λ by v_λ^r with,*

$$D_W v_\lambda^r = \lambda v_\lambda^r.$$

Analogously, we denote the left eigenvector of D_W with eigenvalue λ by v_λ^l with,

$$\left(v_\lambda^l\right)^H D_W = \lambda \left(v_\lambda^l\right)^H.$$

Lemma 3.6. *The eigenvalues λ of D_W are either real or appear in complex conjugate pairs.*

Proof. The characteristic polynomial of D_W is defined by

$$\begin{aligned}\det\left(D_W-\lambda I\right) &= \det\left(\Gamma_5\right)\det\left(D_W-\lambda I\right)\det\left(\Gamma_5\right)\\ &= \det\left(\Gamma_5 D_W\Gamma_5-\lambda I\right)\\ &= \det\left(D_W^H-\lambda I\right)=\det\left(D_W-\bar{\lambda}I\right),\end{aligned}$$

where $\bar{\lambda}$ denotes the complex conjugate of λ. Hence, λ is an eigenvalue of D_W if and only if $\bar{\lambda}$ is also an eigenvalue. □

Lemma 3.7. *For any right eigenvector v with eigenvalue λ, i.e., fulfilling $D_W v=\lambda v$, the vector $(\Gamma_5 v)$ is the left eigenvector to the eigenvalue $\bar{\lambda}$, satisfying*

$$\left(\Gamma_5 v\right)^H D_W=\bar{\lambda}\left(\Gamma_5 v\right)^H.$$

Proof. Using the Γ_5-hermiticity of D_W we have

$$\begin{aligned}\left(\Gamma_5 v\right)^H D_W &= v^H\Gamma_5 D_W\\ &= v^H D_W^H\Gamma_5\\ &= \bar{\lambda}v^H\Gamma_5=\bar{\lambda}\left(\Gamma_5 v\right)^H.\end{aligned}$$

□

Combining Lemmas 3.6 and 3.7 we have eigentriples for any pair of complex conjugate eigenvalues

$$\left(\lambda,v_\lambda^r,v_\lambda^l\right)\quad\text{and}\quad\left(\bar{\lambda},v_{\bar{\lambda}}^r,v_{\bar{\lambda}}^l\right)$$

that satisfy

$$v_\lambda^l=\Gamma_5 v_{\bar{\lambda}}^r\quad\text{and}\quad v_{\bar{\lambda}}^l=\Gamma_5 v_\lambda^r.$$

Due to this fact the bi-orthogonality of the right and left eigenvectors of the Wilson-Dirac operator D_W are given as follows.

Lemma 3.8. *Let $(\lambda_i,v_{\lambda_i}^r,v_{\lambda_i}^l)$ denote the eigentriples of D_W. Then we have with $c_{ij}\in\mathbb{C}\setminus\{0\}$,*

$$\left(\Gamma_5 v_{\lambda_i}^l\right)^H v_{\lambda_j}^r=\begin{cases}c_{ij}, & \text{if}\quad\lambda_i=\bar{\lambda}_j\\ 0, & \text{else}\end{cases}.$$

Proof. We have

$$\left(\Gamma_5 v^l_{\lambda_i}\right)^H v^r_{\lambda_j} = \frac{1}{\overline{\lambda_i}} \left(\Gamma_5 v^l_{\lambda_i}\right)^H D_W v^r_{\lambda_j} \text{ and } \left(\Gamma_5 v^l_{\lambda_i}\right)^H v^r_{\lambda_j} = \frac{1}{\lambda_j} \left(\Gamma_5 v^l_{\lambda_i}\right)^H D_W v^r_{\lambda_j}.$$

Hence the difference of these two equations is given by

$$0 = \left(\frac{1}{\overline{\lambda_i}} - \frac{1}{\lambda_j}\right) \left(\Gamma_5 v^l_{\lambda_i}\right)^H D_W v^r_{\lambda_j}.$$

□

Remark 3.9. *Any vector $w \in \mathbb{C}^m$ that fulfills $(\Gamma_5 w)^H w = 0$ is said to have* zero chiral weight. *Due to Lemma 3.8 all right and left eigenvectors eigenvalues $\lambda \in \mathbb{C} \setminus \mathbb{R}$ of D_W have zero chiral weight. Further we observe that we have for any eigenvector v of D_W with zero chiral weight*

$$0 = (\Gamma_5 v)^H D_W v = v^H (\Gamma_5 D_W) v.$$

That is, their inner product with the hermitian indefinite operator $\Gamma_5 D_W$ is zero.

Furthermore, we find a relation between the singular values of D_W and the eigenvalues of the hermitian operator $(\Gamma_5 D_W)$.

Lemma 3.10. *If a unitary eigendecomposition of $(\Gamma_5 D_W)$ is given by*

$$(\Gamma_5 D_W) Z = Z\Lambda, \tag{3.5}$$

a singular value decompostion of D_W is given by

$$D_W V = U\Sigma, \quad \textit{with} \quad V = Z \quad \textit{and} \quad U = \Gamma_5 Z \, \textit{sign}(\Lambda).$$

Proof. Multiplication of (3.5) by Γ_5 yields

$$D_W Z = \Gamma_5 Z \Lambda.$$

As Z and $\Gamma_5 Z$ are unitary and $\text{sign}(\Lambda)\Lambda > 0$ we have the singular decomposition

$$D_W Z = \Gamma_5 Z \text{sign}(\Lambda) \Lambda.$$

□

The Wilson formulation of the Lattice Dirac operator is not the only formulation in Lattice Quantum Chromodynamics, but it plays a crucial role in other formulations as well (e.g., the chiral domain wall and overlap formulations). Both systems can be reduced to solutions of the aforementioned Wilson formulation. As such, we focus on this formulation.

3.1.2 Quantum Electrodynamics

In the development of algorithms in Lattice Gauge Theory, the operators arising in Quantum Electrodynamics (QED) are oftentimes used as a first test-bed, because of the similar spectral behavior and symmetries.

Schwinger model

The Dirac operator of QED acts on a $d = 2$-dimensional euclidean space and 2 spin components, with the generators of the Clifford-Algebra now given by the so-called Pauli-matrices

$$\gamma_1 = \sigma_1 = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \quad \text{and} \quad \gamma_2 = \sigma_2 = \begin{pmatrix} & i \\ -i & \end{pmatrix}. \tag{3.6}$$

The gauge field A_μ of QED in the Schwinger formulation is given as a continuous real-valued function, i.e., the gauge configurations $\mathcal{U}$ are a subset of the complex numbers with modulus one. Thus the Dirac operator of QED is a 2×2 matrix of first order partial differential operators.

With the notion of finite covariant differences given in Definition 3.3, we can write the Wilson-stabilized discretization of the Schwinger model for a given Gauge configuration $\mathcal{U}$ of $U(1)$.

Definition 3.11. *The Wilson-Schwinger operator S_W is defined by*

$$S_W = \frac{1}{2}\left(\sum_{\mu=1}^{2} \sigma_\mu \otimes (\partial_\mu + \partial^\mu) - \partial_\mu \partial^\mu\right).$$

As before, a mass term that reflects certain physical properties of the fermions involved is added to the Wilson-Schwinger operator. Thus, whenever we talk about the Wilson-Schwinger operator S_W we think of the massive operator $S_W + mI$, for a given mass m.

Spectral properties of the Wilson-Schwinger operator

The discussion of spectral properties of the Wilson-Schwinger operator is mostly based on the σ_3-symmetry of the operator. Analogously, to the formulation in QCD we define σ_3 by

$$\sigma_3 = i\sigma_1\sigma_2 = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

Hence, we have the projections

$$\frac{I - \sigma_3}{2} = \begin{pmatrix} & \\ & 1 \end{pmatrix} \quad \text{and} \quad \frac{I + \sigma_3}{2} = \begin{pmatrix} 1 & \\ & \end{pmatrix}$$

that project any spinor ψ onto its left- and right-handed components, respectively.

The continuum Schwinger operator $\mathcal{S}$ satisfies

$$\sigma_3 \mathcal{S}^* = \mathcal{S}\sigma_3.$$

Again, this symmetry carries over to the Wilson formulation. That is, with

$$\Sigma_3 = \sigma_3 \otimes I_{N^2} = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \in \mathbb{C}^{2N^2 \times 2N^2}$$

we obtain

$$\Sigma_3 S_W^H = S_W \Sigma_3.$$

Using this symmetry, we find important properties of eigenvalues and eigenvectors of the Wilson-Schwinger operator S_W that help us in the development of an adaptive algebraic multigrid method for this system.

As before we denote the right eigenvector to the eigenvalue λ by v_λ^r and the corresponding left eigenvector by v_λ^l.

Lemma 3.12. *The eigenvalues λ of S_W are either real or appear in complex conjugate pairs.*

Proof. Analogously to Lemma 3.6 □

Lemma 3.13. *For any right eigenvector v with eigenvalue λ, i.e., fulfilling $S_W v = \lambda v$, the vector $(\Sigma_3 v)$ is the left eigenvector to the eigenvalue $\bar{\lambda}$ satisfying*

$$(\Sigma_3 v)^H S_W = \bar{\lambda} (\Sigma_3 v)^H .$$

Proof. Analogously to Lemma 3.7. □

Combining Lemmas 3.12 and 3.13 we have eigentriples for any pair of complex conjugate eigenvalues

$$\left(\lambda, v_\lambda^r, v_\lambda^l\right) \quad \text{and} \quad \left(\bar{\lambda}, v_{\bar{\lambda}}^r, v_{\bar{\lambda}}^l\right)$$

that satisfy

$$v_\lambda^l = \Sigma_3 v_{\bar{\lambda}}^r \quad \text{and} \quad v_{\bar{\lambda}}^l = \Sigma_3 v_\lambda^r.$$

Due to this fact the bi-orthogonality of the right and left eigenvectors of the Wilson-Schwinger operator S_W are given as follows.

Lemma 3.14. *Let $(\lambda_i, v_{\lambda_i}^r, v_{\lambda_i}^l)$ denote the eigentriples of S_W. Then we have with $c_{ij} \in \mathbb{C} \setminus \{0\}$,*

$$\left(\Sigma_3 v_{\lambda_i}^l\right)^H v_{\lambda_j}^r = \begin{cases} c_{ij}, & \text{if} \quad \lambda_i = \bar{\lambda}_j \\ 0, & \text{else} \end{cases} .$$

Proof. Analogously to Lemma 3.8. □

Remark 3.15. *Any vectors $w \in \mathbb{C}^m$ that fulfill $(\Sigma_3 w)^H w = 0$ are said to have* zero chiral weight. *Due to Lemma 3.14 all right and left eigenvectors to complex eigenvalues of S_W have zero chiral weight. Further we observe that we have for any eigenvector v of S_W with zero chiral weight*

$$0 = (\Sigma_3 v)^H S_W v = v^H (\Sigma_3 S_W) v.$$

That is, their inner product with the hermitian indefinite operator $\Sigma_3 S_W$ is zero.

Furthermore, we find a relation between the singular values of S_W and the eigenvalues of the hermitian operator $(\Sigma_3 S_W)$.

Lemma 3.16. *If a unitary eigendecomposition of $(\Sigma_3 S_W)$ is given by*

$$(\Sigma_3 S_W) Z = Z\Lambda,$$

a singular value decompostion of S_W is given by

$$S_W V = U\Sigma, \quad \textit{with} \quad V = Z \quad \textit{and} \quad U = \Sigma_3 Z \, \textit{sign}(\Lambda).$$

Proof. Analogously to Lemma 3.10. □

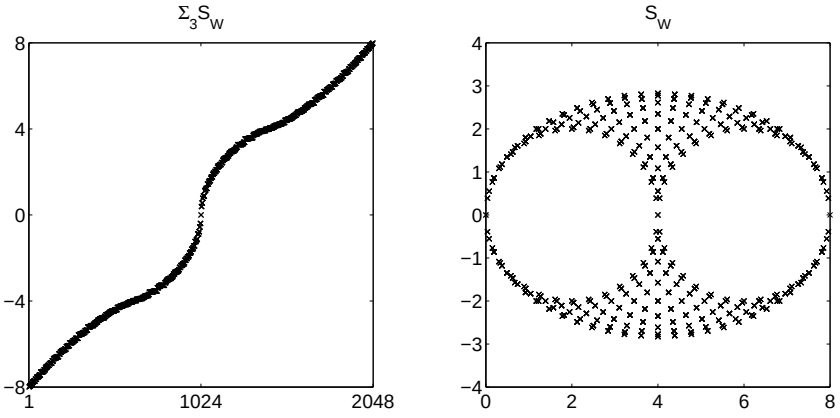
In Figure 3.2 the eigenvalue distributions of $\Sigma_3 S_W$ and S_W are shown first for the free case, i.e., with a constant gauge configuration $\mathcal{U} = \{U_\mu^x\}$, with $U_\mu^x = 1$ for all μ and x on a 32×32 and for a physical gauge configuration $\mathcal{U}$ at temperature $\beta = 5$. In Figure 3.3 we present the entry-wise modulus of the spin-components s_1, s_2 of the eigenvectors $v_\lambda, v_{\bar{\lambda}}$ associated to the smallest, in modulus, complex conjugate pair of eigenvalues $\lambda, \bar{\lambda}$ of the Wilson-Schwinger operator on a 64×64 lattice with a gauge configuration at temperature $\beta = 5$.

3.1.3 The Gauge Laplace operator

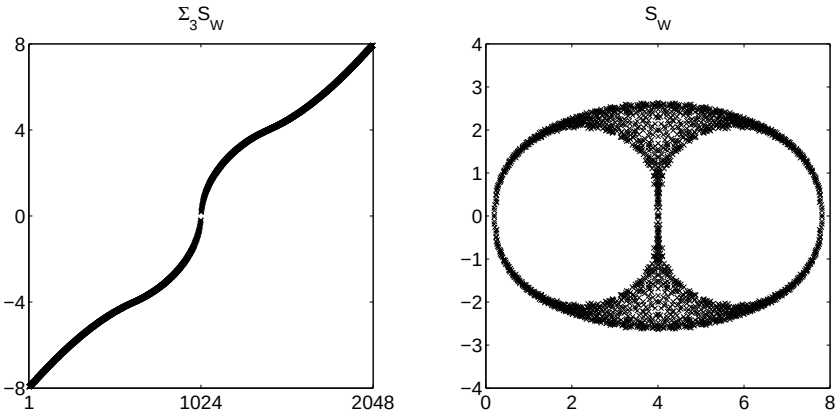
The Gauge Laplace operator is the discrete covariant cousin of the well-known Laplace operator and can be defined accordingly on the lattice using finite covariant differences.

Definition 3.17. *For any gauge configuration $\mathcal{U} = \{U_\mu^x\} \subset SU(n)$ define*

$$A(\mathcal{U}) = \sum_\mu \hat{\partial}_\mu \hat{\partial}^\mu.$$



(a) $\mathcal{U} = \{U_\mu^x\}$, with $U_\mu^x = 1$



(b) $\mathcal{U}$ at $\beta = 5$

Figure 3.2: Eigenvalue distributions of $\Sigma_3 S_W$ and S_W for a constant gauge configuration $\mathcal{U} = \{U_\mu^x\}$, with $U_\mu^x = 1$ in 3.2(a) and a physical gauge configuration $\mathcal{U}$ at temperature $\beta = 5$ in 3.2(b) on a 32×32 lattice.

The resulting operator $A(\mathcal{U})$ is called the Gauge Laplace *operator and resembles the 5-point discretization of Δ with the links replaced by the discrete gauge configuration $\mathcal{U}$ in $SU(n)$. In the special case of $\mathcal{U} \subset U(1)$ we obtain that the 5-point finite difference discretization of Δ equals the Gauge Laplace operator $A(\mathcal{U})$ in the case $U_\mu^x = 1$ for all x, μ. That is, in case there is no*

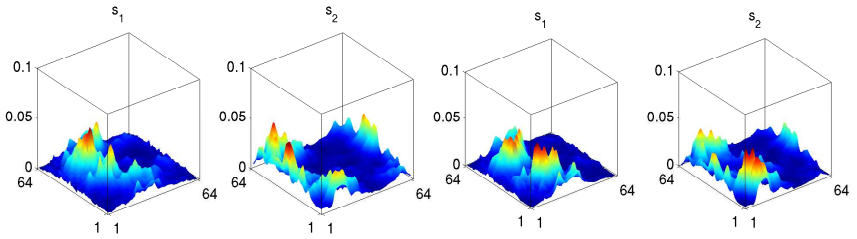


Figure 3.3: Entry-wise modulus of the spin components s_1, s_2 of the eigenvectors corresponding to the smallest, in modulus, pair of complex conjugate eigenvalues of a Wilson-Schwinger operator on a 64×64 grid with a gauge configuration at temperature $\beta = 5$.

background gauge field A_μ. We further denote this special case by $A(\mathcal{U}_0)$.

The Gauge Laplace operator is yet another simplification of the Wilson-Dirac and Wilson-Schwinger operator that is used to test methods for either one of these operators. In order to understand the role of the Gauge Laplace operator we observe that the block-diagonal of the Wilson-Dirac operator is given by 4 copies of the Gauge Laplace operator $A(\mathcal{U})$, with $\mathcal{U} \subset SU(3)$. Using the basis for the Clifford-Algebra given in (3.3) and denoting the spatial directions by $x_1, \ldots, x_4$, permutation with respect to the spin components $1, \ldots, 4$ yields

$$D_W\psi = \begin{pmatrix} A(\mathcal{U}) & & i\hat{\partial}_{x_3} + \hat{\partial}_{x_4} & i\hat{\partial}_{x_1} - \hat{\partial}_{x_2} \\ & A(\mathcal{U}) & i\hat{\partial}_{x_1} + \hat{\partial}_{x_2} & -i\hat{\partial}_{x_3} + \hat{\partial}_{x_4} \\ i\hat{\partial}_{x_3} - \hat{\partial}_{x_4} & i\hat{\partial}_{x_1} - \hat{\partial}_{x_2} & A(\mathcal{U}) & \\ i\hat{\partial}_{x_1} + \hat{\partial}_{x_2} & -i\hat{\partial}_{x_3} - \hat{\partial}_{x_4} & & A(\mathcal{U}) \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}.$$

A similar reordering can be found for the Wilson-Schwinger operator. With a gauge configuration $\mathcal{U} \subset U(1)$, using the Pauli matrices (3.6) in original numbering as generators for the algebra and denoting the space directions by x, y, we obtain by permutation with respect to the spin components $1, 2$,

$$S_W\psi = \begin{pmatrix} A(\mathcal{U}) & \left(\hat{\partial}_x + i\hat{\partial}_y\right) \\ -\left(\hat{\partial}_x - i\hat{\partial}_y\right) & A(\mathcal{U}) \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}.$$

Motivated by this observation we start the development of algorithms for Quantum Dynamics with the Gauge Laplace operator.

Spectral properties of the Gauge Laplace operator

By Definition 3.17 the Gauge Laplace operator $A(\mathcal{U})$ on the lattice, associated with the background gauge field configuration $\mathcal{U}$ is hermitian, i.e.,

$$(A(\mathcal{U}))^H = A(\mathcal{U}).$$

Hence, the spectrum of the operator is real. For masses $m > m_c$, where m_c denotes the smallest eigenvalue of $A(\mathcal{U})$ the Gauge Laplace operator is positive definite.

An important issue to consider when developing solvers for the Gauge Laplace system is the local character of its algebraically smooth error, i.e., its near-kernel. Figure 3.4 contains plots of the modulus, real and imaginary part, of the eigenvector of the smallest eigenvalue of a Gauge Laplace operator

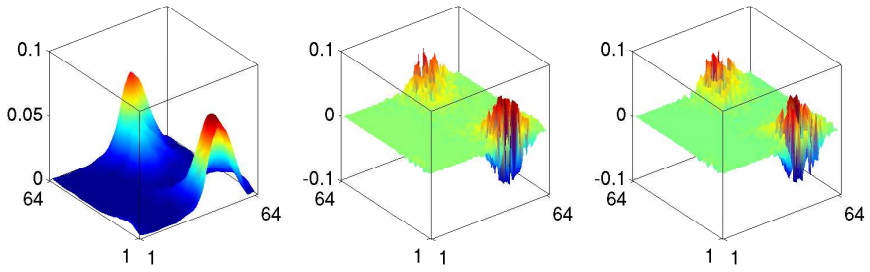


Figure 3.4: Modulus, real and imaginary part of the eigenvector to the smallest eigenvalue of $A(\mathcal{U})$ for gauge configuration $\mathcal{U}$ at temperature $\beta = 5$ on a 64×64 lattice.

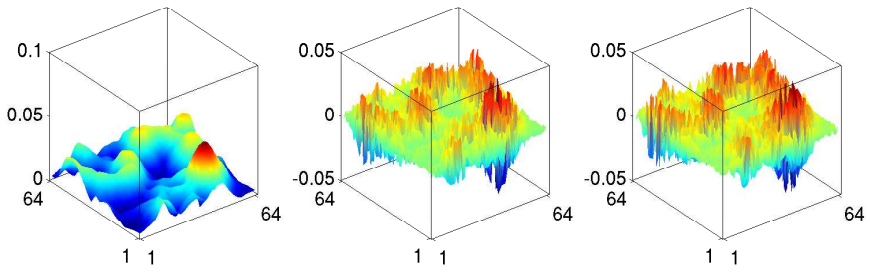


Figure 3.5: Modulus, real and imaginary part of a random complex vector (N(0, 1) distributed) after 100 Gauss-Seidel iterations with right-hand-side zero for a Gauge Laplace operator $A(\mathcal{U})$ on a 64×64 lattice with a gauge configuration $\mathcal{U}$ at temperature $\beta = 5$.

$A(\mathcal{U})$ on a 64×64 lattice and a gauge configuration $\mathcal{U}$ at temperature $\beta = 5$. In Figure 3.5 we provide a plot of an initially random (N$(0, 1)$ distributed) vector that we smoothed using 100 Gauss-Seidel relaxations for a Gauge Laplace operator $A(\mathcal{U})$ on a 64×64 lattice with a gauge configuration at temperature $\beta = 5$. Clearly, we see that due to the random background gauge field, the eigenvectors associated with small eigenvalues tend to be locally supported and highly oscillatory, both properties are challenging problems in the development of algebraic multigrid methods for such systems.

3.2 Markov Chains

Another application under consideration in this thesis is the so-called class of Markov chain problems. The origin of these problems is the evolution of stochastic processes in time.

Definition 3.18. *A stochastic process is called a* finite homogeneous Markov chain *if it has $m \in \mathbb{N}$ states $s_1, \ldots, s_m$ and associated transition probabilities $p(s_i \to s_j) \in [0, 1]$ which are time independent.*

For sake of simplicity we say Markov chain instead of finite homogeneous Markov chain in the course of this thesis. Time independence means that the transition probabilities $p(s_i \to s_j)$ do not change during the evolution of the stochastic process.

Definition 3.19. *A vector $x \in \mathbb{R}^m$ is called a* probability distribution vector *if*

$$x_i \in [0, 1] \quad \textit{and} \quad \|x\|_1 = 1.$$

In the Markov chain context the component i in a probability distribution x is the probability of the Markov chain to be in state s_i.

Definition 3.20. *Let a Markov chain with states $s_1, \ldots, s_m$ and transition probabilities $p(s_i \to s_j)$ be given. We define the* transition matrix $A = (a_{ij})_{ij} \in [0, 1]^{N \times N}$ *by*

$$a_{ij} = p(s_j \to s_i).$$

Due to the fact that $\sum_{j=1}^{m} p(s_i \to s_j) = 1$ any transition matrix has column-sum 1*, i.e., it is* column stochastic*.*

With the equivalent description of the Markov chain process by its transition matrix A one can write down the time evolution of the process. Given a probability distribution x^k at time step k we get the probability distribution at time step $k + 1$ by

$$x^{k+1} = Ax^k.$$

Definition 3.21. *The probability distribution x is called the* steady-state vector *of the Markov chain given by its transition matrix A if*

$$Ax = x.$$

Hence, steady-state vectors correspond to eigenvectors of the transition matrix with eigenvalue one.

In order to discuss existence and uniqueness of steady-state vectors of Markov chains we need some further definitions and notations.

Definition 3.22. *A vector is $x \in \mathbb{R}^m$ is called* non-negative *if*

$$x_i \geq 0,\ i = 1, \ldots, m.$$

Analogously a matrix $A \in \mathbb{R}^{m \times m}$ is called non-negative *if*

$$a_{ij} \geq 0,\ i, j = 1, \ldots, m.$$

To simplify notation we write $x \geq 0$ and $A \geq 0$, respectively.

Definition 3.23. *The spectral radius $\rho(A)$ of the matrix $A \in \mathbb{R}^{m \times m}$ is defined by*

$$\rho(A) = \max_{1 \leq i \leq m} |\lambda_i|,$$

where λ_i are the eigenvalues of A.

Definition 3.24. *A matrix $A \in \mathbb{R}^{m \times m}$ is called* reducible *if there exists a permutation $\pi \in \mathbb{R}^{m \times m}$ of $\{1, \ldots, m\}$ such that*

$$\pi A \pi^T = \begin{pmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{pmatrix}.$$

Otherwise it is called irreducible.

The transition matrix A of a Markov chain is irreducible if no probability sink exists, i.e., there exists a path in the corresponding directed graph of the matrix from each state s_i to any other state s_j.

Theorem 3.25. *Let $A \geq 0$ be irreducible. Then $\rho(A)$ is an eigenvalue of A and there exists a unique (up to scaling) right eigenvector $x > 0$ with $Ax = \rho(A)x$.*

Proof. See for example [BP94, page 27]. □

Corollary 3.26. *For any Markov chain with an irreducible transition matrix A, there exists a unique steady-state vector $x > 0$.*

Proof. By Definition 3.20 we have $A \geq 0$. Furthermore, as A is column-stochastic we have $\rho(A) = 1$. Due to Theorem 3.25 a unique steady-state vector x with $Ax = x$ exists. □

We present an algorithm in chapter 5 based on the bootstrap algebraic multigrid approach developed in chapter 4 that is able to compute steady-state solutions of Markov chain problems. The Markov chain problem is of particular interest as it is not only a first non-symmetric test for the algorithms we develop, but also requires to resolve a certain error component up to arbitrary accuracy in the multigrid hierarchy. This is an important feature in the construction of black-box multigrid methods for singular matrices.

Chapter 4

General Framework for AMG

As discussed in chapter 2, algebraic multigrid methods construct a multigrid hierarchy using a setup algorithm. The focus of this chapter is on the derivation of robust multigrid setup algorithms based on the adaptive definitions of interpolation and restriction operators. In section 4.1, we first present a modified version of the adaptive reduction-based algebraic multigrid approach, and develop a generalized two-grid convergence theory for this method that allows us to handle complex-valued systems. Although, the work on adaptive reduction-based algebraic multigrid leads to several new results, the results shown in section 4.1.2 suggest that the proposed method has limited success in treating all challenges of the systems arising in Lattice Gauge Theory. Hence, we consider developing a much more generally applicable method in a bootstrap algebraic multigrid framework. This includes the introduction and analysis of a least squares based definition of interpolation in section 4.2 and the description of a collection of adaptive multigrid techniques in section 4.3 that can be used to further improve the performance of algebraic multigrid methods. Since the main focus of this chapter is on the development of interpolation in algebraic multigrid, we postpone the discussion of smoothers and algebraic coarsening to chapter 5, where we propose algorithms based on the adaptive techniques developed in this chapter for the problems outlined in chapter 3. Note that, although the emphasis of this work is on problems outlined in chapter 3, the ideas developed in this chapter yield general tools for algebraic multigrid setup algorithms. The individual newly devised or refined adaptive techniques are introduced in a modular way.

In recent years, significant effort has been invested on improving the range of applicability of black-box multigrid techniques. There are many approaches to achieving robust multigrid solvers for wide classes of matrices. Adaptive multigrid methods [BFM$^+$04, BFM$^+$06, MMM06] offer many ad-

vantages because of the efficiency they inherit from the algebraic multigrid approaches on which they are based [RS87, VMB96, VBM01]. The key idea behind adaptive multigrid algorithms is to experimentally use the multigrid smoother itself to expose those error components that must be accurately accounted for in the coarse-grid correction process, the so-called "algebraically smooth" errors that the given smoother is slow to reduce. In its simplest form, this amounts to simply iterating many times with a fixed stationary iterative (relaxation) method on the homogeneous problem, $Au = 0$, with a random initial guess. The dominant error left after many relaxation sweeps must, by definition, reflect the algebraically smooth error components of the problem. These error components can then be built into the coarse-grid correction process by introducing them into the range of interpolation, e.g., by means of the smoothed aggregation framework introduced in section 2.6. In practice, exposing these error components by simple relaxation alone is very inefficient and, so, the process is accelerated by a multilevel relaxation process that exposes the local and global characteristics of these slow-to-converge error components simultaneously.

In what follows, we first introduce a variant of adaptive reduction-based algebraic multigrid, as devised in [MMM06]. Then a generalized two-grid convergence proof is given for this new formulation and the capabilities and limitations of this approach for the application to the Gauge Laplace operator of Lattice Gauge Theory, as introduced in section 3.1.3, are discussed in section 4.1.2.

Next, we present a least squares based formulation of interpolation in section 4.2, first mentioned by Brandt in [Bra02]. For this formulation, we present its connection to a least squares formulation in element-free AMG$\not e$, proposed by Falgout in [Fal02], which generalizes least squares interpolation in an AMG$\not e$ framework. We further analyze the convergence theory for least squares interpolation as rigorous as possible to date.

In addition, we derive bootstrap techniques in section 4.3 that are mutually beneficial to the least squares interpolation, but not limited to this approach. These techniques allow for a multiscale setup process based on coarsest grid eigensolves and lead to an observation that allows treatment of the misfit in interpolation of eigenvectors corresponding to small eigenvalues on the coarsest grid, where they can be represented.

A thorough study of algorithms composed of the adaptive techniques described in this chapter is given in chapter 5.

4.1 Adaptive Reduction-based AMG Interpolation

The reduction-based algebraic multigrid (AMGr) approach, introduced in section 2.5, can be used to define an adaptive reduction-based method, as discussed in [MMM06].

Recall that the reduction-based framework, assumes that a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ is given, and then tries to approximate the "ideal" interpolation (2.38) by replacing A_{ff}^{-1} with a sparse approximation to it. Adaptivity in this framework can be implemented by enforcing the sparse approximation to fulfill certain constraints, i.e., given a known or computed representative e of algebraically smooth error, we define $\tilde{A}_{ff}^{-1}$, such that

$$\tilde{A}_{ff}^{-1} A_{fc} e_c = A_{ff}^{-1} A_{fc} e_c. \tag{4.1}$$

In cases where algebraically smooth error can be represented by $e = \mathbf{1}$, the constant vector, it is easily seen that for diagonally dominant matrices with constant diagonal, $D = \mathrm{diag}(A)$ fulfills the assumptions of Theorem 2.40 guaranteeing two-grid convergence of the approach when using $\mathcal{F}$-relaxation (2.40) with D_{ff}^{-1} for this particular approximation of A_{ff}^{-1}.

However, the assumptions made in Theorem 2.40 were focused on only satisfying (2.41). In a more general setting without any assumptions on the nature of the representatives of algebraically smooth error, enforcing (2.42) appears to be of equal importance. Consider for example a hermitian positive definite (hpd) operator that is diagonally dominant with a constant diagonal, but its eigenvector corresponding to the smallest eigenvalue is not well represented by $\mathrm{diag}(A)$ (e.g., Gauge Laplace operator $A(\mathcal{U})$). Then, it is easily seen that for a complex valued representative e we obtain a complex valued approximation of A_{ff} if defining the approximation by matching the action of A_{ff}^{-1} on this representative as in (4.1). In this case assumption (2.42) is meaningless.

Hence, there are several reasons why one may want to replace the choice of one approximation D to A_{ff} for both, interpolation and $\mathcal{F}$-relaxation by two different approximations D_s and D_P. Then $\mathcal{F}$-relaxation is given by

$$M = \omega \begin{pmatrix} D_s^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

and interpolation by

$$P = \begin{pmatrix} -D_P^{-1} A_{fc} \\ I \end{pmatrix}.$$

In what follows, we present a modified convergence analysis for this generalization of the adaptive reduction-based algebraic multigrid approach that uses D_P and D_s.

The following analysis uses the observation made in section 2.5 that the Galerkin operator with ideal interpolation P corresponds to the Schur complement $S_{cc}(A)$. In what follows, we relate the reduction-based definition of interpolation to the ideal interpolation via the relation of their Galerkin operators.

Proposition 4.1. *Let P be defined as in (2.38) with A given in the same permuted block-notation. Then for $F = \begin{pmatrix} I \\ 0 \end{pmatrix}$ we have*

$$\mathcal{R}(P) \perp_A \mathcal{R}(F).$$

Proof. For any $x, y \in \mathbb{C}^m$ we have

$$\begin{aligned} \langle Py, Fx \rangle_A &= \langle APy, Fx \rangle_2 \\ &= \langle \begin{pmatrix} 0 \\ S_{cc}(A) \end{pmatrix} y, Fx \rangle_2 \\ &= \langle \begin{pmatrix} 0 \\ S_{cc}(A) \end{pmatrix} y, \begin{pmatrix} I \\ 0 \end{pmatrix} x \rangle_2 = 0. \end{aligned}$$

□

As we know that any matrix of the 2×2 block form $\begin{pmatrix} I & W \\ 0 & I \end{pmatrix}$ has full rank, we are able to split $\mathbb{C}^m$ into the direct sum

$$V = \mathcal{R}(P) \oplus \mathcal{R}(F).$$

Definition 4.2 (Strengthened Cauchy-Bunyakowski-Schwarz Inequality)**.** *Given a direct decomposition of $V = \mathbb{C}^m$ into $V_1 \oplus V_2 = V$ and a hermitian positive definite linear operator $A : V \to V$, we define $\gamma^2 \in [0, 1)$ as the smallest constant such that*

$$\langle v_1, v_2 \rangle_A^2 \leq \gamma^2 \|v_1\|_A^2 \|v_2\|_A^2, \ v_1 \in V_1, v_2 \in V_2. \tag{4.2}$$

The constant γ^2 in Definition 4.2 can be interpreted as the cosine of the abstract angle between the subspaces V_1 and V_2. Furthermore, we can relate γ^2 to the spectral equivalence of $A_c = P^H A P$ for an arbitrary interpolation operator P and the Schur complement $S_{cc}(A)$ of Definition 2.39 using the fact that the Galerkin operator A_c of ideal interpolation is the Schur complement $S_{cc}(A)$.

For this purpose we first recall a variational property that is fulfilled by the Schur complement $S_{cc}(A)$ (cf. [Axe94, Theorem 3.8]).

Proposition 4.3. *Let* $x = \begin{pmatrix} x_f \\ x_c \end{pmatrix}$. *The Schur complement* $S_{cc}(A)$ *fulfills the variational principle*

$$\langle S_{cc}(A)x_c, x_c\rangle_2 = \inf_{x_f} \langle Ax, x\rangle_2.$$

Proof. As $\langle Ax, x\rangle_2 \geq 0$ for all $x \in \mathbb{C}^m$ we have

$$\langle A_{ff}x_f, x_f\rangle_2 = \langle A \begin{pmatrix} x_f \\ 0 \end{pmatrix}, \begin{pmatrix} x_f \\ 0 \end{pmatrix}\rangle_2 \geq 0.$$

Thus we get

$$\begin{aligned} \langle Ax, x\rangle_2 &= \langle \begin{pmatrix} A_{ff}x_f + A_{fc}x_c \\ A_{cf}x_f + A_{cc}x_c \end{pmatrix}, \begin{pmatrix} x_f \\ x_c \end{pmatrix}\rangle_2 \\ &= \langle A_{ff}\left(x_f + A_{ff}^{-1}A_{fc}x_c\right), x_f + A_{ff}^{-1}A_{fc}x_c\rangle_2 + \langle S_{cc}(A)x_c, x_c\rangle_2 \\ &\geq \langle S_{cc}(A)x_c, x_c\rangle_2. \end{aligned}$$

□

Now, the infimum is attained for $x_f = -A_{ff}^{-1}A_{fc}x_c$ which can also be seen as a harmonic extension from $\mathcal{C}$-variables to $\mathcal{F}$-variables coinciding with ideal interpolation.

Theorem 4.4. *Let* $S_{cc}(A)$ *be defined as in* (2.39) *and* γ^2 *given by* (4.2) *for* $V_1 = \mathcal{R}(F)$ *and* $V_2 = \mathcal{R}(P)$. *Then we have*

$$\left(1 - \gamma^2\right) P^H AP \leq S_{cc}(A) \leq P^H AP.$$

Proof. Given the subspaces $\mathcal{R}(F)$ and $\mathcal{R}(P)$ we can write the Schur complement as

$$\langle S_{cc}(A)x_c, x_c\rangle_2 = \inf_w \langle Fw + Px, Fw + Px\rangle_A.$$

Thus, assuming that (4.2) is fulfilled with $\gamma \in \mathbb{R}$ and writing $P^H AP = A_c$, we obtain for any real t

$$\begin{aligned} \langle S_{cc}(A)x, x\rangle_2 &= \inf_w \inf_t \langle tFw + Px, tFw + Px\rangle_2 \\ &= \inf_w \left(\langle A_c x, x\rangle_2 - \frac{\langle Fw, Px\rangle_A}{\langle Fw, Fw\rangle_A}\right) \\ &\geq \inf_w \left(\langle A_c x, x\rangle_2 - \gamma^2 \langle A_c x, x\rangle_2\right) \\ &= \left(1 - \gamma^2\right) \langle A_c x, x\rangle_2. \end{aligned}$$

□

With these observations we can bound $\|E_{tg}\|_A$ by means of Theorem 2.27.

Theorem 4.5. *Assume that interpolation P is defined, such that there exists $\gamma^2 \in [0,1)$ with*

$$\left(1-\gamma^2\right) P^H A P \leq S_{cc}(A) \leq A_c.$$

Furthermore, assume that there exist $c_1, c_2 \in \mathbb{R}$ such that the hermitian $\mathcal{F}$-relaxation M_{ff} fulfills

$$c_1 M_{ff} \leq A_{ff} \leq c_2 M_{ff}. \tag{4.3}$$

Then, we have

$$\|E_{tg}\|_A \leq 1 - \frac{1-\gamma^2}{\mu}, \quad \mu = \max\left\{\frac{1}{c_1\left(2-c_1\right)}, \frac{1}{c_2\left(2-c_2\right)}\right\}.$$

Proof. First, we observe that

$$c_1 M_{ff} \leq A_{ff} \leq c_2 M_{ff} \Rightarrow c_1 I \leq M_{ff}^{-1/2} A_{ff} M_{ff}^{-1/2} \leq c_2 I.$$

This in turn gives $(1-c_2)I \leq I - M_{ff}^{-1/2} A_{ff} M_{ff}^{-1/2} \leq (1-c_1)I$. It follows that

$$\rho(I - M_{ff}^{-1} A_{ff}) = \rho(I - M_{ff}^{-1/2} A_{ff} M_{ff}^{-1/2}) \leq \max\{|1-c_1|, |c_2-1|\} < 1, \tag{4.4}$$

showing that M_{ff} defines a convergent smoother for A_{ff}. This follows, since $I - M_{ff}A_{ff}$ is selfadjoint in $\langle ., . \rangle_{A_{ff}}$, implying

$$\rho(I - M_{ff}A_{ff}) = \|I - M_{ff}A_{ff}\|_{A_{ff}}.$$

In addition by using the identity of Theorem 2.27 (cf. [FVZ05, Theorem 4.2]) we obtain

$$\|E_{tg}\|_A \leq 1 - \frac{1}{K}, \quad K \leq \frac{1}{1-\gamma^2} \sup_{w \neq 0} \frac{\langle \widetilde{M}_{ff} w, w \rangle_2}{\langle A_{ff} w, w \rangle_2}, \tag{4.5}$$

with $\widetilde{M}_{ff} = M_{ff}^H \left(M_{ff}^H + M_{ff} - A_{ff}\right)^{-1} M_{ff}$ denoting the symmetrized version of M_{ff}. As $\sup_{w \neq 0} \frac{w^H \widetilde{M}_s w}{w^H A_{ff} w}$ is the smallest α for which $\widetilde{M}_{ff} \leq \alpha A_{ff}$ we obtain,

$$\begin{aligned}\widetilde{M}_{ff} \leq \tilde{\alpha} A_{ff} &\iff M_{ff}\left(2M_{ff} - A_{ff}\right)^{-1} M_{ff} \leq \tilde{\alpha} A_{ff} \\ &\iff A_{ff}^{-\frac{1}{2}} M_{ff} A_{ff}^{-\frac{1}{2}} \left(2A_{ff}^{-\frac{1}{2}} M_{ff} A_{ff}^{-\frac{1}{2}} - I\right)^{-1} A_{ff}^{-\frac{1}{2}} M_{ff} A_{ff}^{-\frac{1}{2}} \leq \tilde{\alpha} I.\end{aligned}$$

From (4.3) we have $\frac{1}{c_2} I \leq A_{ff}^{-\frac{1}{2}} M_{ff} A_{ff}^{-\frac{1}{2}} \leq \frac{1}{c_1} I$, thus,

$$\widetilde{M}_{ff} \leq \tilde{\alpha} A_{ff} \Longleftrightarrow \tilde{\alpha} \geq \max_{t \in \left[\frac{1}{c_2}, \frac{1}{c_1}\right]} \frac{t^2}{2t-1}. \tag{4.6}$$

Note, that $\frac{1}{c_2} > \frac{1}{2}$ by (4.4). Maximizing (4.6) with respect to t, we see that

$$\widetilde{M}_{ff} \leq \tilde{\alpha} A_{ff} \Longleftrightarrow \tilde{\alpha} \geq \max\left(\frac{1}{c_1\left(2-c_1\right)}, \frac{1}{c_2\left(2-c_2\right)}\right).$$

Thus,

$$\sup_w \frac{w^H \widetilde{M}_{ff} w}{w^H A_{ff} w} = \max\left(\frac{1}{c_1\left(2-c_1\right)}, \frac{1}{c_2\left(2-c_2\right)}\right) = \alpha \tag{4.7}$$

and we get

$$K \leq \frac{\alpha}{(1-\gamma^2)}.$$

Combining this with (4.5) we finally obtain

$$\|E_{tg}\|_A \leq 1 - \frac{1-\gamma^2}{\mu}.$$

□

Note, that assumption (4.3) on the $\mathcal{F}$-smoother means that it is a convergent iteration on the $\mathcal{F}$-equations and that there always exists a $\gamma^2 \in [0,1)$ due to Theorem 4.4. Hence, we can guarantee two-grid convergence in this framework. In the case that more specific knowledge about the spectral relationship of D_s and A_{ff} are given, we can minimize μ and get a more precise bound on two-grid convergence.

Corollary 4.6. *Let the hermitian and positive-definite matrix D_s be given and the assumptions of Theorem 4.5 be satisfied. Further, assume that there are positive constants, λ and Λ, such that $\lambda D_s \leq A_{ff} \leq \Lambda D_s$. Define the smoothing operator M as*

$$M = \omega \begin{pmatrix} D_s^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad \omega = \frac{2}{\Lambda + \lambda}.$$

Then,

$$\|E_{tg}\|_A \leq 1 - \frac{4\lambda\Lambda}{(\Lambda+\lambda)^2}\left(1-\gamma^2\right).$$

Proof. By the definition of M we have $c_1 = \frac{2\lambda}{\Lambda+\lambda}$ and $c_2 = \frac{2\Lambda}{\Lambda+\lambda}$, such that with (4.7) we obtain

$$\frac{1}{\alpha} = c_1 (2 - c_1) = c_1 c_2 = \frac{4\lambda\Lambda}{(\Lambda + \lambda)^2}.$$

□

Utilizing the analysis of [FVZ05], we can also derive an estimate for the convergence of a two-grid method that uses full smoothing, i.e., smoothing on both $\mathcal{F}$ and $\mathcal{C}$ variable sets, rather than smoothing only on $\mathcal{F}$ variables. Hence, this result also applies to Jacobi or Gauss-Seidel smoothing, which are of interest in a practical implementation of this method for hermitian positive definite operators A. In general, full smoothing is expected to yield faster convergence of the resulting method over using mere $\mathcal{F}$-relaxation.

To analyze this case, consider a two-grid method given by its error propagator,

$$E_{tg} = I - B_{tg}^{-1} A.$$

Generalizing the above approach, we can interpret the two-grid method with full smoothing in the same framework as it was used for the analysis of the $\mathcal{F}$-smoothing case. Instead of using a smoother M_{ff} that fulfills

$$\|I - M_{ff}^{-1} A_{ff}\|_{A_{ff}} \leq 1,$$

we consider a smoother M, such that

$$\|I - M^{-1} A\|_A \leq 1.$$

The two-grid preconditioner B_{tg}^{-1} can than be written as

$$B_{tg}^{-1} = \begin{bmatrix} I & P \end{bmatrix} \hat{B}_{tg}^{-1} \begin{bmatrix} I \\ P^H \end{bmatrix}$$

with

$$\hat{B}_{tg}^{-1} = \begin{bmatrix} I & -M^{-H} A P \\ 0 & I \end{bmatrix} \begin{bmatrix} \bar{M}^{-1} & 0 \\ 0 & A_c^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -P^H A M^{-1} & I \end{bmatrix}.$$

Assuming that $\widetilde{M}$ gives a convergent smoother for A with

$$A \leq \widetilde{M} \leq \kappa A$$

and that P is chosen such that

$$P^H A P \leq \nu S_{cc}(A),$$

we obtain by [FVZ05, Theorem 5.1] the bound

$$\|E_{tg}\|_A \leq 1 - \frac{1}{\nu\kappa}.$$

With the two-grid analysis complete, we proceed with a description of the adaptivity used in the modified reduction-based algebraic multigrid context.

4.1.1 Adaptivity in the modified AMGr framework

As proposed in [MMM06], we use an adaptive scheme to define D_P and, with it the interpolation operator, P_D. In order to keep the operators in the multigrid hierarchy sparse, we choose D_P to be diagonal. More specifically, we choose D_P such that D_P^{-1} matches the action of A_{ff}^{-1} on a given vector

$$u = \begin{pmatrix} u_f \\ u_c \end{pmatrix},$$

which represents algebraically smooth error, i.e., we require

$$-D_P^{-1} A_{fc} u_c = u_f = -A_{ff}^{-1} A_{fc} u_c$$

for a given u_c. The key issue to consider when attempting to design an efficient adaptive AMGr solver in this setting, is then reduced to development of an efficient scheme for computing the prototype, u, used to define D_P. The classical adaptive methods [BFM$^+$04, BFM$^+$05, BFM$^+$06] use repeated application of the given relaxation scheme (or the resulting solver) to compute (or improve) the prototype. In general, the two main drawbacks of this approach are that, first, there is no theoretically founded stopping criterion available for such an approach that guarantees its optimality; and, second, the number of setup iterations in such a classical adaptive process to compute a sufficiently accurate approximation of the prototype depends on the condition number of the matrix, i.e., the number of required setup iterations increases if the condition number increases, e.g., by increasing the number of variables in the discretization.

For the Gauge Laplace operators, the smallest eigenvector is often not a good local representative of the algebraically smooth error, which further compounds the difficulty of developing an adaptive scheme for this system. Thus, instead of the above αAMGr type of interpolation that is built to fit the smallest eigenvector, we consider developing a different approach that rather fits an appropriate linear combination of eigenvectors to small eigenvalues.

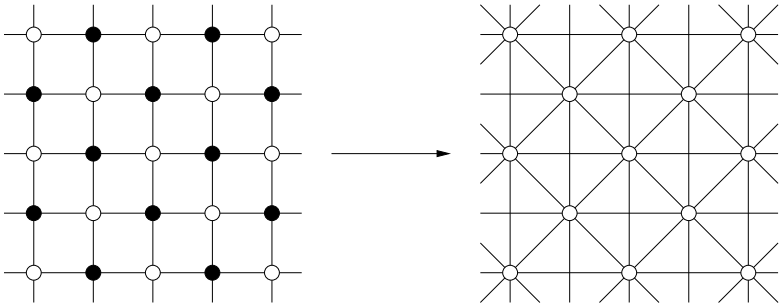


Figure 4.1: G_A before and after the "odd-even" reduction.

4.1.2 Numerical Results

In this section, we show the capabilities and limitations of the adaptive reduction-based algebraic multigrid approach for solving the Gauge Laplace systems described in chapter 3. Most of the results can also be found in [BFKM09].

As we limit the discussion in what follows to 5-point discretizations of the systems at hand, we introduce an "odd-even" reduction preprocessing step that reduces the number of variables by a factor of 2 yielding a 9-point Schur complement operator. Numbering the variables on the grid in an odd-even fashion we can write the permuted A as

$$A = \begin{pmatrix} I & A_{oe} \\ A_{eo} & I \end{pmatrix}.$$

The odd-even reduced operator A_{oo} is then given as the exact Schur complement of this variable splitting

$$A_{oo} = S_{oo}(A) = I - A_{oe}A_{eo}.$$

The transition of sparsity from A to A_{oo} is illustrated in Figure 4.1. Here circles with the same color are related to their odd or even numbering on the grid, respectively. Note, that given a solution for the odd-even reduced system A_{oo}, a solution for the original system A is known. For our numerical tests, we consider Gauge Laplace operators of varying size, mass, and temperature to test the reduction-based algebraic multigrid method. As a benchmark for later tests of our method applied to the Gauge Laplace system, we first consider the $\beta = \infty$ case with Dirichlet boundary conditions, which gives the standard 5-point discrete shifted-Laplace operator,

$$\mathcal{L} = -\Delta - (2\pi^2 - m)I,$$

obtained using a central-difference discretization. Here, the eigenvector associated with the smallest eigenvalue is known and has global support; specifically, this vector is the restriction of $\sin(\pi x)\sin(\pi y)$ to the grid points, and the smallest eigenvalue can be determined by the choice of shift, m. This problem was a first test case in the development of the original adaptive algebraic multigrid setup process [Mac04]. To illustrate the performance of the original adaptive process for such problems, we consider this problem with the shift chosen so that the system becomes increasingly ill-conditioned for fixed problem sizes. As the numerical results provided in Table 4.1 illustrate, such an adaptive setup procedure produces an effective solver for this model problem provided that a sufficient amount of work is done to expose the eigenvector associated with the smallest eigenvalue of the system matrix (i.e., a sufficient amount of work is done to ensure that the weak approximation property [Bra86, MR82] is satisfied by P, built using this computed vector for the given shift).

n_{rel} \ $\lambda_{\min}$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
5	.06	.02	.04	.37	.85	.98
25	.07	.02	.05	.05	.38	.86
50	.07	.02	.05	.05	.17	.66
100	.07	.02	.06	.06	.06	.16
500	.07	.02	.06	.06	.06	.06
exact	.07	.02	.06	.06	.06	.06

Table 4.1: Odd-even reduced 5-pt discretization of the Laplace operator with Dirichlet boundary conditions shifted to a fixed smallest eigenvalue. V(2, 2)-cycle asymptotic convergence rates with Gauss-Seidel smoother, using Gauss-Seidel relaxation applied to a positive random initial guess in the adaptive setup phase.

Next, we report the results of this original adaptive setup applied to a highly disordered system. The numerical results in Table 4.2 correspond to this scheme applied to a Gauge Laplace operators with randomly configured gauge field. Here, we take $\beta = 5$ and $N = 64$ and again vary the minimal eigenvalue and number of relaxations used to approximate the eigenvector corresponding to the smallest eigenvalue of the fine-level system. As the numerical results in Table 4.2 demonstrate, in contrast to the $\beta = \infty$ case, here increasing the number of relaxations used in the adaptive process eventually leads to degradation in performance of the resulting solver based on this single vector. Further, we see that this degradation is more severe in cases where the minimal eigenvalue is $O(10^{-3})$ or $O(10^{-4})$. This is consistent in all

tests, except for the last column where the minimal eigenvalue is shifted to be $O(10^{-6})$. In this case, using the exact eigenvector to the smallest eigenvalues does provide the best overall solver. This is to be expected as the weak approximation property implies that P must be able to reproduce this vector very accurately. Due to the local support of the eigenvectors to the smallest eigenvalues for this problem we see that using the minimal eigenvector is, in general, a suboptimal choice for the vector in the adaptive setup scheme. While each of these vectors is supported locally, their support does not, in general, overlap exactly. In such cases, a linear combination of these vectors may give a better approximation to the slow-to-converge vectors of the system matrix.

To test this approach, we consider an "artificial" adaptive process that uses a linear combination of the eigenvectors associated with the k smallest eigenvalues of the system matrix, weighted by the reciprocal of their eigenvalues, as the vector to be fit in the adaptive setup phase. We choose $k = 10$ as this gives good performance in our numerical tests. Results for this approach are shown in Table 4.2 in the line labeled "LC". Here, we see that the performance of the stand alone MG solver based on this approach is not, in general, better than that of the solver based on P defined using a prototype computed using relaxation. As the eigenvectors to the smallest eigenvalues can be local, using relaxation (or a linear combination of the eigenvectors to the ten smallest eigenvalues computed exactly) does not produce an AMGr-style P that satisfies the weak approximation property [Bra86, MR82], which requires accuracy in the computed prototype proportional to its Rayleigh quotient. However, both methods produce a P that leads to an effective variational MG preconditioner. The results in Table 4.3 are for various problem sizes and choices of β. Here, P is defined using the prototype computed by using relaxation and also by taking a linear combination of the eigenvectors to the ten smallest eigenvalues. As before, we see that both solvers perform well as a preconditioner for CG. Overall, our proposed AMGr-style method, based on a single prototype, is not expected to produce an optimal stand-alone solver for these systems. Our numerical results suggest that the approach does, however, have potential for dramatically improving CG performance for cases where the more expensive multiple-vector type adaptive methods (e.g., αSA) are not applicable. Even though the results for the adaptive reduction-based algebraic multigrid approach do not look bad for the Gauge Laplace system, especially as a preconditioner for a Krylov subspace iteration like CG, we felt that the nature of its adaptivity does not allow us to cope with the difficulties encountered in these problems with random background gauge fields. Herein especially the outlook on even more complex systems like the Wilson-Schwinger of QED and Wilson operator of Lattice-QCD con-

$n_{rel} \setminus \lambda_{\min}$	10^{-1}		10^{-2}		10^{-3}		10^{-4}		10^{-6}	
5	.4	(9)	.79	(15)	.97	(19)	.99	(21)	.99	(25)
25	.32	(9)	.53	(11)	.83	(14)	.98	(15)	.99	(18)
50	.31	(8)	.55	(11)	.72	(12)	.95	(14)	.99	(17)
100	.28	(8)	.52	(10)	.65	(13)	.9	(14)	.99	(17)
300	.32	(8)	.48	(10)	.53	(10)	.54	(10)	.89	(13)
500	.33	(8)	.5	(10)	.6	(11)	.6	(11)	.62	(11)
exact	.31	(8)	.53	(10)	.61	(12)	.61	(11)	.62	(12)
LC	.35	(8)	.43	(9)	.67	(11)	.67	(12)	.62	(12)
CG	*	(44)	*	(75)	*	(107)	*	(231)	*	(435)

Table 4.2: Odd-even reduced Gauge Laplace operator with periodic boundary conditions shifted to a fixed smallest eigenvalue. V(2, 2)-cycle asymptotic convergence rates with Gauss-Seidel smoother, using Gauss-Seidel applied to a complex-valued random initial guess in the adaptive setup phase. In parentheses, we report the iteration count for preconditioned CG to reduce the initial residual by a relative factor of 10^8. For the line labeled "LC", a linear combination of the eigenvectors associated with the ten smallest eigenvalues of the system matrix, weighted by the reciprocal of their eigenvalues, as the vector to be fit in the adaptive setup phase. The line labeled CG contains iteration counts of the Conjugate Gradient method applied to this system as a stand-alone solver; again the (relative) residual is reduced to 10^{-8} in these tests.

β \ N	32	64	128	256
1	11 / 12	10 / 14	15 / 15	11 / 14
5	12 / 15	11/ 15	15 / 15	14 / 16
10	7 / 11	13 / 15	17 / 16	19 / 17

Table 4.3: Odd-even reduced Gauge Laplace operators of various sizes and temperatures β, shifted to have smallest eigenvalue $\frac{1}{N^2}$. AMGr 2-level V(2, 2) preconditioner with Gauss-Seidel smoothing for CG using both a linear combination of the eigenvectors to the ten smallest eigenvalues, scaled by their associated inverse Rayleigh quotients to define P (shown first) as well as using relaxation to define the prototype in the definition of interpolation (shown second).

vinced us to keep on looking for other adaptive processes that are capable to resolve the problems we encountered with the reduction-based approach. In the next section, we present an adaptive approach to define interpolation that tears down the limits of adaptive algebraic multigrid found here and

and also provides a way to compute representatives of algebraically smooth error in a multiscale fashion.

.

4.2 Least Squares AMG Interpolation

As discussed in section 4.1, in many important applications several vectors are needed in order to define effective multigrid interpolation operators. While there is a framework in adaptive smoothed aggregation [BFM+04] that is able to overcome this problem, the setup it uses computes the necessary vectors sequentially and thus can be impracticably expensive.

In contrast to the smoothed aggregation framework, Brandt proposed a least squares formulation of interpolation in [Bra02]. In this section we explore this idea and develop a precise mathematical formulation. Later, in chapter 5, we present results that demonstrate its potential to solve the applications discussed in chapter 3.

Assuming that the set of variables Ω is partitioned into a subset of coarse variables $\mathcal{C}$ and the set of $\mathcal{F}$ variables, the task of defining interpolation in algebraic multigrid is reduced to the task of defining interpolation weights p_{ij} for each $i \in \mathcal{F}$.

Definition 4.7. *Let $\Omega = \mathcal{C} \cup \mathcal{F}$. For each $i \in \mathcal{F}$ define the set of interpolatory points $\mathcal{C}_i$ as the set*

$$\mathcal{C}_i = \{j \in \mathcal{C},\ p_{ij} \neq 0\}.$$

The least squares approach of interpolation is a way to (adaptively) define interpolation weights p_{ij}. It can also be further used to define appropriate sets of interpolatory points $\mathcal{C}_i$ within the bootstrap setup process.

The basic idea of using a least squares fit to define interpolation is very simple.

Definition 4.8. *For $i \in \mathcal{F}$, given a set of interpolatory points $\mathcal{C}_i$ and a set of linearly independent vectors $u^{(1)}, \ldots, u^{(k)}$, we define the interpolation weights $p_i = (p_{ij})_{j \in \mathcal{C}_i}$ as the minimizer of the least squares functional*

$$\mathcal{L}(p_i) = \sum_l \omega_l (u_i^{(l)} - p_i R u^{(l)})^2 \rightarrow min, \tag{4.8}$$

where R is a restriction operator on the coarse variables, such that $RP = I$ and ω_l are weights used in the fit that are applied to weight each vector, to be discussed later on. In the following the resulting interpolation is called least squares *interpolation or LS interpolation.*

Remark 4.9. *If a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ is given and interpolation is given as the identity on the $\mathcal{C}$ set of variables, a simple choice of R in Definition 4.8, is*

$$R = \begin{pmatrix} 0 & I \end{pmatrix}.$$

In the following, if not mentioned explicitly, we assume that R has this form. Our discussion of least squares interpolation focuses on the classical choice of coarse grid variables $\mathcal{C} \subset \Omega$. In this case, the weighted least squares problem (4.8) *can be written as*

$$\mathcal{L}(p_i) = \sum_l \omega_l (u_i^{(l)} - \sum_{j \in \mathcal{C}_i} p_{ij} u_j^{(l)})^2 \to \min. \tag{4.9}$$

In what follows, we make extensive use of canonical injections and restrictions.

Definition 4.10. *Given a set of variables $\Omega = \{1, \ldots, m\}$ and a subset $\Omega_i \subset \Omega$ denote the canonical injection of Ω_i into Ω by $I_{|\Omega_i}$, i.e., as a restriction of the identity on the set of variables Ω_i. Further, denote the canonical restriction $I_{|\Omega_i}^T v$ of a vector $v \in \mathbb{C}^m$, i.e., with variables according to Ω, onto its variables belonging to Ω_i by v_{Ω_i}.*

With this notation, R as given in Remark 4.9, can be written as

$$R = I_{|\mathcal{C}}^T.$$

Now, we can formulate the least squares problem (4.9) in a classical linear algebra context as follows. Introducing

$$W = \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_k \end{pmatrix} \in \mathbb{R}^{k \times k} \quad \text{and} \quad U = \begin{pmatrix} u^{(1)} & \cdots & u^{(k)} \end{pmatrix} \in \mathbb{C}^{m \times k},$$

we have

$$\sum_l \omega_l (u_i^{(l)} - \sum_{j \in \mathcal{C}_i} p_{ij} u_j^{(l)})^2 = \| U_{\{i\}} W^{\frac{1}{2}} - p_i U_{\mathcal{C}_i} W^{\frac{1}{2}} \|_2^2.$$

That is, the weights of the individual terms in the sum of (4.9) are reflected by scaling, i.e., weighting, the test vectors instead. This in turn yields a least squares problem as

$$\mathcal{L}(p_i) = \| U_{\{i\}} W^{\frac{1}{2}} - p_i U_{\mathcal{C}_i} W^{\frac{1}{2}} \|_2^2 \to \min. \tag{4.10}$$

In this formulation it is easy to write down the minimizer and necessary conditions for its uniqueness.

Lemma 4.11. *Assuming that $U_{\mathcal{C}_i} W^{\frac{1}{2}}$ has full-rank, p_i is uniquely determined by*

$$p_i = U_{\{i\}} W U_{\mathcal{C}_i}^H \left(U_{\mathcal{C}_i} W U_{\mathcal{C}_i}^H \right)^{-1}.$$

Proof. The derivative of the least squares functional $\mathcal{L}$ with respect to p_{ij} is

$$\frac{\partial}{\partial p_{ij}} \mathcal{L}(p_i) = 2 \sum_l \omega_l \left(u_i^{(l)} - p_i \left(u_{\mathcal{C}_i}^{(l)} \right) \right) \left(u_{\mathcal{C}_i}^{(l)} \right)_j .$$

A little algebra shows that $\nabla \mathcal{L}(p_i) = 0$ yields

$$p_i U_{\mathcal{C}_i} W U_{\mathcal{C}_i}^H = U_{\{i\}} W U_{\mathcal{C}_i}^H.$$

With $\operatorname{rank}(U_{\mathcal{C}_i} W^{\frac{1}{2}}) = |\mathcal{C}_i|$, we obtain that $U_{\mathcal{C}_i} W U_{\mathcal{C}_i}^H$ is non-singular. Thus, we have

$$p_i = U_{\{i\}} W U_{\mathcal{C}_i}^H \left(U_{\mathcal{C}_i} W U_{\mathcal{C}_i}^H \right)^{-1}.$$

□

Note that the derivation of the unique minimizer of (4.10) is also easily seen by forming the normal equations that are equivalent to the derivation in Lemma 4.11.

Remark 4.12. *Due to Lemma 4.11 the restricted vectors $U_{\mathcal{C}_i}$ have to form a generating system of the local linear space in order to get a unique minimizer for $i \in \mathcal{F}$. Furthermore, this also implies a lower bound on the number k of vectors used in the least squares fit, that is*

$$k \geq \max_{i \in \mathcal{F}} |\mathcal{C}_i|$$

to obtain a unique least squares interpolation operator P.

The vectors used in the definition of least squares interpolation are further referred to as test vectors. The test vectors appearing in this variational definition of interpolation should, according to the fundamental principle of the complementarity of smoothing and coarse-grid correction, represent algebraically smooth error. We discuss in section 4.3 how such representatives can be computed by exploiting the multigrid hierarchy itself.

The least squares fit of the interpolation weights p_{ij} are computed as the best weights used to fit the set of test vectors $\{u^{(1)}, \ldots, u^{(k)}\}$ in a weighted least squares sense. It is a natural requirement that the weights, ω_l, are defined according to the "smoothness" of $u^{(l)}$. In the following we present some possible choices of weights that fulfill this requirement.

Assuming that the algebraically smooth vectors associated with the given smoother S are also the eigenvectors with small eigenvalues of A, the choice

$$\omega_l^{\frac{1}{2}} = \frac{\langle u^{(l)}, u^{(l)} \rangle_2}{\langle u^{(l)}, u^{(l)} \rangle_A}, \tag{4.11}$$

i.e., the inverse Rayleigh-Quotient, then biases the least-square fit towards the algebraically smoothest vectors. This assumption holds for Richardson and ω-Jacobi, but the situations for Gauss-Seidel and other smoothers is intricate and often not provable. However for many problems numerical evidence suggests that this simplifying assumption is appropriate.

Taking into account a more accurate definition of algebraic smoothness with respect to the smoother, the two-grid analysis in [FVZ05] suggests to use

$$\omega_l^{\frac{1}{2}} = \frac{\langle u^{(l)}, u^{(l)} \rangle_2}{\langle u^{(l)}, u^{(l)} \rangle_{\tilde{S}}}$$

instead, with $\tilde{S}$ defined as in (2.22). As such formulation takes more explicit account for the smoother.

There is at least one other choice of weights that is particularly interesting. Denoting the residual of $u^{(l)}$ as $r^{(l)} = Au^{(l)}$ we can define the weights

$$\omega_l^{\frac{1}{2}} = \frac{\langle u^{(l)}, u^{(l)} \rangle_2^{\frac{1}{2}}}{\langle Au^{(l)}, Au^{(l)} \rangle_2^{\frac{1}{2}}} = \frac{\|u^{(l)}\|_2}{\|r^{(l)}\|_2}. \tag{4.12}$$

As before, denoting the interpolatory window for $i \in \mathcal{F}$ by w_i we can localize this particular form of weighting as follows

$$\left(\omega_l^i\right)^{\frac{1}{2}} = \frac{\|I_{|w_i}^T u^{(l)}\|_2}{\|I_{|w_i}^T r^{(l)}\|_2} = \frac{\|I_{|w_i}^T u^{(l)}\|_2}{\langle A^H I_{|w_i} I_{|w_i}^T Au^{(l)}, u^{(l)} \rangle_2^{\frac{1}{2}}}.$$

Remark 4.13. *There are some noteworthy observations to the choices of weights in* (4.11) *and* (4.12).

- *Weighting according to* (4.12) *is equivalent to*

$$\omega_l^{\frac{1}{2}} = \frac{\langle u^{(l)}, u^{(l)} \rangle_2^{\frac{1}{2}}}{\langle A^H Au^{(l)}, u^{(l)} \rangle_2^{\frac{1}{2}}}.$$

  In words, this choice of weighting is the square root of the inverse Rayleigh-Quotient for the normal equations $A^H A$.

- *For an eigenvector v of A with corresponding eigenvalue λ the weight in* (4.12) *is $|\lambda|^{-1}$, which is the same weight one gets in* (4.11) *if A is hermitian positive definite.*

- *Choosing weights according to* (4.12) *is straight-forwardly applicable to non-hermitian and indefinite operators A, whereas the weighting defined in* (4.11) *does not reflect the spectral properties the weighting process is intended to account for.*

Note, that in the case that A is singular with its kernel $\mathcal{N}(A)$ spanned by $v^{(1)}, \ldots, v^{(k_0)}$ multigrid theory, as reviewed in section 2.2, requires that these vectors are exactly represented in $\mathcal{R}(P)$, i.e.,

$$\text{There exists a } v_c^{(i)} \in \mathbb{C}^{m_c}, \quad \text{such that} \quad v^{(i)} = P v_c^{(i)}. \tag{4.13}$$

Lemma 4.14. *Let A be a singular hermitian positive semi-definite matrix with its kernel $\mathcal{N}(A)$ spanned by $v^{(1)}, \ldots, v^{(k_0)}$. Assume that $v^{(1)}, \ldots, v^{(k_0)}$ are contained in the set of test vectors $u^{(1)}, \ldots, u^{(k)}, k \geq k_0$. Further, assume that for each $i \in \mathcal{F}$ the set of interpolatory points $\mathcal{C}_i$ fulfills $|\mathcal{C}_i| \geq k_0$. Then least squares interpolation with weights defined according to* (4.11) *or* (4.12) *fulfills* (4.13)*.*

Proof. Both (4.11) and (4.12) yield the weight

$$\omega_l = \infty \quad \text{if} \quad u^l \in \mathcal{N}(A).$$

Hence, these test vectors enter the the least squares fit (4.8) as fixed constraints, i.e., they have to be fulfilled exactly in order to get a finite fit value (4.8). As $|\mathcal{C}_i| \geq k_0$, it is possible to fulfill all kernel vector constraints, i.e., the resulting interpolation fulfills (4.13). □

Here Lemma 4.14 holds if the weighting process chosen in the definition of least squares interpolation yields infinite weight for vectors in $\mathcal{N}(A)$ and $|\mathcal{C}_i| \geq k_0$ for all $i \in \mathcal{F}$.

4.2.1 Generalized Least Squares Interpolation

As observed in chapter 2, many variants of algebraic multigrid can be described by the AMG$\not\in$ framework, introduced in [HV01], by using an appropriate definition of an extension map.

In a similar way least squares interpolation can be generalized to the AMG$\not\in$ framework. Based on an idea due to Falgout [Fal02], the extension map E appearing in AMG$\not\in$ interpolation (cf. section 2.3) can be defined by a least squares fit.

Below we summarize some of our ideas and observations regarding this approach, show that the approach reduces to the Classical AMG formulation in a special case, and demonstrate how this can be interpreted as a direct least squares fit similar to (4.8) by using the observation that for any error e we have the residual equality $r = Ae$.

We start by recalling the basics of AMG$\not e$ interpolation as introduced in section 2.3.

Assuming that a $\mathcal{C}$-$\mathcal{F}$ partitioning of variables is given, the definition of interpolation weights in the AMG$\not e$ context proceeds by construction of certain local matrices A_i for each grid point $i \in \mathcal{F}$ (including possibly several unknowns associated with this grid point) that correspond to a fine-grid neighborhood of point i and its interpolatory set $\mathcal{C}_i$.

The local matrices are built using extension mappings. With the notation of $\mathcal{C}_\mathcal{I}$ and $\chi_\mathcal{I}$ of Definition 2.36 for any subset $\mathcal{I} \subset \mathcal{F}$, we have the permuted matrix

$$A = \begin{pmatrix} A_{ff} & A_{fc} & A_{f\chi} & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}.$$

Then, the AMG$\not e$ extension map is defined by

$$\begin{aligned} E : \mathcal{I} \cup \mathcal{C}_\mathcal{I} &\to \mathcal{I} \cup \mathcal{C}_\mathcal{I} \cup \chi_\mathcal{I} \\ \begin{pmatrix} v_f \\ v_c \end{pmatrix} &\to \begin{pmatrix} v_f \\ v_c \\ E_{\chi f} v_f + E_{\chi c} v_c \end{pmatrix}. \end{aligned}$$

Here, E couples unknowns from $\mathcal{I} \cup \mathcal{C}_\mathcal{I}$ to unknowns in $\chi_\mathcal{I}$ and the local matrices $A_\mathcal{I}$ are defined by

$$\begin{aligned} A_\mathcal{I} &= \begin{pmatrix} I & 0 & E_{\chi f}^H \\ 0 & I & E_{\chi c}^H \end{pmatrix} A_{\mathcal{I} \cup \mathcal{C}_\mathcal{I} \cup \chi_\mathcal{I}} \begin{pmatrix} I & 0 \\ 0 & I \\ E_{\chi f} & E_{\chi c} \end{pmatrix} \\ &= \begin{pmatrix} A_{ff} + A_{f\chi} E_{\chi f} & A_{fc} + A_{f\chi} E_{\chi c} \\ * & * \end{pmatrix}. \end{aligned}$$

Given the local matrix A_i, interpolation is then given by the local *harmonic extension* corresponding to $i \in \mathcal{I}$

$$p_i = -\left(\hat{A}_{ff}^{-1} \hat{A}_{fc} \right)_{.,i}$$

with $\hat{A}_{ff} = A_{ff} + A_{f\chi} E_{\chi f}$ and $\hat{A}_{fc} = A_{fc} + A_{f\chi} E_{\chi c}$.

The general idea of using least squares interpolation in this setting is to define E by a minimization over several test vectors $u^{(1)}, \ldots, u^{(k)}$ that represent algebraically smooth error in a least squares sense:

$$\sum_l \omega_l \| u_\chi^{(l)} - \begin{pmatrix} E_{\chi f} & E_{\chi c} \end{pmatrix} \begin{pmatrix} u_f^{(l)} \\ u_c^{(l)} \end{pmatrix} \|^2 \to \min_E \ . \tag{4.14}$$

The minimizing extension map is then used to define interpolation in the standard AMG$\not e$ sense. Again, the weights ω_l should bias the least squares fit towards locally algebraically smooth error. Ideally one could think about using a local weighting process that involves the local matrix $A_\mathcal{I}$. However, since $A_\mathcal{I}$ is unknown a priori, this results in a nonlinear minimization problem. For now we assume the simpler choice of ω_l as defined in (4.11) or (4.12).

We note that, the interpolation only requires the product $A_{f\chi}E$. It is thus sufficient to formulate a least squares minimization for

$$A_{f\chi}E = \begin{pmatrix} A_{f\chi}E_{\chi f} & A_{f\chi}E_{\chi c} \end{pmatrix} = \begin{pmatrix} w_f & w_c \end{pmatrix} = w,$$

whose minimizer yields the same interpolation operator as the minimizer of (4.14). This least squares problem for w is then given by

$$\sum_l \omega_l \| A_{f\chi} u_\chi^{(l)} - \begin{pmatrix} w_f & w_c \end{pmatrix} \begin{pmatrix} u_f^{(l)} \\ u_c^{(l)} \end{pmatrix} \|^2 \to \min_w, \tag{4.15}$$

as the following Lemma shows.

Lemma 4.15. *Assume that* $k > |\mathcal{I}| + |\mathcal{C}_\mathcal{I}|$ *and that* $U = \begin{pmatrix} u_f^{(1)} & \ldots & u_f^{(k)} \\ u_c^{(1)} & \ldots & u_c^{(k)} \end{pmatrix}$ *has full rank. Then the linear least squares problems* (4.14) *and* (4.15) *have unique solutions* $\hat{E}$ *and* $\hat{w}$ *and*

$$\hat{w} = A_{f\chi}\hat{E} \ .$$

Proof. With $b = \begin{pmatrix} u_\chi^{(1)} & \ldots & u_\chi^{(k)} \end{pmatrix}$ the solution of (4.14) fulfills the normal equation given by

$$bU^H = EUU^H. \tag{4.16}$$

Since U has full rank, $|\mathcal{I}| + |\mathcal{C}_\mathcal{I}|$, the linear operator of the normal equations has full rank such that (4.16) has a unique solution

$$\hat{E} = bU^H \left(UU^H\right)^{-1} \ .$$

Analogously, we find the unique solution of (4.15) as

$$\hat{w} = A_{f\chi} b U^H \left(UU^H\right)^{-1} = A_{f\chi}\hat{E} \,.$$

□

Now, by using the residual equation $r = Au$ for the homogeneous system $Au = 0$, we further reformulate the least squares fit.

Lemma 4.16. *With the residual equation*

$$r_f = (Au)_f = A_{ff}u_f + A_{fc}u_c + A_{f\chi}u_\chi$$

for any $u \in \mathbb{C}^m$ and the definitions $\hat{A}_{ff} = A_{ff} + A_{f\chi}E_{\chi f}$ and $\hat{A}_{fc} = A_{fc} + A_{f\chi}E_{\chi c}$, with $\|\hat{A}_{ff}\|_2 = c < \infty$ for all $\mathcal{I} \subset \mathcal{F}$, we obtain a bound on (4.15) *by minimizing a modified least squares problem*

$$\sum_l \omega_l \|u_f^{(l)} - \hat{A}_{ff}^{-1} r_f^{(l)} + \hat{A}_{ff}^{-1}\hat{A}_{fc}u_c^{(l)}\|^2 \to \min_E \,. \tag{4.17}$$

Proof. Simple algebraic manipulation yields

$$\begin{aligned}
& \sum_l \omega_l \|A_{f\chi}u_\chi^{(l)} - A_{f\chi}E_{\chi c}u_c^{(l)} - A_{f\chi}E_{\chi f}u_f^{(l)}\|^2 \\
= & \sum_l \omega_l \|r_f^{(l)} - A_{ff}u_f^{(l)} - A_{fc}u_c^{(l)} - A_{f\chi}E_{\chi f}u_f^{(l)} - A_{f\chi}E_{\chi c}u_c^{(l)}\|^2 \\
= & \sum_l \omega_l \|\hat{A}_{ff}u_f^{(l)} - r_f^{(l)} + \hat{A}_{fc}u_c^{(l)}\|^2 \\
\leq & \|\hat{A}_{ff}\|_2^2 \sum_l \omega_l \|u_f^{(l)} - \hat{A}_{ff}^{-1} r_f^{(l)} + \hat{A}_{ff}^{-1}\hat{A}_{fc}u_c^{(l)}\|^2 .
\end{aligned}$$

The boundedness of $\|\hat{A}_{ff}\|_2^2$ yields the desired result. □

Remark 4.17. *The inclusion of the residual term $\hat{A}_{ff}^{-1} r_f^{(l)}$ can be interpreted as a local relaxation that enforces the equation to be fulfilled exactly in the local neighborhood. Brandt mentioned the use of local relaxation to enforce zero residuals in [Bra02].*

To simplify (4.17) even further we can use a result in [Fal02, HV01]. Therein, it is shown that we can omit connections from points in $\mathcal{I}$ to points in $\chi_\mathcal{I}$, represented by $E_{\chi f}$, that is, we can take $E_{\chi f} = 0$.

Lemma 4.18. *Given* $E = \begin{pmatrix} E_{\chi f} & E_{\chi c} \end{pmatrix}$ *and* $\tilde{E} = \begin{pmatrix} 0 & \tilde{E}_{\chi c} \end{pmatrix}$ *with*

$$\tilde{E}_{\chi c} = E_{\chi f} \left(-\hat{A}_{ff}^{-1} \hat{A}_{fc} \right) + E_{\chi c}$$

we have

$$-\hat{A}_{ff}^{-1} \hat{A}_{fc} = -A_{ff}^{-1} \left(A_{fc} + A_{f\chi} \tilde{E}_{\chi c} \right).$$

Proof. Direct computation yields the desired result. □

In words, extension maps E and $\tilde{E}$ of Lemma 4.18 give the same interpolation weights. Hence, we can further manipulate the least squares fit by taking $\tilde{E}_{\chi f} = 0$, yielding the new map $\tilde{E}$

$$\begin{aligned} & \sum_l \omega_l \| u_f^{(l)} - \hat{A}_{ff}^{-1} r_f^{(l)} + \hat{A}_{ff}^{-1} \hat{A}_{fc} u_c^{(l)} \|^2 \to \min_E \\ \iff & \sum_l \omega_l \| u_f^{(l)} - A_{ff}^{-1} r_f^{(l)} + A_{ff}^{-1} \left(A_{fc} + A_{f\chi} \tilde{E}_{\chi c} \right) u_c^{(l)} \|^2 \to \min_{\tilde{E}}. \end{aligned}$$

Using these results, we are able to see that the AMGe least squares interpolation is equivalent to a modified direct least squares interpolation.

Since the block-row P_{fc} of interpolation P is given by

$$P_{fc} = -A_{ff}^{-1} \left(A_{fc} + A_{f\chi} \tilde{E}_{\chi c} \right),$$

we obtain

$$\begin{aligned} & \sum_l \omega_l \| u_f^{(l)} - A_{ff}^{-1} r_f^{(l)} + A_{ff}^{-1} \left(A_{fc} + A_{f\chi} \tilde{E}_{\chi c} \right) u_c^{(l)} \|^2 \to \min_{\tilde{E}} \\ \iff & \sum_l \omega_l \| u_f^{(l)} - A_{ff}^{-1} r_f^{(l)} - P_{fc} u_c^{(l)} \|^2 \to \min_{P_{fc}}. \end{aligned} \tag{4.18}$$

Remark 4.19. *Some interesting observations can be made utilizing* (4.18).

1. *Assume* $r_f^{(l)} = 0$ *for* $l = 1, \ldots, k$ *and* $\mathcal{I} = \{i\}$ *for all* $i \in \mathcal{F}$, *then* (4.18) *reduces to* (4.8).

2. *For* $\mathcal{I} = \{i\}$ *and* $a_{ii} \neq 0, i \in \mathcal{F}$, *we obtain a modified least squares fit similar to* (4.10) *that includes the residual correction*

$$\mathcal{L}(p_i) = \| U_{\{i\}} W^{\frac{1}{2}} - \frac{1}{a_{ii}} (AU)_{\{i\}} W^{\frac{1}{2}} - p_i U_{\mathcal{C}_i} W^{\frac{1}{2}} \|_2^2 \to \min. \tag{4.19}$$

3. *Formulation* (4.18) *allows one to define least squares interpolation in situations where there exists a natural blocking of variables, e.g., in systems of partial differential equations, where several variables of the discretization are assumed to be at the same physical location, i.e., at the same grid point in the artificial grid given by the graph of the matrix* G_A.

4. *The AMG¢ formulation of least squares interpolation also allows one to generalize the notion of algebraic distance to systems of PDEs. Algebraic distance is discussed in detail in section 4.2.2.*

5. *Moreover, the definition of interpolation weights in a least squares framework allows one to analyze the given system of equations in order to find appropriate interpolatory points and thus generalizes the notion of "strength-of-connection" used in classical AMG as defined in section 2.4 to an adaptive framework. This can be done by using the least squares functional* $\mathcal{L}$.

4.2.2 Choice of $\mathcal{C}_i$ – Algebraic Distances

As mentioned in the introduction of section 4.2, the task of building interpolation in algebraic multigrid, assuming that a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ is given, is not only to find appropriate interpolation weights, but also to define a good set of interpolatory points $\mathcal{C}_i$ for each grid point $i \in \mathcal{F}$. The least squares framework of computing interpolation can be used to guide the choice of interpolatory points, as we discuss in the following section.

Definition 4.20. *We say that interpolation* P *has* caliber γ *if*

$$|\{j \in \mathcal{C} : p_{ij} \neq 0\}| = |\mathcal{C}_i| \leq \gamma, \quad \textit{for all } i \in \mathcal{F}.$$

First, recall that in classical AMG for the M-matrix case the set of interpolatory points is defined by using the notion of "strength-of-connection".

Definition 4.21. *Variable* $j \in \Omega$ *is strongly connected to variable* $i \in \Omega$ *if*

$$-a_{ij} > \theta \max_{k \neq i} \left(-a_{ik}\right),$$

for $\theta \in (0, 1)$ *fixed.*

This definition is based on the assumption that algebraically smooth error varies slowly along strong connections and that it is best represented using

strongly connected variables. Hence, the set $\mathcal{C}_i$ is defined as a subset of the strongly connected neighbors (in the graph neighborhood) of $i \in \mathcal{F}$.

More generally, as mentioned in [BL04, BL09], one could use the least squares functional (4.9) in order to generalize the notion of two variables as algebraically close if it is possible to represent algebraically smooth error along their connection, i.e., they allow a mutual representation of algebraically smooth error by one-another.

Definition 4.22. *Given a set of test vectors $u^{(1)}, \ldots, u^{(k)}$ and $i, j \in \Omega$ we define the* algebraic distance $d^{\alpha}_{i \leftarrow j}$ *from j to i as*

$$d^{\alpha}_{i \leftarrow j} = \min_{\delta_{ij}} \sum_{l} \omega_l \left(u_i^{(l)} - \delta_{ij} u_j^{(l)} \right)^2, \tag{4.20}$$

with weights ω_l defined as in (4.11) *or* (4.12).

Remark 4.23. *Note that,* (4.20) *coincides with one-sided least squares interpolation* (4.9) *with $\mathcal{C}_i = \{j\}$, i.e., interpolation from j to i. Again, we could modify Definition 4.22 by adding the residual term if $a_{ii} \neq 0$ for $i \in \Omega$. In that case algebraic distance reads as*

$$d^{\alpha}_{i \leftarrow j} = \min_{\delta_{ij}} \sum_{l} \omega_l \left(u_i^{(l)} - \frac{1}{a_{ii}} r_i^{(l)} - \delta_{ij} u_j^{(l)} \right)^2,$$

with $r^{(l)} = Au^{(l)}$.

Computing $d^{\alpha}_{i \leftarrow j}$ for all $(i, j) \in \Omega^2$ is prohibitive, but as we are only interested in local interpolation the computation of algebraic distance can be to the near graph neighborhood of i. An obvious way to define $\mathcal{C}_i$ using algebraic distance would involve only the algebraically closest points to i. On the other hand, one could choose $\mathcal{C}_i$ as the subset of a graph neighborhood of i with $|\mathcal{C}_i| \leq \gamma$ that yields the minimal least squares functional (4.9). This approach would require to compute least squares interpolation for any combination of subsets. We choose to introduce a compromise between the two approaches. First mentioned by Brandt and Livne in [BL04] we define a *greedy strategy* to build $\mathcal{C}_i$ and simultaneously least squares interpolation to $i \in \mathcal{F}$.

As formulated in Algorithm 3, the greedy strategy builds $\mathcal{C}_i$ one point at a time. Starting with the variable $j \in \mathcal{N}_i \subset \mathcal{C}$ with smallest algebraic distance, we add the variable $k \in \mathcal{N}_i$ that yields the best least squares fit in each iteration. The method terminates once a prescribed accuracy or the caliber of interpolation, i.e., the maximal number of interpolatory points, is reached.

Algorithm 3 gr_ls_interpolation {Computes Least Squares Interpolation}

Input: $A, \mathcal{U} = \{u^{(1)}, \ldots, u^{(k)}\}, \gamma, c$
Output: P
$U^{(l)} = Ru^{(l)},\ u^{(l)} \in \mathcal{U}$
$\omega_l^{\frac{1}{2}} = \frac{\langle u^l, u^l \rangle}{\langle Au^l, u^l \rangle},\ u^{(l)} \in \mathcal{U}$
for $i \in \mathcal{F}$ **do**
  Let $\mathcal{N}_i \subset \mathcal{C}$ neighboring C-points of i
  Set $\mathcal{C}_i = \emptyset$
  while $\mathcal{L}(P_i) > \gamma$ and $|\mathcal{J}_i| < c$ and $\mathcal{N}_i \neq \emptyset$ **do**
    Find $g^* = \underset{g \in \mathcal{N}_i}{\operatorname{argmin}} \left(\sum_{l=1}^{k} \omega_l \left(u_i^{(l)} - \sum_{j \in \{\mathcal{C}_i \cup g\}} p_{ij} U_j^{(l)} \right)^2 \right)$
    Set $\mathcal{N}_i = \mathcal{N}_i \setminus g^*$ and $\mathcal{C}_i = \mathcal{C}_i \cup g^*$
  end while
end for

Although it cannot be guaranteed that the best set $\mathcal{C}_i$ with $|\mathcal{C}_i| \leq \gamma$ is found by the greedy strategy it obviously yields improving least squares fits. Moreover, using a QR-decomposition of the operator involved in the least squares fit adding and removing points to $\mathcal{C}_i$ can be done in an update/downdate scheme of the QR-decomposition (cf. [GVL89, section 12.6]).

The least squares functional can further be used to control accuracy and sparsity of interpolation and hence the resulting algebraic multigrid hierarchy, assuming that we use the Galerkin definition of coarse-grid operators. We present a thorough experimental study of the potential to control accuracy and sparsity by the least squares functional in chapter 5.

4.2.3 Least Squares Interpolation – Theory

In order to gain deeper understanding of least squares interpolation, we analyze the method in a two-grid framework. In what follows, we assume that A is hermitian positive definite. Our purpose is to discuss the assumptions that lead to a weak approximation property

$$\| (I - PR)\, e\|_2^2 \leq \eta \|e\|_A^2$$

for least squares interpolation as defined in section 4.2.

Given a set of test vectors $u^{(1)}, \ldots, u^{(k)} \in \mathbb{C}^m$ and corresponding weights

$\omega_1, \ldots, \omega_k \in \mathbb{R}^+$ we write

$$W = \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_k \end{pmatrix} \in \mathbb{R}^{k\times k} \quad \text{and} \quad U = \begin{pmatrix} u^{(1)} & \cdots & u^{(k)} \end{pmatrix} \in \mathbb{C}^{m\times k}.$$

Again, given the set of variables Ω, denote the canonical injection of a subset $\Omega_i \subset \Omega$ into Ω by $I_{|\Omega_i}$ and write shortly v_{Ω_i} for any vector v restricted to Ω_i, i.e., $v_{\Omega_i} = I^T_{|\Omega_i} v$.

Further, we assume that a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ and for each $i \in \mathcal{F}$ a set of interpolatory points $\mathcal{C}_i \subset \mathcal{C}$ are given. Additionally, we define the interpolatory windows $w_i = \{i\} \cup \mathcal{C}_i$ and assume that when it comes to numbering the elements of w_i, i is the first.

The following auxiliary result concerns the localization of the weak approximation measure analogous to the derivation in AMGe made in Definition 2.32. Again, given an interpolation operator P, we choose a restriction operator R, such that $RP = I$. In the case, that we have given a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$ and P is the identity on the coarse variable set $\mathcal{C}$, we can choose $R = I^T_{|\mathcal{C}}$.

Lemma 4.24. *For interpolation P with $I^T_{|\mathcal{C}} P = I$, we have*

$$\| (I - PR)\, e\|_2^2 = \sum_{i\in\mathcal{F}} \| I^T_{|\{i\}} (I - PR)\, e\|_2^2. \tag{4.21}$$

Proof. Introducing the trivial partition of unity $I = \sum_{i\in\Omega} I_{|\{i\}} I^T_{|\{i\}}$ in (4.21) yields

$$\begin{aligned} \| (I - PR)\, e\|_2^2 &= \langle (I - PR)\, e, (I - PR)\, e\rangle_2 \\ &= \sum_{i\in\Omega}\sum_{j\in\Omega} \langle I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e, I_{|\{j\}} I^T_{|\{j\}} (I - PR)\, e\rangle_2 \\ &= \sum_{i\in\Omega} \langle I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e, I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e\rangle_2. \end{aligned}$$

As $(PR)_{|\mathcal{C}} = I$ we obtain

$$\begin{aligned} & \sum_{i\in\Omega} \langle I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e, I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e\rangle_2 \\ = & \sum_{i\in\mathcal{F}} \langle I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e, I_{|\{i\}} I^T_{|\{i\}} (I - PR)\, e\rangle_2 \\ = & \sum_{i\in\mathcal{F}} \| I^T_{|\{i\}} (I - PR)\, e\|_2^2. \end{aligned}$$

□

Now, we consider the connection between the above weak approximation property and least squares interpolation, in particular we define corresponding notions of local test vectors and a local weak approximation property.

The i-th row of the interpolation operator P that is defined as the solution of the least squares problem (4.8) is given as

$$\underset{p_i\in\mathbb{C}^{|\mathcal{C}_i|}}{\operatorname{argmin}}\|U_{\{i\}}W^{\frac{1}{2}}-p_iU_{\mathcal{C}_i}W^{\frac{1}{2}}\|_2^2=\underset{p_i\in\mathbb{C}^{|\mathcal{C}_i|}}{\operatorname{argmin}}\|\begin{pmatrix}1 & -p_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}\|_2^2. \tag{4.22}$$

With this formulation of the least squares problem we can derive an equivalent problem that involves a local operator similar to the derivation used in AMGe theory (cf. [BCF+00, HV01]).

Lemma 4.25. *Assuming that for each $i\in\mathcal{F}$ the least squares problem* (4.22) *has a unique solution p_i with corresponding minimal functional value $c_{\mathcal{L}}^{(i)}=\mathcal{L}(p_i)$ we then have*

$$\|\begin{pmatrix}1 & -p_i\end{pmatrix}e\|_2^2\leq c_{\mathcal{L}}^{(i)}\langle A_{w_i}^{(U)}e,e\rangle_2,\quad \textit{for all } e\in\mathbb{C}^m,$$

with $A_{w_i}^{(U)}=\left(U_{w_i}WU_{w_i}^H\right)^{-1}$.

Proof. According to (4.22) and the definition of the operator norm we have

$$\begin{aligned}c_{\mathcal{L}}^{(i)} &= \min_{q_i\in\mathbb{C}^{|w_i|}}\|\begin{pmatrix}1 & -q_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}\|_2^2\\ &= \|\begin{pmatrix}1 & -p_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}\|_2^2=\max_{e\in\mathbb{C}^{|k|}}\frac{\|\begin{pmatrix}1 & -p_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}e\|_2^2}{\|e\|_2^2}\end{aligned}$$

Thus if the number of test vectors $k\geq|w_i|$ for all $i\in\mathcal{F}$ and $\operatorname{rank}(U_{w_i}W^{\frac{1}{2}})\geq|w_i|$, then $\left(U_{w_i}WU_{w_i}^H\right)\in\mathbb{C}^{|w_i|\times|w_i|}$ is non-singular, i.e., $\left(U_{w_i}WU_{w_i}^H\right)^{-1}$ exists. Let $U_{w_i}W^{\frac{1}{2}}=Y_{w_i}\Sigma_{w_i}Z_{w_i}^H$ be the singular value decomposition of $U_{w_i}W^{\frac{1}{2}}\in\mathbb{C}^{|w_i|\times|k|}$, with unitary $Y_{w_i}\in\mathbb{C}^{|w_i|\times|w_i|}$, $\Sigma_{w_i}=\begin{pmatrix}\Sigma_{w_i}^+ & 0\end{pmatrix}\in\mathbb{C}^{|w_i|\times|k|}$, $\Sigma_{w_i}^+\in\mathbb{C}^{|w_i|\times|w_i|}$ and unitary $Z_{w_i}^H\in\mathbb{C}^{|k|\times|k|}$.

Then, changing the dimension of e as appropriate below and using the fact that Y_{w_i} and Z_{w_i} are unitary matrices we have

$$\begin{aligned}\|\begin{pmatrix}1 & -p_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}\|_2^2 &= \max_{e\neq0}\frac{\|\begin{pmatrix}1 & -p_i\end{pmatrix}U_{w_i}W^{\frac{1}{2}}e\|_2^2}{\|e\|_2^2}\\ &= \max_{e\neq0}\frac{\|\begin{pmatrix}1 & -p_i\end{pmatrix}Y_{w_i}\begin{pmatrix}\Sigma_{w_i}^+ & 0\end{pmatrix}Z_{w_i}^He\|_2^2}{\|Z_{w_i}^He\|_2^2}\\ &= \max_{e\neq0}\frac{\|\begin{pmatrix}1 & -p_i\end{pmatrix}Y_{w_i}\begin{pmatrix}\Sigma_{w_i}^+ & 0\end{pmatrix}e\|_2^2}{\|e\|_2^2}.\end{aligned}$$

Writing $e = \begin{pmatrix} e_{w_i}^T & \tilde{e}^T \end{pmatrix}^T$ we obtain

$$\frac{\| \begin{pmatrix} \Sigma_{w_i}^+ & 0 \end{pmatrix} e\|_2^2}{\|e\|_2^2} \geq \frac{\|\Sigma_{w_i}^+ e_{w_i}\|_2^2}{\|e_{w_i}\|_2^2 + \|\tilde{e}^2\|_2}.$$

Thus, we have

$$\begin{aligned}
\| \begin{pmatrix} 1 & -p_i \end{pmatrix} U_{w_i} W^{\frac{1}{2}} \|_2^2 &= \max_{e \neq 0} \frac{\| \begin{pmatrix} 1 & -p_i \end{pmatrix} Y_{w_i} \Sigma_{w_i}^+ e\|_2^2}{\|e\|_2^2} \\
&= \max_{e \neq 0} \frac{\| \begin{pmatrix} 1 & -p_i \end{pmatrix} Y_{w_i} \Sigma_{w_i}^+ e\|_2^2}{\| \left(\Sigma_{w_i}^+\right)^{-1} Y_{w_i}^H Y_{w_i} \Sigma_{w_i}^+ e\|_2^2} \\
&= \max_{e \neq 0} \frac{\| \begin{pmatrix} 1 & -p_i \end{pmatrix} e\|_2^2}{\| \left(\Sigma_{w_i}^+\right)^{-1} Y_{w_i}^H e\|_2^2} \\
&= \max_{e \neq 0} \frac{\| \begin{pmatrix} 1 & -p_i \end{pmatrix} e\|_2^2}{\langle Y_{w_i} \left(\Sigma_{w_i}^+\right)^{-2} Y_{w_i}^H e, e\rangle_2}.
\end{aligned}$$

The result follows by the observation that

$$\left(U_{w_i} W U_{w_i}^H\right)^{-1} = Y_{w_i} \left(\Sigma_{w_i}^+\right)^{-2} Y_{w_i}^H.$$

□

Given an overlapping decomposition of $\Omega = \cup_i \Omega_i$ such that for each Ω_i a local hermitian positive definite linear operator A_{Ω_i} exists with its corresponding local energy form $\langle ., . \rangle_{A_{\Omega_i}}$, assume the following summability property holds

$$\sum_i \langle v_{\Omega_i}, v_{\Omega_i} \rangle_{A_{\Omega_i}} \leq \sigma \langle v, v \rangle_A, \tag{4.23}$$

where σ does only depend on the overlap of Ω_i. Assuming that such local energy forms exist, we can formulate the last assumption needed to prove a weak approximation property for least squares interpolation by comparing the local operators implied by the least squares formulation of interpolation $A_{w_i}^U$ with the given local operators A_{w_i} that fulfill 4.23. Note that the assumption of the existence of local energy forms that fulfill (4.23) is a natural one in the context of finite element discretizations and was used in the proof of the weak approximation property for the AMGe method, as reviewed in section 2.3.

Theorem 4.26. *Let P be the least squares interpolation defined by a set of test vectors $u^{(1)}, \ldots, u^{(k)}$ and corresponding weights $\omega_1, \ldots, \omega_k$ and $R = I_{|\mathcal{C}}^T$ the canonical restriction to $\mathcal{C}$. Assume that the following holds:*

1. *The interpolatory windows w_i form an overlapping decomposition*

$$\Omega = \cup_{i\in\mathcal{F}} w_i = \cup_{i\in\mathcal{F}} \left(\{i\} \cup \mathcal{C}_i\right).$$

2. *We have given associated local operators A_{w_i} with corresponding energy forms $\langle ., . \rangle_{A_{w_i}}$ such that,*

$$\sum_{i\in\mathcal{F}} \langle v_{w_i}, v_{w_i} \rangle_{A_{w_i}} \leq \sigma \langle v, v \rangle_A. \tag{4.24}$$

3. *For each $i \in \mathcal{F}$, the least squares problem* (4.22) *is uniquely solvable with minimizer $I^T_{|_{\{i\}}} P$ and minimum*

$$c^{(i)}_{\mathcal{L}} = \mathcal{L}(I^T_{|_{\{i\}}} P).$$

4. *for each $i \in \mathcal{F}$ there exists a number $\eta^{(U)}_i > 0$ such that,*

$$\sup_{e\neq 0} \frac{\langle e, e \rangle_{A^{(U)}_{w_i}}}{\langle e, e \rangle_{A_{w_i}}} \leq \eta^{(U)}_i. \tag{4.25}$$

We then have

$$\| \left(I - PR\right) e \|_2^2 \leq \eta \langle e, e \rangle_A, \quad \eta = \sigma \max_{i\in\mathcal{F}} \left(\eta^{(U)}_i c^{(i)}_{\mathcal{L}} \right).$$

Proof. By Lemma 4.24 we have

$$\| \left(I - PR\right) e \|_2^2 = \sum_{i\in\mathcal{F}} \| I^T_{|_{\{i\}}} \left(I - PR\right) e \|_2^2.$$

Now, using Lemma 4.25 for each part of the sum we obtain

$$\| \left(I - PR\right) e \|_2^2 = \sum_{i\in\mathcal{F}} c^{(i)}_{\mathcal{L}} \langle e, e \rangle_{A^{(U)}_{w_i}}.$$

Further, using assumption (4.25) on $A^{(U)}_{w_i}$ yields

$$\| \left(I - PR\right) e \|_2^2 \leq \sum_{i\in\mathcal{F}} \eta^{(U)}_i c^{(i)}_{\mathcal{L}} \langle e, e \rangle_{A_{w_i}}.$$

Finally, using the summability property (4.24) of the local energy forms we obtain

$$\| \left(I - PR\right) e \|_2^2 \leq \sigma \max_{i\in\mathcal{F}} \left(\eta^{(U)}_i c^{(i)}_{\mathcal{L}} \right) \langle e, e \rangle_A.$$

□

Next we discuss the assumptions of Theorem 4.26. First, we consider the existence of local energy forms induced by local operators A_{w_i} that fulfill (4.24) in a general framework and than estimates on the spectral equivalence (4.25) of the local energy forms and the local operators $A_{w_i}^{(U)}$ that are induced by the least squares formulation.

We start analyzing this issue by taking a closer look at particular choices of test vectors and corresponding weights that yield local operators $A_{w_i}^{(U)}$ in the least squares formulation and test if they fulfill the summability property (4.24). In this case the second assumption (4.25) is fulfilled with $\eta_i^{(U)} = 1$ for all $i \in \mathcal{F}$. Furthermore, we gain insight on a particular choice of local energy forms that can be used to measure arbitrary choices of test vectors and corresponding weights and their respective resulting local operators.

A possible approach to define local energy forms that also fits into the least squares formulation is as follows. Given the eigendecomposition of $A = V\Lambda V^H$, we define the set of test vectors written in matrix form by $U = V$, i.e., the complete set of eigenvectors of A. Even though this choice is not feasible in practice, it is an interesting choice in theory. Associated with these test vectors we choose the weights as suggested in section 4.2, namely $W^{\frac{1}{2}} = \Lambda^{-1}$. Thus, for each $i \in \mathcal{F}$ and associated interpolatory window w_i, we get the corresponding local set of test vectors

$$U_{w_i} = I_{|_{w_i}}^T U = I_{|_{w_i}}^T V.$$

Lemma 4.27. *With the choice of test vectors $U = V$ and weights $W^{\frac{1}{2}} = \Lambda^{-1}$, we obtain for $A_{w_i}^{(U)}$ from Lemma 4.25,*

$$A_{w_i}^{(U)} = S_{w_i w_i}(A^H A) = S_{w_i w_i}(A^2),$$

where $S_{w_i w_i}(A^H A)$ denotes the Schur complement of $A^H A$ onto the subset of variables w_i.

Proof. By the definition of the test vectors and weights we have

$$A_{w_i}^{(U)} = \left(U_{w_i} W U_{w_i}^H\right)^{-1} = \left(I_{|_{w_i}}^T V \Lambda^{-1} \Lambda^{-H} V^H I_{|_{w_i}}\right)^{-1} = \left(I_{|_{w_i}}^T \left(A^H A\right)^{-1} I_{|_{w_i}}\right)^{-1}.$$

A little algebra shows that we have for any hermitian positive definite linear operator X and non-overlapping splitting of variables $\Omega = w_i \cup \chi$ the following block structure of X^{-1} (cf. [Axe94, p. 93]),

$$X^{-1} = \begin{pmatrix} (S_{w_i w_i}(X))^{-1} & -X_{w_i w_i}^{-1} X_{w_i \chi} \left(S_{\chi\chi}(X)\right)^{-1} \\ (S_{w_i w_i}(X))^{-1} X_{w_i \chi} X_{\chi\chi}^{-1} & (S_{\chi\chi}(X))^{-1} \end{pmatrix}.$$

Thus, we obtain

$$A_{w_i}^{(U)} = S_{w_i w_i}(A^H A) = S_{w_i w_i}(A^2).$$

□

In what follows, we need some basic results for Schur complements and harmonic extensions.

Definition 4.28. *Given a subset of variables $w \subset \Omega$ and denoting the complement of w by χ, i.e., $\chi = \Omega \setminus w$, we permute A according to this splitting*

$$A = \begin{pmatrix} A_{ww} & A_{w\chi} \\ A_{\chi w} & A_{\chi\chi} \end{pmatrix}.$$

Then we define the harmonic extension P_w *of w in Ω by*

$$P_w = \begin{pmatrix} I_w \\ -A_{\chi\chi}^{-1} A_{\chi w} \end{pmatrix}.$$

Note, that we introduced a specific harmonic extension, $P_{\mathcal{F}}$, earlier in section 2.5 as the ideal interpolation operator given a splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$.

Lemma 4.29. *The harmonic extension P_w fulfills the following equation*

$$AP_w = I_{|w} S_{ww}(A).$$

Proof. We have

$$AP_w = \begin{pmatrix} A_{ww} & A_{w\chi} \\ A_{\chi w} & A_{\chi\chi} \end{pmatrix} \begin{pmatrix} I_w \\ -A_{\chi\chi}^{-1} A_{\chi w} \end{pmatrix} = \begin{pmatrix} A_{ww} - A_{w\chi} A_{\chi\chi}^{-1} A_{\chi w} \\ 0 \end{pmatrix}.$$

□

Theorem 4.30. *Given a subset of all variables $w \subset \Omega$ and its complement χ we write $A \in \mathbb{C}^{m \times m}$ in block permuted form as*

$$A = \begin{pmatrix} A_{ww} & A_{w\chi} \\ A_{\chi w} & A_{\chi\chi} \end{pmatrix}.$$

Then the Schur complement $S_{ww}(A^H A)$ is given as

$$S_{ww}(A^H A) = S_{ww}(A^H) \left(P_w^H P_w\right)^{-1} S_{ww}(A), \tag{4.26}$$

with P_w the harmonic extension from Definition 4.28.

Proof. Using the result of Lemma 4.29 we obtain

$$P_w = A^{-1} I_{|_w} S_{ww}(A).$$

Plugging this formulation of P_w into the right hand side of (4.26) and using the equality $(S_{ww}(A))^H = S_{ww}(A^H)$ yields

$$\begin{aligned} & S_{ww}(A^H) \left(P_w^H P_w\right)^{-1} S_{ww}(A) \\ = \; & S_{ww}(A^H) \left(S_{ww}(A^H) I_{|_w}^T A^{-H} A^{-1} I_{|_w} S_{ww}(A)\right)^{-1} S_{ww}(A) \\ = \; & \left(I_{|_w}^T \left(A^H A\right)^{-1} I_{|_w}\right)^{-1}. \end{aligned}$$

□

Again using the fact that

$$S_{ww}(A) = P_w^H A I_w$$

we can write the Schur complement $S_{ww}(A^H A)$ in yet another form using the orthogonal projection $\pi_2(P_w)$ onto $\mathcal{R}(P_w)$,

$$S_{ww}(A^H A) = I_{|_w} A^H P_w \left(P_w^H P_w\right)^{-1} P_w^H A I_{|_w}^T = I_{|_w} A^H \pi_2(P_w) A I_{|_w}^T.$$

Hence, the summability property (4.24) that we have to prove, can be written as

$$\langle \sum_{i \in \mathcal{F}} \pi_2(P_{w_i}) A e_{w_i}, e_{w_i} \rangle_A \leq \sigma \langle e, e \rangle_A. \tag{4.27}$$

This inequality (4.27) is always fulfilled with $\sigma = |\mathcal{F}| \|A\|_2$, as a sum of orthogonal projections, but since $|\mathcal{F}|$ usually depends on the size of A, we are not satisfied by this trivial bound.

Lemma 4.31. *We can write* (4.27) *equivalently as*

$$\langle Ae, e \rangle_2 \leq \sigma \langle \left(\sum_{i \in \mathcal{F}} \pi_2(P_{w_i}) \right)^{-1} e, e \rangle_2. \tag{4.28}$$

Proof. With $B = \left(\sum_{i \in \mathcal{F}} \pi_2(P_{w_i})\right)$ and some exchanges of variables we obtain,

$$\begin{aligned} & \langle ABAe, e \rangle_2 && \leq \sigma \langle Ae, e \rangle_2 \\ \iff \; & \langle A^{\frac{1}{2}} B A^{\frac{1}{2}} y, y \rangle_2 && \leq \sigma \langle y, y \rangle_2 \\ \iff \; & \langle B^{\frac{1}{2}} A B^{\frac{1}{2}} y, y \rangle_2 && \leq \sigma \langle y, y \rangle_2 \\ \iff \; & \langle Az, z \rangle_2 && \leq \sigma \langle B^{-1} z, z \rangle_2. \end{aligned}$$

□

In what follows, we need a result of subspace correction methods.

Lemma 4.32. *Assuming* $\mathbb{C}^m = \sum_{i\in\mathcal{F}} \mathcal{R}(P_{w_i})$, *we obtain*

$$\langle \left(\sum_{i\in\mathcal{F}} \pi_2(P_{w_i}) \right)^{-1} e, e\rangle_2 = \inf_{e=\sum_{i\in\mathcal{F}} e_{w_i}} \|P_{w_i} e_{w_i}\|_2^2. \tag{4.29}$$

Proof. For the sake of brevity we refer to [XZ02, Lemma 2.4]. □

Thus, taking any decomposition of e that fulfills

$$\sum_{i\in\mathcal{F}} e_{w_i} = e$$

and inserting it in the left-hand-side of (4.28) we get

$$\langle Ae, e\rangle_2 = \langle \sum_{i\in\mathcal{F}} \sum_{j\in\mathcal{F}} P_{w_j}^H A P_{w_i} e_{w_i}, e_{w_j}\rangle_2.$$

Define $\gamma_{w_i w_j}$ to be the smallest constant such that

$$\langle P_{w_j}^H A P_{w_i} e_{w_i}, e_{w_j}\rangle_2 \leq \gamma_{w_i w_j} \|P_{w_i} e_{w_i}\|_2 \|P_{w_j} e_{w_j}\|_2. \tag{4.30}$$

Lemma 4.33. *The bounds* $\gamma_{w_i w_j}$ *in* (4.30) *satisfy*

$$\gamma_{w_i w_j} \leq \|A\|_2.$$

Proof. Using the strengthened Cauchy-Bunyakowski-Schwarz-inequality of Definition 4.2 we know that

$$\langle P_{w_i} e_{w_i}, P_{w_j} e_{w_j}\rangle_A \leq \hat{\gamma}_{w_i w_j} \|P_{w_i} e_{w_i}\|_A \|P_{w_j} e_{w_j}\|_A$$

with $\hat{\gamma}_{w_i w_j} \leq 1$. Due to the fact that $\|x\|_A \leq \|A\|_2^{\frac{1}{2}} \|x\|_2$ we obtain

$$\langle P_{w_i} e_{w_i}, P_{w_j} e_{w_j}\rangle_A \leq \hat{\gamma}_{w_i w_j} \|A\|_2 \|P_{w_i} e_{w_i}\|_2 \|P_{w_j} e_{w_j}\|_2.$$

As $\gamma_{w_i w_j}$ is the smallest constant fulfilling (4.30), we have

$$\gamma_{w_i w_j} \leq \hat{\gamma}_{w_i w_j} \|A\|_2 \leq \|A\|_2.$$

□

With this we can bound σ in terms of $\gamma_{w_i w_j}$.

Lemma 4.34. *Introducing the symmetric and entry-wise positive matrix* $\Gamma = \left(\gamma_{w_i w_j}\right)_{i,j}$*, we obtain*

$$\sigma \leq \rho(\Gamma).$$

Proof. We have

$$\rho(\Gamma) \leq \|\Gamma\|_\infty = \max_{i \in \mathcal{F}} \sum_{j \in \mathcal{F}} \gamma_{w_i w_j}.$$

Thus,

$$\begin{aligned} \langle Ae, e\rangle_2 &= \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{F}} \langle AP_{w_i} e_{w_i}, P_{w_j} e_{w_j}\rangle_2 \\ &\leq \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{F}} \gamma_{w_i w_j} \|P_{w_i} e_{w_i}\|_2 \|P_{w_j} e_{w_j}\|_2 \\ &\leq \rho(\Gamma) \sum_{i \in \mathcal{F}} \|P_{w_i} e_{w_i}\|_2^2. \end{aligned}$$

Finally, plugging this result into (4.29) and taking the infimum over all decompositions $\sum_w P_w e_w = e$ yields $\sigma \leq \rho(\Gamma)$. □

Corollary 4.35. *In the case where* $\gamma_{w_i w_j}$ *are exponentially decaying, i.e., given a function* $\delta(.,.)$ *that measures the distance between two windows* w_i, w_j *on the artificial grid* G_A *that fulfills*

$$\delta(w_i, w_j) \geq \alpha |i - j|,$$

we assume that there exists a $\tau \in [0, 1)$ *such that,*

$$\gamma_{w_i w_j} \leq \|A\|_2 \tau^{\delta(w_i, w_j)}.$$

Then we have

$$\sigma \leq \frac{2\|A\|_2}{1 - \tau}.$$

Proof. Denoting $\tilde{\Gamma}_{i,j} = \|A\|_2 \tau^{\delta(w_i, w_j)}$ we have $\sigma(\Gamma) \leq \sigma(\tilde{\Gamma})$. The hypothesis then follows from Lemma 4.34 and application of the geometric sum. □

Using the specific properties of P_w we can give a more precise description of $\gamma_{w_i w_j}$ as follows.

Lemma 4.36. *For* $\gamma_{w_i w_j}$ *defined by* (4.30) *we have*

$$\gamma_{w_i w_j} = \sup_{e_{w_i}, e_{w_j}} \frac{\langle [A^{-2}]_{w_j w_j}^{-\frac{1}{2}} [A^{-1}]_{w_j w_i} [A^{-2}]_{w_i w_i}^{-\frac{1}{2}} e_{w_i}, e_{w_j}\rangle_2}{\|e_{w_i}\|_2 \|e_{w_j}\|_2},$$

where $[X]_{v,v'}$ *denotes the* v, v' *block of the matrix* X*.*

Proof. By definition of $\gamma_{w_i w_j}$ we have

$$\gamma_{w_i w_j} = \sup_{e_{w_i}, e_{w_j}} \frac{\langle AP_{w_i} e_{w_i}, P_{w_j} e_{w_j} \rangle_2}{\|P_{w_i} e_{w_i}\|_2 \|P_{w_j} e_{w_j}\|_2}. \tag{4.31}$$

Writing the canonical injection of w into $\mathbb{C}^m$ as $I_{|_w}$ and using the definition of P_w we have

$$AP_w = I_{|_w} S_{ww}(A) \quad \text{and} \quad P_w = A^{-1} I_{|_w} S_{ww}(A).$$

Thus, we obtain with the abbreviation $S_{ww} = S_{ww}(A)$,

$$\|P_w e_w\|_2 = \langle I_{|_w}^T A^{-2} I_{|_w} S_{ww} e_w, S_{ww} e_w \rangle_2 = \langle \left[A^{-2}\right]_{ww} S_{ww} e_w, S_{ww} e_w \rangle_2,$$

such that we can write (4.31) as

$$\gamma_{w_i w_j} = \sup_{e_{w_i}, e_{w_j}} \frac{\langle A^{-1} I_{|_{w_i}} S_{w_i w_i} e_{w_i}, I_{|_{w_j}} S_{w_j w_j} e_{w_j} \rangle_2}{\langle [A^{-2}]_{w_i w_i} S_{w_i w_i} e_{w_i}, S_{w_i w_i} e_{w_i} \rangle_2^{\frac{1}{2}} \langle [A^{-2}]_{w_j w_j} S_{w_j w_j} e_{w_j}, S_{w_j w_j} e_{w_j} \rangle_2^{\frac{1}{2}}}.$$

Then, the change of variables $y_{w_i} = S_{w_i w_i} e_{w_i}$ and $y_{w_j} = S_{w_j w_j} e_{w_j}$ yields

$$\gamma_{w_i w_j} = \sup_{y_{w_i}, y_{w_j}} \frac{\langle [A^{-1}]_{w_j w_i} y_{w_i}, y_{w_j} \rangle_2}{\langle [A^{-2}]_{w_i w_i} y_{w_i}, y_{w_i} \rangle_2^{\frac{1}{2}} \langle [A^{-2}]_{w_j w_j} y_{w_j}, y_{w_j} \rangle_2^{\frac{1}{2}}}.$$

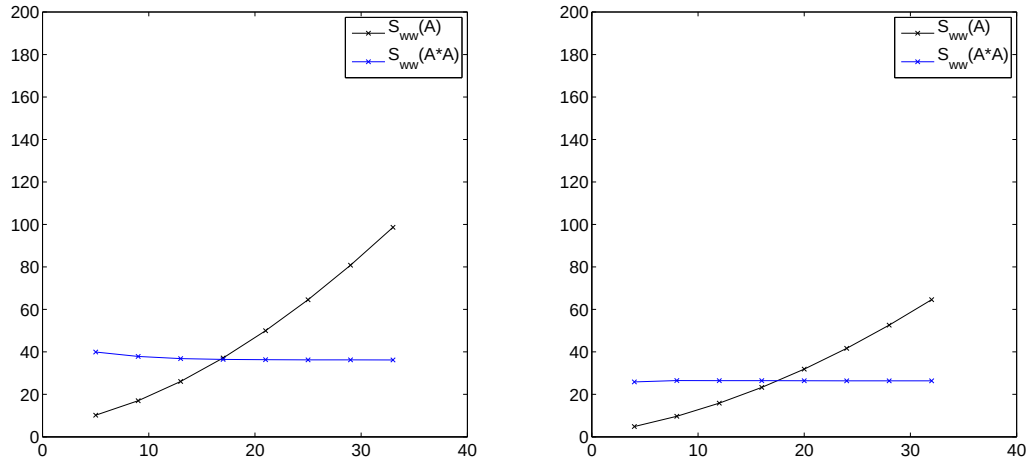
With another change of variables $z_{w_i} = [A^{-2}]_{w_i w_i}^{\frac{1}{2}} y_{w_i}$ and $z_{w_j} = [A^{-2}]_{w_j w_j}^{\frac{1}{2}} y_{w_j}$ we obtain

$$\gamma_{w_i w_j} = \sup_{z_{w_i}, z_{w_j}} \frac{\langle [A^{-2}]_{w_j w_j}^{-\frac{1}{2}} [A^{-1}]_{w_j w_i} [A^{-2}]_{w_i w_i}^{-\frac{1}{2}} z_{w_i}, z_{w_j} \rangle_2}{\langle z_{w_i}, z_{w_i} \rangle_2^{\frac{1}{2}} \langle z_{w_j}, z_{w_j} \rangle_2^{\frac{1}{2}}}.$$

□

Unfortunately, proving a bound on σ for this particular choice of test vectors that is independent of the problem-size (for problems discretized with varying mesh-size h) turns out to be very intricate. Numerical experiments suggest that at least for discretizations of the Laplace operator Δ, which we discuss in section 5.1 in more detail, such a bound can be found and that the estimate by means of $\rho(\Gamma)$ is fairly sharp. In Figure 4.2, we present bounds on σ in

$$\langle \left(\sum_{i \in \mathcal{F}} A_{w_i}^{(U)} \right) e, e \rangle_2 \leq \sigma \langle Ae, e \rangle_2 \tag{4.32}$$



(a) Spectral bound σ (full-coarsening) (b) Spectral bound σ (red-black)

Figure 4.2: Bounds σ in the spectral equivalence of the sum of local matrices $A_w^{(U)}$ and A in (4.32) a linear operator A arising in discretization of the Laplace operator on an equidistant $N \times N$ grid. The local operators $A_w^{(U)}$ are defined by either $U = V\Lambda^{-1}$ implying $A_w^{(U)} = S_{ww}(A^2)$ or $U = V\Lambda^{-\frac{1}{2}}$ implying $A_w^{(U)} = S_{ww}(A)$. Using two different coarse-variable sets, full-coarsening in 4.2(a) and red-black in 4.2(b) and associated interpolatory windows w. For an illustration of full-coarsening see Figure 5.2 and for red-black or odd-even see Figure 4.1.

for different choices of coarse-variable sets and interpolatory windows. These results justify the choice of test vectors $U = V\Lambda^{-1}$ and show further that the choice of test vectors $U = V\Lambda^{-\frac{1}{2}}$ which gives $A_w^{(U)} = S_{ww}(A)$, does not yield a constant bound when increasing the problem size N. Unfortunately, we were not yet successful in providing a theoretical bound nor have we been able to classify the matrices for which one can find such a bound for (4.32) independent of the size of the problem with the choice $U = V\Lambda^{-1}$. Our ongoing research aims to gain a deeper understanding of this assumption in the proof to ensure a weak approximation property for least squares interpolation.

Let us now consider the other assumption made in Theorem 4.26. Assuming that local energy forms are given for each interpolatory window $w_i = \{i\} \cup \mathcal{C}_i, i \in \mathcal{F}$, i.e., we have

$$\sum_{i \in \mathcal{F}} \langle v_{w_i}, v_{w_i} \rangle_{A_{w_i}} \leq \sigma \langle v, v \rangle_A.$$

Given the eigendecomposition of A by $A = V\Lambda V^H$ it is natural to view

each test vector as a linear combination of the eigenvectors $v_1, \ldots, v_m$, i.e.,

$$u^{(j)} = \sum_i \beta_{ij} v_i, \quad j = 1, \ldots, k.$$

In this way, U can be written as

$$U = VB, \quad B = (\beta_{ij})_{i,j} \in \mathbb{C}^{m \times k}.$$

Hence, adding weights, again denoted by W, we have

$$A_{w_i}^U = \left(V_{w_i} B W B^H V_{w_i}^H\right)^{-1}.$$

Further assume that each A_{w_i}, corresponding to the given local energy forms, can be written as

$$A_{w_i} = \left(V_{w_i} W_{w_i} V_{w_i}^H\right)^{-1},$$

that is, it can be written as an operator $A_{w_i}^{(U)}$ arising from the least squares formulation with the test vectors $U = V$ and a choice of weights that (might) depend on the respective set w_i. Then, we can formulate (4.25) as follows,

$$\sup_{e \neq 0} \frac{\langle e, e \rangle_{A_{w_i}^{(U)}}}{\langle e, e \rangle_{A_{w_i}}} = \frac{\langle V_{w_i} B W B^H V_{w_i}^H e, e \rangle_2}{\langle V_{w_i} W_{w_i} V_{w_i}^H e, e \rangle_2}.$$

With a change of variables $y = V_{w_i}^H e$, we obtain

$$\sup_{y \neq 0} \frac{\langle B W B^H y, y \rangle_2}{\langle W_{w_i} y, y \rangle_2} = \sup_{z \neq 0} \frac{\langle W_{w_i}^{-\frac{1}{2}} B W B^H W_{w_i}^{-\frac{1}{2}} z, z \rangle_2}{\langle z, z \rangle_2}.$$

Thus, the assumption on spectral equivalence of the local operators can be expressed in terms of the spectral radius of a matrix that only involves the local weights W, W_{w_i} and the coefficient matrix B of the test vectors used in the definition of least squares interpolation.

Having completed the convergence analysis discussion, we proceed to describe the details of the adaptive setup process, namely the multiscale computation of appropriate test vectors used in least squares interpolation.

4.3 Bootstrap AMG Techniques

There has been a long debate about how to compute test vectors and incorporated them into adaptive algebraic multigrid methods. The most prevalent approach is to compute one component at a time as proposed in the

smoothed aggregation framework [BFM+04]. An alternate approach proposed by Brandt in [Bra00] is to use a *bootstrap* adaptivity. One part of this bootstrap setup is the use of least squares interpolation defined in section 4.2. Tightly coupled to this is the computation of an appropriate set of test vectors $u^{(1)}, \ldots, u^{(k)}$.

In this section, we introduce these ideas, namely the computation of test vectors and their multigrid enhancement and the coarse grid treatment of almost zero modes.

4.3.1 Test Vectors

The test vectors used in the least squares definition of interpolation and coarse-grid generation should represent algebraically smooth error. Thus, an initial set of test vectors can be calculated by applying the given multigrid smoother.

Given the linear operator A and the corresponding linear system to solve $Au = f$, we assume the existence of a smoother S for this operator, as defined in Definition 2.2. The original idea of adaptive algebraic multigrid methods is to use the current multigrid method to expose error components that are not treated well enough by the multigrid hierarchy. That is, if no initial multigrid hierarchy is given, the smoother S is used to compute such an error component. In the original idea this process is repeated one component at a time. In the bootstrap algebraic multigrid approach, we choose instead of only one initial test vector a set of test vectors $u^{(1)}, \ldots, u^{(k)}$. Then the first set of test vectors that is used to define least squares interpolation is given by $S^\nu u^{(1)}, \ldots, S^\nu u^{(k)}$, where we apply the smoother ν times to the homogeneous equations

$$Au^{(l)} = 0, \quad l = 1, \ldots, k.$$

The smoother removes error components at the low end of the spectrum of its error propagator $S = I - \mathcal{M}A$, hence the vectors $S^\nu u^{(l)}$, assuming that the smoother defines a converging iteration, are rich in algebraically smooth error, i.e., low-energy vectors of $\mathcal{M}A$. In order to get a well-defined least squares fit for each fine-grid point $i \in \mathcal{F}$, the result in Remark 4.12 suggests we choose the number of test vectors k, such that

$$k \geq \max_{i \in \mathcal{F}}(|\mathcal{C}_i|), \tag{4.33}$$

where $\mathcal{C}_i$ is the set of interpolatory points for $i \in \mathcal{F}$. Our numerical experience suggests that it is beneficial to use more test vectors as presented by (4.33), both to guarantee the well-definiteness of the least squares fits and also to

improve the representation of the algebraically smooth error in the span of the test vectors. In chapter 5.1 we show numerical evidence of this heuristic.

Motivated by the fundamental principle of multigrid, the interplay of smoothing and coarse-grid correction, interpolation P must accurately represent the low end of the spectrum of $\mathcal{M}A$ to yield an efficient multigrid method. Using weights in the least squares process, as explained in section 4.2, biases the fit towards these algebraically smooth vectors. When starting with random initial vectors it can be expected that the weights defined in section 4.2 after a few relaxations do not differ largely and the weighting can be omitted, but in the case one has exposed the algebraically smooth error very accurately, the weighting process takes care of the proper representation of these vectors in the least squares built interpolation.

Given the multigrid hierarchy of operators $A_0, A_1, \ldots, A_L$ and corresponding interpolation operators $P_{l+1}^l, l = 0, \ldots, L-1$, we have a chain of subspaces of $\mathbb{C}^m$ spanned by the columns of the composite interpolation operators $P_l = P_1^0 \cdot \ldots \cdot P_l^{l-1},\ l = 1, \ldots, L$

$$\mathbb{C}^m = \operatorname{span}(P_1) \supset \ldots \supset \operatorname{span}(P_L).$$

Thus, for any given vector $x_l \in \mathbb{C}^{m_l}$ we have

$$\langle x_l, x_l \rangle_{A_l} = \langle P_l x_l, P_l x_l \rangle_A.$$

Furthermore, defining $T_l = P_l^H P_l$ we obtain

$$\frac{\langle x_l, x_l \rangle_{A_l}}{\langle x_l, x_l \rangle_{T_l}} = \frac{\langle P_l x_l, P_l x_l \rangle_A}{\langle P_l x_l, P_l x_l \rangle_2}.$$

This observation motivates the bootstrap process outlined below.

Lemma 4.37. *On any level l, given a vector $v^{(l)} \in \mathbb{C}^{m_l}$ and $\lambda^{(l)} \in \mathbb{C}$, such that,*

$$A_l v^{(l)} = \lambda^{(l)} T_l v^{(l)}, \tag{4.34}$$

we have

$$rq(P_l v^{(l)}) = \frac{\langle P_l v^{(l)}, P_l v^{(l)} \rangle_A}{\langle P_l v^{(l)}, P_l v^{(l)} \rangle_2} = \lambda^{(l)}.$$

Proof. The hypothesis follows by simple algebraic transformations using that $v^{(l)}$ is an eigenvector of the generalized eigenvalue problem (4.34) and the equations $A_l = P_l^H A P_l$ and $T_l = P_l^H P_l$. □

Lemma 4.37 has practical implications as it allows one to connect the eigenvectors and eigenvalues of the operators in the multigrid hierarchy with the eigenvectors and eigenvalues of the finest grid operator A. This allows us to compute approximations to eigenvectors corresponding to small eigenvalues, using the multigrid hierarchy, i.e., on coarse grids and leads to the following bootstrap procedure.

Once the initial multigrid hierarchy has been set up, we compute a set of vectors $\mathcal{V}_L = \{v_i^{(L)}\}$ on the coarsest grid, such that,

$$A_L v_i^{(L)} = \lambda_i^L T_L v_i^{(L)}, \lambda_i^{(L)} \in \mathbb{C}.$$

As the size of A_L is small, the eigenpairs $(v_i^{(L)}, \lambda_i^L)$ can be computed directly. Keeping in mind the relation between eigenvectors of the generalized eigenvalue problem (4.34) and the finest grid Rayleigh-Quotients observed in Lemma 4.37, a subset of the set of eigenpairs with the lowest eigenvalues can be used to build approximations to the eigenvectors with small eigenvalues on increasingly finer grids. At any coarse level l, we use the existing interpolation operator P_l^{l-1} to transfer these vectors to the next finer grid. Assuming that v_l is a solution to (4.34) on level l, i.e.,

$$A_l v_l = \lambda_l T_l v_l \Leftrightarrow \left(P_l^{l-1}\right)^H A_{l-1} P_l^{l-1} v_l = \lambda_l \left(P_l^{l-1}\right)^H T_{l-1} P_l^{l-1} v_l.$$

With this we obtain that $v_{l-1} = P_l^{l-1} v_l$ is an approximation to the generalized eigenvalue problem on grid $l-1$. Then on grid $l-1$ we apply a smoothing iteration to the eigenproblem

$$\left(A_{l-1} - \lambda_{l-1} T_{l-1}\right) v_{l-1} = 0. \tag{4.35}$$

Simultaneously, we also update the approximation to λ_{l-1} by

$$\lambda_{l-1} = \frac{\langle A_{l-1} v_{l-1}, v_{l-1} \rangle_2}{\langle T_{l-1} v_{l-1}, v_{l-1} \rangle_2}.$$

In fact, this procedure resembles an inverse Rayleigh-Quotient iteration found in eigenvalue computations (cf. [Wil65]), the only difference is that instead of computing the inverse of $A_{l-1} - \lambda_{l-1} T_{l-1}$ we approximate it by the application of an appropriate smoothing iteration.

Algorithm 4 below describes the process of enriching the set of test vectors. The process starts on the coarsest grid and proceeds to increasingly finer ones. On each grid we enhance the accuracy of the coarse-grid representations of finest-grid eigenvectors by relaxation on the eigenproblem (4.35).

Instead of enhancing the set of test vectors and recompute least squares interpolation only on the finest grid we can think of more sophisticated setup

Algorithm 4 Updating the set of test vectors {Multigrid eigensolver}

if The current level is the coarsest level **then**
  Take $\mathcal{V}_L = \{v_i^{(L)}\}_{i=1,\dots,k_u}$, s.t. $A_L v_i^{(L)} = \lambda_i^{(L)} T_L v_i^{(L)}$, $|\lambda_1^{(L)}| \leq \cdots \leq |\lambda_{k_u}^{(L)}|$
else
  Given $\mathcal{V}_{l+1}$ and P_{l+1}^l from the initial setup
  $v_i^{(l)} = P_{l+1}^l v_i^{(l+1)}, \quad \lambda_i^{(l)} = \lambda_i^{(l+1)}, \quad i = 1, \dots, k_u$
  Relax on $\left(A_l - \lambda_i^{(l)} T_l\right) v_i^{(l)} = 0$
  Calculate $\lambda_i^{(l)} = \frac{\langle A_l v_i^{(l)}, v_i^{(l)} \rangle_2}{\langle T_l v_i^{(l)}, v_i^{(l)} \rangle_2}$
end if

cycles. On each intermediate grid l we might want to consider to recompute the hierarchy $A_{l+1}, \dots, A_L$ using a subset or all of the improved representatives of algebraically smooth error contained in $\mathcal{V}_l$ before advancing further. In order to determine how to proceed we define a tool that allows us to measure the quality of the given hierarchy in terms of the approximation quality of algebraically smooth error computed in Algorithm 4.

Definition 4.38. *Given eigenvalues of the generalized eigenvalue problem*

$$A_j v^{(j)} = \lambda^{(j)} T_j v^{(j)},$$

on grids $j = l_1, l_2$, $l_1 < l_2$ with $\lambda^{(l_1)}, \lambda^{(l_2)}$, respectively, we define the eigenvalue approximation measure $\tau_\lambda^{(l_1,l_2)}$ *by*

$$\tau_\lambda^{(l_1,l_2)} = \frac{|\lambda^{(l_1)} - \lambda^{(l_2)}|}{|\lambda^{(l_2)}|}.$$

The eigenvalue approximation measure $\tau_\lambda^{(l_1,l_2)}$ is a useful tool to decide if a particular eigenvalue and correspondingly its eigenvector is accurately approximated by interpolation from grid l_1 to l_2 in the multigrid hierarchy. In chapter 5, we explain how this approach can be used to measure and control the accuracy of interpolation in the bootstrap setup.

In Figure 4.3 a possible setup cycle is visualized, at each blue dot, one has to decide whether to recompute P or advance to the next finer grid in the multigrid hierarchy. The illustrated cycle resembles a W-cycle, but any other cycling strategy can be applied to the setup.

4.3.2 Almost Zero Modes

Another adaptive multigrid technique first outlined by Brandt and Livne in [BL04] is the treatment of so-called *almost zero modes* (AZM). It is com-

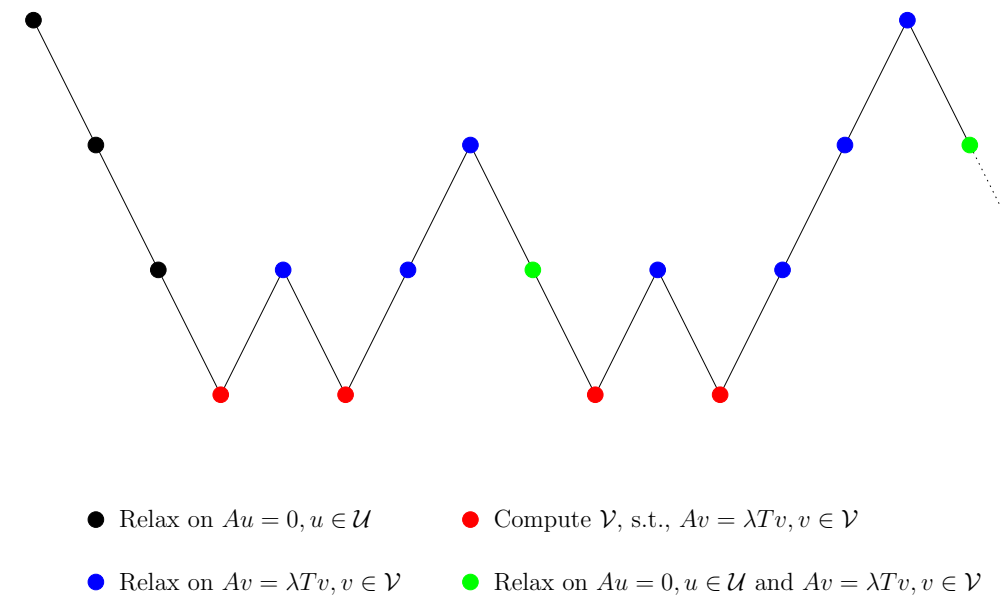


Figure 4.3: Bootstrap AMG setup W-cycle.

mon knowledge that algebraic multigrid methods can serve as excellent preconditioners for Krylov subspace methods, even if their stand-alone convergence is slow. This occurs when the algebraic multigrid error propagator has only a few small eigenvalues. These eigenvalues slow down the algebraic multigrid method as a stationary process, but a Krylov subspace (e.g., conjugate gradient) method can remove these outlying eigenvalues in just a few additional iterations.

The approach of AZMs computes the solution using the multigrid hierarchy, i.e., replacing additional work on the finest grid by the Krylov subspace method with additional work on coarse grids if possible.

In order to analyze this situation we consider a two-grid setting. Let $v_1, \ldots, v_m$ denote the eigenvectors of A with corresponding eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$ with,

$$Av_i = \lambda_i v_i.$$

Given an interpolation operator P, we define coarse grid matrices $A_c = P^H AP$ and $T_c = P^H P$. Then we denote the eigenvectors of the generalized eigenvalue problem

$$A_c x = \lambda T_c x$$

by $v_1^c, \ldots, v_{m_c}^c$ with associated eigenvalues $\lambda_1^c \leq \lambda_2^c \leq \ldots \leq \lambda_{m_c}^c$. Further we assume that

$$\|Pv_i^c\|_2 = 1$$

and also assume that the given interpolation fulfills

$$v_i \approx P v_i^c$$

for the algebraically smoothest vectors, i.e., the eigenvectors associated with small eigenvalues.

In that case analyzing the two-grid coarse-grid correction given by

$$\left(I - P A_c^{-1} P^H A\right) e \tag{4.36}$$

yields insight into the relationship of the respective eigenvalues λ_i and λ_i^c.

Lemma 4.39. *Let* $e = \sum_i \alpha_i v_i$, $P^H v_i = \sum_j \mu_{ij} v_j^c$ *and* $P v_j^c = \sum_k \gamma_{jk} v_k$. *With this* (4.36) *yields*

$$\left(I - P A_c^{-1} P^H A\right) e = \sum_i \left(\alpha_i - \sum_j \sum_k \alpha_k \frac{\lambda_k \mu_{kj} \gamma_{ji}}{\lambda_j^c} \right) v_i \tag{4.37}$$

Proof. We have

$$\begin{aligned}
\left(I - P A_c^{-1} P^H A\right) e &= \sum_i \alpha_i v_i - \sum_i \alpha_i \lambda_i \left(P A_c^{-1} P^H v_i\right) \\
&= \sum_i \alpha_i v_i - \sum_i \alpha_i \lambda_i P A_c^{-1} \left(\sum_j \mu_{ij} v_j^c \right) \\
&= \sum_i \alpha_i v_i - \sum_i \alpha_i \lambda_i P \left(\sum_j \frac{\mu_{ij}}{\lambda_j^c} v_j^c \right) \\
&= \sum_i \alpha_i v_i - \sum_i \sum_j \alpha_i \frac{\lambda_i \mu_{ij}}{\lambda_j^c} P v_j^c \\
&= \sum_i \alpha_i v_i - \sum_i \sum_j \sum_k \alpha_i \frac{\lambda_i \mu_{ij} \gamma_{jk}}{\lambda_j^c} v_k \\
&= \sum_i \alpha_i v_i - \sum_i \sum_j \sum_k \alpha_k \frac{\lambda_k \mu_{kj} \gamma_{ji}}{\lambda_j^c} v_i \\
&= \sum_i \left(\alpha_i - \sum_j \sum_k \alpha_k \frac{\lambda_k \mu_{kj} \gamma_{ji}}{\lambda_j^c} \right) v_i .
\end{aligned}$$

□

Assume now that μ_{ii} dominates the linear combination in $P^H v_i = \sum_j \mu_{ij} v_j^c$ and similarly γ_{jj} in $P v_j^c = \sum_k \gamma_{jk} v_k$. This is reasonable due to the assumption that interpolation approximately reproduces eigenvectors with small

eigenvalues on successive grids, i.e., $Pv_i^c \approx v_i$ and $P^H v_i \approx v_i^c$ for small i. In that case we can neglect many terms in (4.37). Assuming further that e is algebraically smooth, i.e., $e \approx \sum_{i=1}^{z} \alpha_i v_i, z < m$ we obtain

$$\left(I - PA_c^{-1}P^H A\right) e \approx \sum_{i=1}^{z} \left(\alpha_i - \alpha_i \frac{\lambda_i \mu_{ii} \gamma_{ii}}{\lambda_i^c}\right) v_i.$$

Hence, the "misfit" for each eigenvector component is given by

$$\left(I - PA_c^{-1}P^H A\right) v_i = \alpha_i \frac{\lambda_i^c - \lambda_i \mu_{ii} \gamma_{ii}}{\lambda_i^c} v_i. \tag{4.38}$$

Keeping in mind that $\mu_{ii} \approx \|P^H v_i\|_2 = 1$ and $\gamma_{ii} \approx \|P v_i^c\|_2 = 1$ and defining the eigenvalue approximation of λ_i by λ_i^c as

$$|\lambda_i - \lambda_i^c|,$$

we observe in (4.38) that the error

$$\frac{\lambda_i^c - \lambda_i \mu_{ii} \gamma_{ii}}{\lambda_i^c}$$

in the direction of a specific eigenvector is large, if the relative eigenvalue approximation

$$\frac{|\lambda_i^c - \lambda_i|}{|\lambda_i^c|}$$

is large. Thus, coarse-grid correction does a poor job in reducing this component of the error. It is easily seen that the absolute error margin that is allowed for each eigenvector component is inversely related to the associated eigenvalue. Therefore, a large relative error in eigenvalue approximation tends to occur more often with small eigenvalues, i.e., almost zero modes.

However, knowing v_i^c and approximately v_i and hence λ_i^c and λ_i in the bootstrap setup process offers a natural multigrid solution to the problem, fixing it on the coarse-grid by adjusting the contribution of the eigenvector v_i^c in the linear combination of the error e^c by the factor

$$\zeta_i = \frac{\lambda_i^c}{\lambda_i^c - \lambda_i \mu_{ii} \gamma_{ii}} \approx \frac{\lambda_i^c}{\lambda_i^c - \lambda_i}.$$

Hence, this treatment of almost zero modes only requires one additional projection onto the one-dimensional subspace spanned by v_i^c on the coarse grid.

The treatment of almost zero modes as proposed in this section requires careful bookkeeping. Note, that the treatment of almost zero modes is also a technique of general purpose and can be used together with the eigenvalue approximation property introduced in Definition 4.38 to measure and gauge the accuracy of any multigrid method with little overhead costs.

4.3.3 Optimally weighted coarse-grid correction

Another idea not directly related to the bootstrap mindset, but also a general purpose technique to improve multigrid methods is the analysis of optimal weighted coarse-grid correction, as it was discussed in [Van95, Bla88]. It is well known that the inclusion of over- and under-weighting in stationary iterative methods can yield huge benefits over the unweighted variants. Motivated by the two-grid analysis and the resulting convergence identity in [FVZ05] (cf. section 2.2) we analyze the optimally weighted coarse-grid correction in the presence of smoothing.

Given an interpolation operator P and Galerkin-based coarse-grid problem, the two-grid error propagation with post-smoothing by the smoother S is given by

$$e \mapsto S\left(I - \pi_A\left(P\right)\right)e. \tag{4.39}$$

We are interested in investigating the version of this error propagation operator where coarse-grid correction is weighted by $\omega \in \mathbb{R}$

$$e \mapsto S\left(I - \omega\pi_A\left(P\right)\right)e.$$

Hence, the problem of optimal weighted coarse-grid correction measured in the A-norm of the two-grid error propagation operator can be formulated as

$$\underset{\omega\in\mathbb{R}}{\operatorname{argmin}}\|S\left(I - \omega\pi_A\left(P\right)\right)e\|_A.$$

In order to derive the optimally weight $\omega^* = \omega^*(e)$ of coarse-grid correction we first state an auxiliary result.

Lemma 4.40. *Let $C \in \mathbb{C}^{m\times m}$ be hermitian and positive definite and $e, w \in \mathbb{C}^m$. Then we have*

$$\langle e - \omega w, e - \omega w\rangle_C \to \min! \quad \textit{for } \omega^* = \frac{\Re(\langle w, e\rangle_C)}{\langle w, w\rangle_C}. \tag{4.40}$$

Proof. We have

$$\langle e - \omega w, e - \omega w\rangle_C \;=\; \langle e, e\rangle_C - 2\omega\Re(\langle w, e\rangle_C) + \omega^2\langle w, w\rangle_C,$$

which yields (4.40). □

Theorem 4.41. *For given $e \in \mathbb{C}^m$ and two-grid error propagation* (4.39) *we have*

$$\underset{\omega\in\mathbb{R}}{\operatorname{argmin}}\|S\left(I - \omega\pi_A\left(P\right)\right)Se\|_A = \frac{\Re\left(\langle\pi_A(P)e, e\rangle_{S^HAS}\right)}{\langle\pi_A(P)e, \pi_A(P)e\rangle_{S^HAS}}. \tag{4.41}$$

Proof. We have $\|S(I - \omega\pi_A(P))e\|_A = \|I - \omega\pi_A(P)e\|_{S^H AS}$. Hence, we obtain (4.41) by application of Lemma 4.40. □

Interestingly enough the computation of $\omega^*(e)$ can be done on-the-fly during the solution phase with minor additional work needed. In order to see this we introduce the following notation

u	Current iterate
$\check{e} = P\left(P^H AP\right)^{-1} P^H(b - Au)$	Coarse-grid correction
$\hat{u} = u + \check{e}$	Iterate after coarse-grid correction
$u^+ = \hat{u} + M(b - A\hat{u})$	Smoothing iteration

The error $e = u - u^*$ used to define $\omega^*(e)$ is not known in the two-grid iteration, but the residual $r = Ae = A(u - u^*) = Au - b$ is. Some simple algebra gives

$$\begin{aligned}\omega^*(e) &= \frac{\langle \pi_A(P)e, e\rangle_{S^H AS}}{\langle \pi_A(P)e, \pi_A(P)e\rangle_{S^H AS}} \\ &= \frac{\langle S\pi_A(P)e, ASe\rangle_2}{\langle S\pi_A(P)e, S\pi_A(P)e\rangle_A}\,.\end{aligned}$$

For $S\pi_A(P)e$ we have

$$S\pi_A(P)e = SP\left(P^H AP\right)^{-1} P^T A(u^* - u) = SP\left(P^H AP\right)^{-1} P^H r = S\tilde{e}\,.$$

Hence, due to the fact that $\tilde{e}$ is available in the two-grid iteration, the only modification in order to compute $S\pi_A(P)$ is an additional application of the smoother to the coarse-grid correction $\tilde{e}$ with right-hand-side zero.

Further we observe that ASe is given by

$$\begin{aligned}ASe &= A(I - MA)(u^* - u) = A(u^* - u - Mr) \\ &= A(u^* - (u + Mr)) = A(u^* - \tilde{u}) = b - A\tilde{u} = \tilde{r}.\end{aligned}$$

This corresponds to the residual after application of smoothing to the current iterate and is available in the two-grid iteration.

With these observations the optimally weighted coarse-grid correction is given in Algorithm 5. Only one additional application of the smoother and an additional multiplication by A have to be done in order to compute the optimally weighted coarse-grid correction.

Another interesting observation relates optimally weighted coarse-grid correction to the smoothed interpolation approaches, e.g., smoothed aggregation (cf. section 2.6). Instead of taking the A-orthogonal projection as

Algorithm 5 amg_solve	{ Optimally weighted coarse-grid correction }
Input: A, f	
Output: u	
$u = \mathcal{S}^{\nu_1}(0, f)$	{Pre-smoothing}
$f_c = P^H (f - Au)$	{Restriction}
$u_c = \left(P^H A P\right)^{-1} f_c$	{Solve coarse-grid system}
$\tilde{e} = P u_c$	{Coarse-grid correction}
$u = \mathcal{S}_l^{\nu_2}(u, f)$	{Post-smoothing}
$r = f - Au$	{Residual after smoothing}
$\tilde{e} = \mathcal{S}_l^{\nu_2}(\tilde{e}, 0)$	{Post-smoothing}
$\omega^*(e) = \frac{\langle r, \tilde{e} \rangle_2}{\langle \tilde{e}, \tilde{e} \rangle_A^2}$	{Optimal Coarse-grid correction weight}
$u^+ = \tilde{u} + \omega^*(e)\tilde{e}$	{Next Iterate}
$r^+ = \tilde{r} - \omega^*(e) A\tilde{e}$	{Corresponding Residual}

the coarse-grid correction, taking the $S^H AS$-orthogonal projection $\pi_{S^H AS}(P)$ yields $\omega^*(e) = 1$ for all $e \in \mathbb{C}^m$ as

$$\|S(I - \pi_{S^H AS}(P))e\|_A \leq \|S(I - \omega \pi_{S^H AS}(P))e\|_A \text{ for all } \omega \in \mathbb{R}.$$

Furthermore, we have

$$S\pi_{S^H AS}(P) = \pi_A(SP)S.$$

Hence a $V(1,1)$-cycle using $\pi_{S^H AS}(P)$ is equivalent to a $V(2,0)$-cycle using the A-orthogonal projection onto $\mathcal{R}(SP)$, i.e., the smoothed interpolation operator.

Chapter 5

Numerical Results

In this final chapter, we present adaptive algebraic multigrid algorithms for the solution of the problems outlined in chapter 3 using the adaptive techniques we introduced in chapter 4. Furthermore, we numerically analyze various ideas of chapter 4 (e.g., algebraic distance, least squares interpolation, multigrid eigensolves, almost zero modes) in more detail.

We start with the prototypical model problem for algebraic multigrid methods, the Laplace operator, and demonstrate the capabilities of the proposed bootstrap approach in a situation with known multigrid solution. As a next step we introduce, using the notion of "gauge-equivalence", a class of problems that can be seen as an intermediate step in the transition from linear operators arising in discretizations of the (gauge-less) Laplace operator to the Gauge Laplace operator introduced in section 3.1.3. The first section is then concluded with the successful application of the bootstrap algebraic multigrid method to Gauge Laplace operators with physical gauge configurations of various temperatures and problem sizes. A roadmap towards the solution of the Wilson operator of Lattice QCD, introduced in section 3.1.1, is also presented.

One of the main difficulties encountered in this transition to the Wilson operator is that it is non-hermitian. Thus, in an intermediate step we develop an adaptive algebraic multigrid method, composed of the techniques introduced in chapter 4, to solve the non-symmetric eigenproblem encountered in Markov chain processes, as outlined in section 3.2. In here we demonstrate the capability of the approach to compute eigenvectors of non-hermitian systems to arbitrary accuracy in an adaptive algebraic multigrid fashion.

In the last section of this chapter we consider the Wilson-Schwinger operator of QED, that we introduced in section 3.1.2. We discuss the difficulties inherent in the problem formulation and possible limitations they might imply on the applicability of an adaptive algebraic multigrid method. We

compose an algorithm from the techniques presented in chapter 4 for this problem and discuss its application to these problems.

Finally, we give an outline on future research, involving further analysis and enhancement of the adaptive algebraic multigrid techniques of chapter 4 to solve the Wilson-Dirac system of Lattice QCD.

5.1 Scalar Elliptic PDEs

In this section we conduct numerical experiments using the techniques of the bootstrap framework we introduced in sections 4.2 and 4.3. Herein, we choose test problems with increasing difficulty to put several individual components of the algorithm to the test and analyze their performance. Starting with the prototypical model problem, the Laplace operator. In section 5.1.1, we focus on least squares interpolation and its variants by applying it to this first simple test problem. Thereafter, we demonstrate the application of the idea of algebraic distance using an anisotropic Laplace test problem in section 5.1.2. Finally, the section is concluded by the transition from Gauge Laplace operators with constant gauge configurations to Gauge Laplace operators with physical, randomized, gauge configurations in section, which appear in the Wilson-Schwinger formulation of Lattice Quantum Electrodynamics, in section 5.1.3.

5.1.1 Laplace operator Δ

We start the numerical analysis of our adaptive techniques by applying it to the two-dimensional partial differential equation

$$\begin{aligned} -u_{xx} - u_{yy} &= f(x,y), \quad (x,y) \in \Omega = (0,1)^2 \\ u(x,y) &= 0, \quad (x,y) \in \delta\Omega. \end{aligned}$$

In order to solve this problem numerically we consider a finite-element discretization, using quadrilateral finite-elements (cf. [SF73]), on a $(N-1) \times (N-1)$ equidistant grid. For each point on the grid the resulting element stiffness matrix is given as shown in Figure 5.1. Summation of the element stiffness matrices yields the linear operator A given by the stencil

$$A = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}.$$

Here, A is symmetric and positive definite (spd).

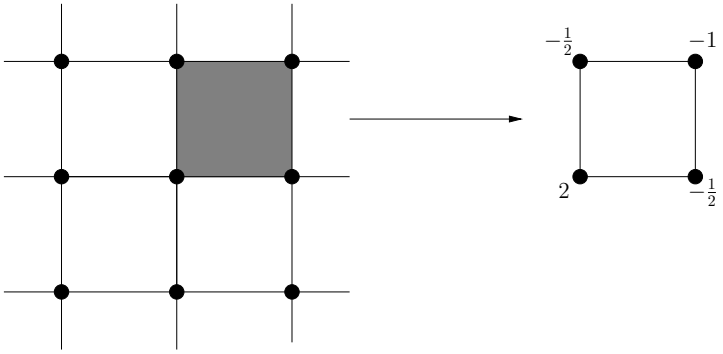


Figure 5.1: The entries of the element stiffness matrix associated with the quadrilateral finite-element discretization of Δu on an equidistant grid.

The results in Figure 2.3 suggest that the Gauss-Seidel iteration (2.5) fulfills a smoothing property, i.e., it reduces components of the error corresponding to large eigenvalues fast for this problem. Hence, we use this smoother throughout the remainder of this section if not stated otherwise.

We choose the splitting of the set of variables $\Omega = \{1, \ldots, N-1\}^2$ into $\mathcal{F}$ and $\mathcal{C}$ sets, such that $\mathcal{C}$ matches "full-coarsening", as depicted in Figure 5.2. In Table 5.1, we show that compatible relaxation yields fast asymptotic convergence rates independent of the problem size for this choice of coarse grid variables. In accordance with this choice of $\mathcal{C}$-variables, we limit the max-

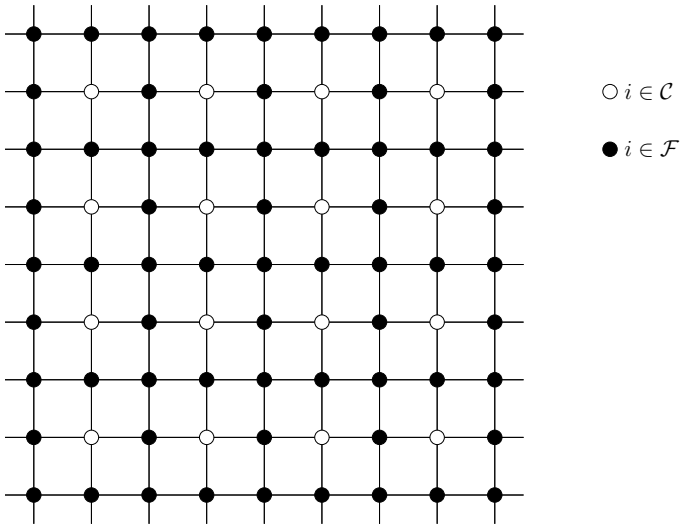


Figure 5.2: Coarsening pattern of full-coarsening on an equidistant two-dimensional grid.

N	31	63	127	255	511	1023
ρ	.2101	.2128	.2139	.2144	.2147	.2148

Table 5.1: Asymptotic compatible Gauss-Seidel convergence rates ρ with $\mathcal{C}_0$ matching "full-coarsening" for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements.

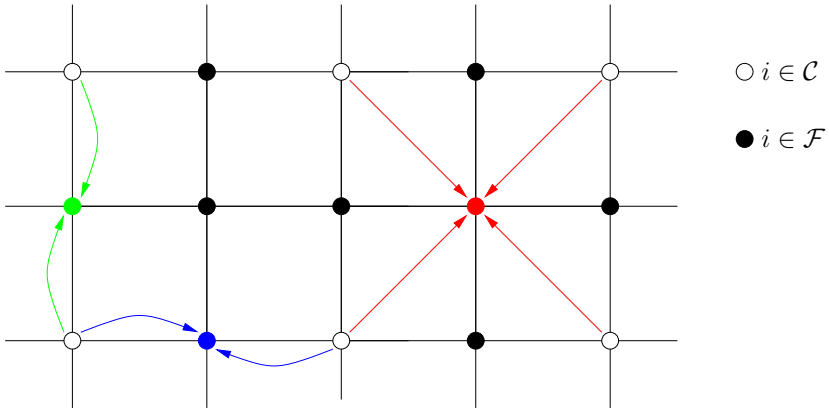


Figure 5.3: Interpolation relations. The drawn interpolation relations depict the maximal number of interpolatory points $\mathcal{C}_i$ for each class of $i \in \mathcal{F}$; in case less interpolatory points $\mathcal{C}_i$ are used, they are a subset of the depicted ones.

imal number of interpolatory points for each $i \in \mathcal{F}$ to 4 and only consider nearest neighbor interpolation (see Figure 5.3). We note that the coarse-level matrices

$$A_{l+1} = \left(P_{l+1}^{l}\right)^{H} A_l P_{l+1}^{l}$$

thus have at most 9 non-zero entries per row, i.e., the sparsity of the finest grid operator A is preserved.

The definition of interpolation operators P is done using least squares interpolation, introduced in section 4.2. We begin our experiments using least squares interpolation with initially random smoothed test vectors. Further, we demonstrate that least squares interpolation benefits from a priori knowledge on the algebraically smooth error, and move on to the bootstrap computation of improved test vectors. We end this section with a discussion of different cycling strategies in the bootstrap setup.

Note, that whenever we present asymptotic convergence rates, ρ, we report

$$\rho = \frac{\|e^{\nu}\|_2}{\|e^{\nu-1}\|_2},$$

in the last iteration where the solver terminates if the method reduced the initial residual by a factor of 10^{12} or applied $\nu = 100$ iterations and failed to converge to this tolerance.

In a purely algebraic setting, we do not assume any a priori knowledge about the algebraically smooth error components.

Hence, we begin our analysis using a set of initially random test vectors

$$u^{(1)}, \ldots, u^{(k)} \in \mathbb{R}^m,$$

where, each entry of these vectors is derived from a normal distribution with expectation zero and variance one (N$(0, 1)$ distribution). Then, we apply to each of these test vectors exactly η iterations of the given smoother, S, and compute least squares interpolation using the resulting approximations to the homogeneous equations. We limit these tests to a two-grid setting, with a fine grid of size 63×63 and accordingly a coarse grid of size 31×31. In Table 5.2 we report results for the asymptotic convergence of the V$(2, 2)$-cycle, varying the number of test vectors k and the number of iterations η applied to each test vector before computing least squares interpolation using weighting according to (4.11). As discussed in section 4.2, the minimal number of test vectors

η \ k	6	7	8	10	12
1	.972 .959	.971 .952	.967 .949	.966 .945	.962 .942
2	.928 .845	.901 .803	.902 .786	.884 .773	.859 .747
3	.834 .645	.802 .613	.782 .591	.750 .551	.725 .517
4	.744 .485	.694 .435	.659 .413	.609 .368	.586 .336
5	.610 .346	.549 .299	.526 .286	.473 .252	.452 .228
6	.500 .255	.422 .212	.408 .206	.366 .179	.346 .163
7	.440 .211	.345 .163	.326 .152	.289 .133	.272 .121
8	.360 .158	.272 .128	.258 .115	.226 .101	.216 .092

Table 5.2: Asymptotic convergence of the V$(2, 2)$ two-grid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a 63×63 grid using quadrilateral finite-elements. Using LS interpolation with k initially random ($N(0, 1)$ distributed) test vectors and η initial smoothings in the setup. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

$k_{\min}$ needed to define a well-posed least squares problem for all $i \in \mathcal{F}$ is given by

$$k_{\min} = \max_{i \in \mathcal{F}} |\mathcal{C}_i| = 4.$$

In practice, to prevent non-unique least squares problems for some $i \in \mathcal{F}$,we choose $k \geq \frac{3}{2} \max_{i \in \mathcal{F}} |\mathcal{C}_i|$. Note that, it is possible to adjust the number of test vectors on-the-fly if one encounters rank-deficient operators U_{w_i}.

As expected we see that increasing the number of test vectors, as well as increasing the number of smoothing iterations applied to the test vectors improves the convergence of the resulting two-grid method. We observe that the use of the residual correction is highly beneficial in this numerical experiment.

However, as can be seen in Table 5.3, with fixed number of test vectors $k = 8$ and smoothing iterations $\eta = 4$, the performance of the V$(2, 2)$-cycle method deteriorates when increasing the grid-size. This observation can be

N	31	63	127	255	511
ρ	.290 .145	.660 .413	.874 .718	.966 .913	.988 .975

Table 5.3: Asymptotic convergence of the V$(2, 2)$ two-grid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements. Using LS interpolation with 8 initially random ($N(0, 1)$ distributed) test vectors and 4 initial smoothings in the setup. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

explained by the fact that the smallest eigenvalue of the system decreases with increasing problem size. Keeping the number of iterations fixed, we are not able to expose the error components that correspond to the small eigenvalues to the same accuracy independently of the problem size, since the convergence of the smoother deteriorates as the smallest eigenvalue goes to zero.

It is known that classical AMG yields a very efficient solver for the problem we discuss here. This can be explained by the fact that interpolation in classical AMG preserves the constant vector $\mathbf{1}$, which is indeed algebraically smooth for discretizations of the Laplace operator. In Table 5.4 we present results, where we add the constant vector to the set of test vectors, and again increase the problem size, but fix the number of test vectors $k = 7 + 1$ and setup iterations $\eta = 4$. We see that by including the constant in the definition of least squares interpolation the method scales much better when increasing the problem size. This demonstrates that the least squares formulation straight-forwardly handles a priori knowledge about algebraically smooth error and benefits from such knowledge.

Motivated by the observation that the input of known algebraically smooth error, i.e., eigenvectors corresponding to small eigenvalues, improves the qual-

N	31	63	127	255	511
ρ	.125 .044	.130 .044	.158 .049	.194 .048	.182 .073

Table 5.4: Asymptotic convergence of the V$(2,2)$ two-grid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements. Using LS interpolation with 7 initially random ($N(0,1)$ distributed) test vectors plus the constant $\mathbf{1}$ and 4 initial smoothings in the setup. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

ity of least squares interpolation, we now consider the multigrid computation of approximations of eigenvectors to small eigenvalues, as was introduced in section 4.3.1. In order to demonstrate this bootstrap technique we move on from the two-grid method to a multigrid method. In Table 5.5, we first repeat the same test that we presented in Table 5.3 for a multigrid method with the coarsest grid always being 7×7. As expected, we clearly see slightly worse

N	31	63	127	255	511
ρ	.299 .159	.679 .440	.891 .750	.971 .925	.988 .976

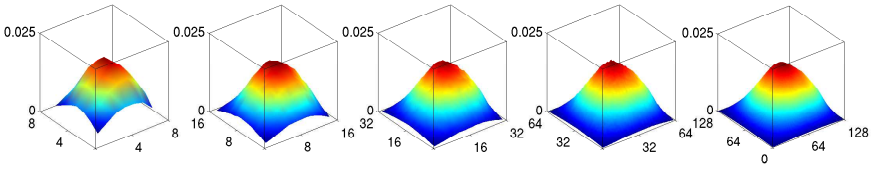
Table 5.5: Asymptotic convergence of the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements. Using LS interpolation with 8 initially random ($N(0,1)$ distributed) test vectors and 4 smoothings in the setup. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

results for the multigrid convergence as compared to the two-grid results presented in Table 5.3. However, the information condensed on the coarsest grid still contains very useful information. In Figure 5.4, we show for the problem of size 127×127 the coarsest grid eigenvector, v, corresponding to the smallest eigenvalue, λ, that fulfills the generalized eigenvalue problem

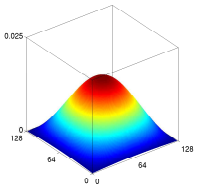
$$A_L v = \lambda T_L v \tag{5.1}$$

and its representations along the multigrid hierarchy according to Algorithm 4. The eigenvector approximations computed by Algorithm 4 resemble the targeted eigenvector on the finest grid, i.e., the eigenvector associated with the smallest eigenvalue of A.

Combining least squares interpolation and the recovery of representatives of algebraically smooth vectors from the multigrid hierarchy leads to a vast



(a) Multigrid representation of the coarsest grid eigenvector



(b) Finest grid eigenvector

Figure 5.4: Comparing the multigrid representation of the coarsest grid eigenvector corresponding to the smallest eigenvalue of the generalized eigenvalue problem (5.1) in 5.4(a) and the associated exact finest grid eigenvector corresponding to the smallest eigenvalue in 5.4(b).

variety of cycling-strategies. The most simple one, depicted in Figure 5.5, resembles a double-V-cycle. This adaptive setup substantially improves the

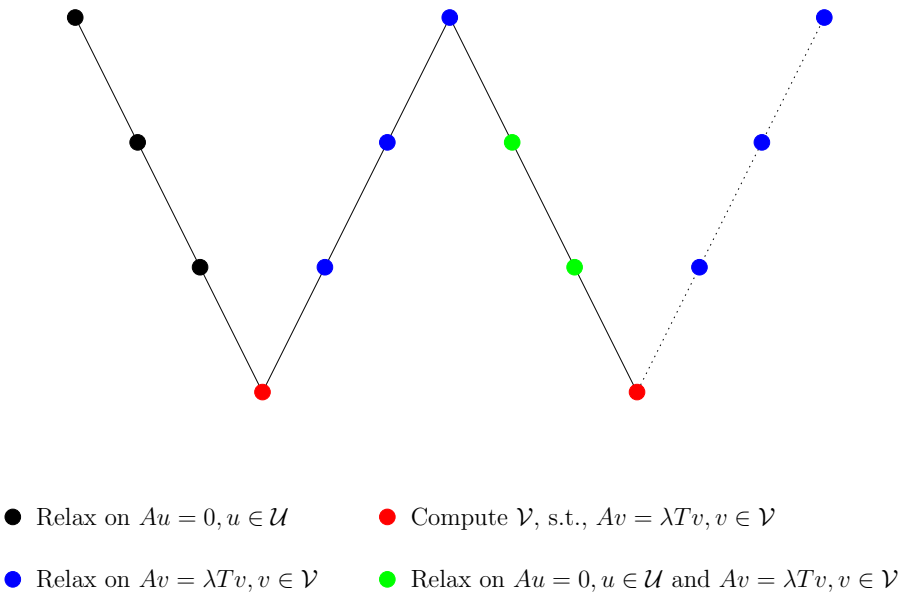


Figure 5.5: Bootstrap AMG setup cycle, double-V (V^2).

multigrid results given in Table 5.5, as can be seen in Table 5.6. In the

N	31	63	127	255	511
ρ	.048 .042	.062 .044	.110 .045	.143 .045	.910 .047

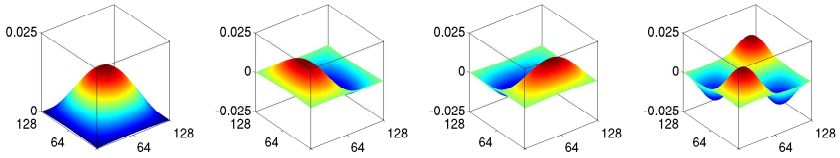
Table 5.6: Asymptotic convergence of the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements. Using LS interpolation with $k = 8$ initially random ($N(0,1)$ distributed) test vectors and $\eta = 4$ smoothings and a double-V-cycle setup (cf. Fig. 5.5) with $|\mathcal{V}| = k$ and 4 relaxations on the eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

adaptive setup we use the same number of iterations η to relax the initial test vectors and to relax on the generalized eigenequation (4.35) on the way back to the finest grid. We choose to compute $|\mathcal{V}| = k$ coarsest grid eigenvectors to the smallest eigenvalues of the generalized eigenproblem (4.34). Note, that although further computations are not needed the last time we visit the coarsest grid, i.e., we do not improve the multigrid hierarchy by these computations, we still may wish to compute coarsest grid eigenvectors and interpolate them to finer grids to measure the quality of the multigrid hierarchy, e.g., by computation of the eigenvalue approximation property. In Figure 5.6, we provide the approximations computed in the second V-cycle of the setup and the associated relative eigenvalue approximation measures $\tau_{\lambda}^{(L,1)}$ of Definition 4.38 are reported in Table 5.7.

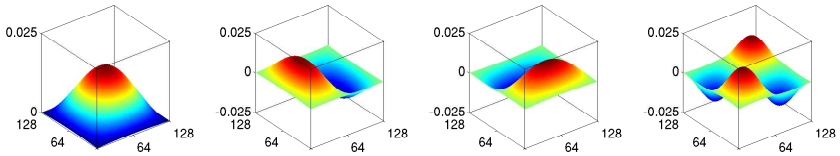
i	1	2	3	4	5	6	7	8
$\tau_{\lambda_i}^{(L,1)}$	.0163	.0316	.0335	.0415	.0776	.0814	.0757	.0762
$\|v_i^L - v_i\|_2$	.0013	.1314	.1311	.0073	.0184	.0214	.6743	.6756

Table 5.7: Relative eigenvalue approximation measures $\tau_{\lambda_i}^{(L,1)}$ of Definition 4.38 and eigenvector approximation estimates of the 8 smallest eigenvalues $\lambda_1 \leq \ldots \leq \lambda_8$ and corresponding eigenvectors $v_1, \ldots, v_8$, after the second V in a double-V-cycle setup (cf. Fig. 5.5) with $|\mathcal{U}| = |\mathcal{V}| = 8$ and 4 relaxations for $\mathcal{U}$ and $\mathcal{V}$. Results for the Laplace operator with Dirichlet boundary conditions, discretized on a 127×127 grid using quadrilateral finite-elements in a multigrid ($L = 5$) setting, including the residual correction (4.19) in the definition of LS interpolation.

We also introduce a W-cycle setup, illustrated in Figure 4.3, in which we recompute the multigrid hierarchy in a W-cycle fashion. Note that this setup resembles what is known in classical AMG literature as an F-cycle, once an initial multigrid hierarchy is computed. That is, the improvement of



(a) Finest grid representation of the coarsest grid eigenvectors



(b) Finest grid eigenvectors

Figure 5.6: Visualization of the finest grid representations of the coarsest grid eigenvectors corresponding to the 4 smallest eigenvalue of the generalized eigenvalue problem (5.1) in 5.6(a) and the associated exact finest grid eigenvectors corresponding to the 4 smallest eigenvalues in 5.6(b).

the multigrid hierarchy propagates from coarse to fine, recomputing the hierarchy, whenever improved information is available. In Table 5.8, we present results for the W-cycle setup using the same test setting as before.

N	31	63	127	255	511
ρ	.046 .042	.062 .044	.110 .045	.115 .045	.360 .046

Table 5.8: Asymptotic convergence of the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for the Laplace operator with Dirichlet boundary conditions, discretized on a $N \times N$ grid using quadrilateral finite-elements. Using LS interpolation with $k = 8$ initially random ($N(0,1)$ distributed) test vectors and $\eta = 4$ smoothings and W-cycle setup (cf. Fig. 5.5) with $|\mathcal{V}| = k$ and 4 relaxations on the eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

The last result to be presented in this section is related to the computational complexity of the adaptive setup. In Figure 5.7, we show the time, t, spent in the setup phase as a function of the number of variables, N^2. One can see that the computational complexity for both setup cycling strategies is $\mathcal{O}(N^2)$, i.e., linear in the number of variables. Therefore, the setup cannot prevent the overall method to achieve optimal complexity. We omit a discussion of the absolute setup time as our implementation is in MATLAB and

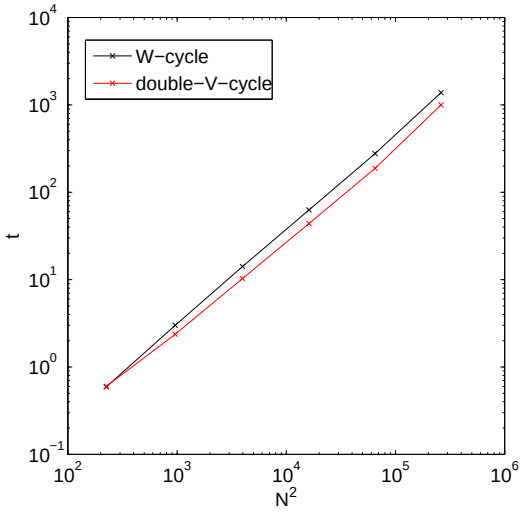


Figure 5.7: Computational complexity, illustrated by the logarithmic dependence of problem-size N^2 against the time the setup takes to finish, of both W- and double-V-cycle bootstrap AMG setups.

not optimized. Hence, a fair comparison to other methods is not appropriate. The discussion of relative cost of the setup in the solution process and the development of an optimized implementation is part of future research.

5.1.2 Anisotropic Laplace – Algebraic Distance

We introduced the concept of algebraic distance in section 4.2 as a way to generalize the notion of strength-of-connection using the least squares interpolation concept. In this section we present first numerical evidence of the discovery of strength-of-connection in an adaptive manner without using any a priori knowledge about the problem for a typical model problem. The typical model problem in classical AMG literature for the use of strength-of-connection is the anisotropic Laplace operator

$$L = \epsilon \partial_{xx} + \partial_{yy}, \quad \epsilon \ll 1. \tag{5.2}$$

We analyze this problem discretized on the unit square $\Omega = [0,1]^2$ with Dirichlet boundary conditions. The local stiffness matrices using stretched quadrilateral finite-elements are given in Figure 5.8 for $\epsilon = \frac{1}{100}$. By summation of the local stiffness matrices we obtain the linear operator A, written

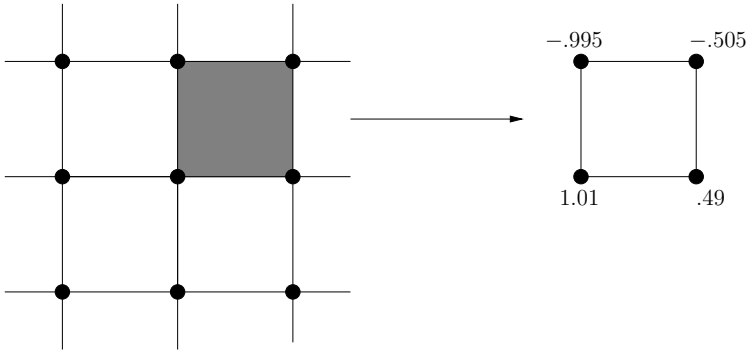


Figure 5.8: The entries of the element stiffness matrix associated with the stretched quadrilateral finite-element discretization of the anisotropic operator (5.2) for $\epsilon = \frac{1}{100}$ on an equidistant grid.

in stencil notation as

$$A = \begin{pmatrix} -0.505 & -1.99 & -0.505 \\ 0.98 & 4.04 & 0.98 \\ -0.505 & -1.99 & -0.505 \end{pmatrix}.$$

Here, we expect the strength of connection of variables on the grid in the x-direction to be much smaller than the strength of connection in the y-direction. We mentioned in section 4.2.2 that algebraic distance can be used to find such relations in a completely algebraic setting. Recall that algebraic distance from j to i for any pair of variables $i, j \in \Omega$, given a set of test vectors $\mathcal{U} = \{u^{(1)}, \ldots, u^{(1)}\}$ and associated weights $\omega_1, \ldots, \omega_k$, is defined by

$$d^{\alpha}_{i \leftarrow j} = \min_{\delta_{ij}} \sum_{l} \omega_l \left(u_i^{(l)} - \delta_{ij} u_j^{(l)} \right)^2 = \min_{\delta_{ij}} \| U_{\{i\}} W^{\frac{1}{2}} - \delta_{ij} U_{\{j\}} W^{\frac{1}{2}} \|_2^2. \quad (5.3)$$

That is, algebraic distance is given as the least squares functional $\mathcal{L}(p_i)$ of (4.10) for one-sided least squares interpolation from $\mathcal{C}_i = j$ to i. We limit the computation of algebraic distance to pairs that are close to each other in the graph G_A of A. For the anisotropic problem at hand, we consider only direct neighbors in the graph G_A, i.e., the algebraic distance matrix

$$D_A = (d_{i \leftarrow j})_{(i,j) \in E_A},$$

where E_A is the set of edges of G_A. Hence, D_A has the same sparsity as A. Even though we only care about the distinction of strong and weak couplings for each $i \in \Omega$ individually, i.e., the relative difference in the entries of the i-th row of D_A, we simplify this analysis by normalizing the algebraic distances

to $i \in \mathcal{F}$, such that the largest distance is 1. That is, we define a normalized algebraic distance $\hat{d}^{\alpha}_{i \leftarrow j}$ by

$$\hat{d}^{\alpha}_{i \leftarrow j} = \frac{1}{\max_k(d^{\alpha}_{i \leftarrow k})} d^{\alpha}_{i \leftarrow j}.$$

Then, using this normalized algebraic distance we define strong and weak couplings using a simple threshold θ similar to the classical formulation of strength-of-connection in Definition 4.21. Ultimately, the threshold can be chosen according to the histogram of connection strength, i.e., if globally strong and weak connections are well separated and can be identified by clusters in the histogram of $d_{i \leftarrow j}$.

In Figure 5.9, we present results that do not assume any a priori knowledge about algebraically smooth error, i.e., the set of test vectors contains 16 initially random $\mathrm{N}(0, 1)$ distributed, vectors, that are smoothed by 4 iterations of Gauss-Seidel before we compute algebraic distances according to (5.3) with ω_l chosen as in (4.11). We illustrate the distribution of the algebraic distances $\hat{d}^{\alpha}_{i \leftarrow j}$ in two histograms, on the bottom-left for the connections $i \leftarrow j$ below the chosen threshold θ and on the bottom-right for the connections $i \leftarrow j$ above the threshold. The connectivity graphs on top of the histograms represent all connections in the respective histogram, i.e., on the top-left, we show all connections $i \leftarrow j$ with $\hat{d}^{\alpha}_{i \leftarrow j} < \theta$ and on the top-right, connections $i \leftarrow j$ with $\hat{d}^{\alpha}_{i \leftarrow j} \geq \theta$. We see that algebraic distance predicts in this setting, with a few exceptions, both the strong couplings and also suggests the chosen threshold $\theta \approx .75$ that is used to distinguish strong and weak couplings by the clustering in the histograms.

We observed before in section 5.1.1 that better information of the algebraically smooth error in terms of approximations to the eigenvectors corresponding to small eigenvalues improves least squares interpolation and the quality of the associated multigrid method. Thus, we conduct a test with algebraic distance assuming exact knowledge of the 16 eigenvectors corresponding to the smallest eigenvalues. In Figure 5.10, we present the results of this test, using the same illustration as before, i.e., on the left, connectivity graph and algebraic distance histogram for algebraically close connections $i \leftarrow j$ and on the right, connectivity graph and histogram for the algebraically far connections $i \leftarrow j$ for this problem. With the improved knowledge of the algebraically smooth error, algebraic distance yields a very accurate result. We obtain an appropriate strong connectivity graph and a very pronounced distinction between strong and weak couplings as illustrated in the histograms.

Although, we only showed a first, and rather simple, test for the application of algebraic distances, we think that the results obtained are very

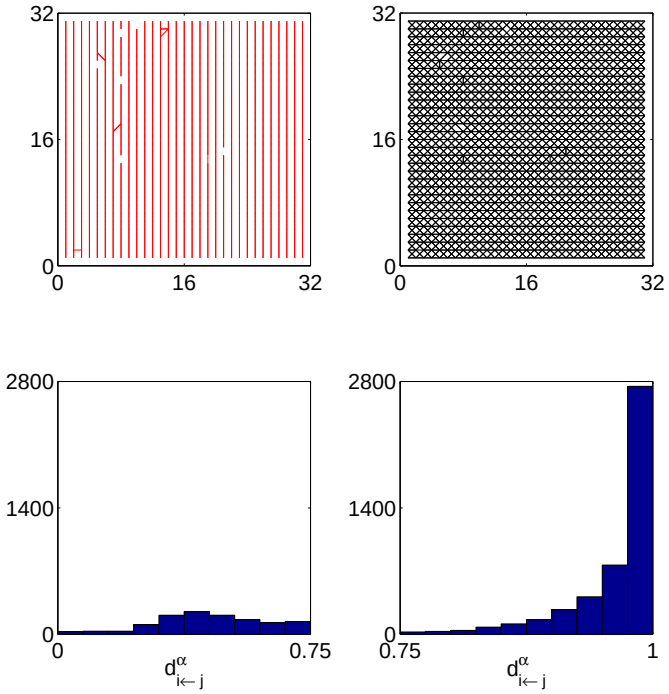


Figure 5.9: Algebraic distance defined by the 16 initially random ($N(0,1)$ distributed) vectors and 4 Gauss-Seidel relaxations for the anisotropic Laplace operator with $\epsilon = \frac{1}{100}$, discretized on a 31×31 grid using stretched quadrilateral finite-elements. The threshold is chosen according to the distribution seen in the histograms below, the corresponding connections in the graph of these subsets are depicted above.

encouraging. We note further that the notion of algebraic distance can be used as a tool in adaptive algebraic multigrid in the definition of interpolation and the discovery of variable relations. The next step in the analysis of algebraic distance could be the application to unstructured grids and variants of anisotropy, e.g., rotated anisotropies.

Another interesting problem of future work for algebraic distance is the discovery of the nature of variables in systems of PDEs, e.g., pressure and velocity in the Stokes problem or spins in Quantum Dynamics. It is common practice to build interpolation in algebraic multigrid for these problems based on the nature of the variable, but if the knowledge about the nature of the variables is not given a priori, algebraic distance might be useful in such problems in an adaptive and purely algebraic setting.

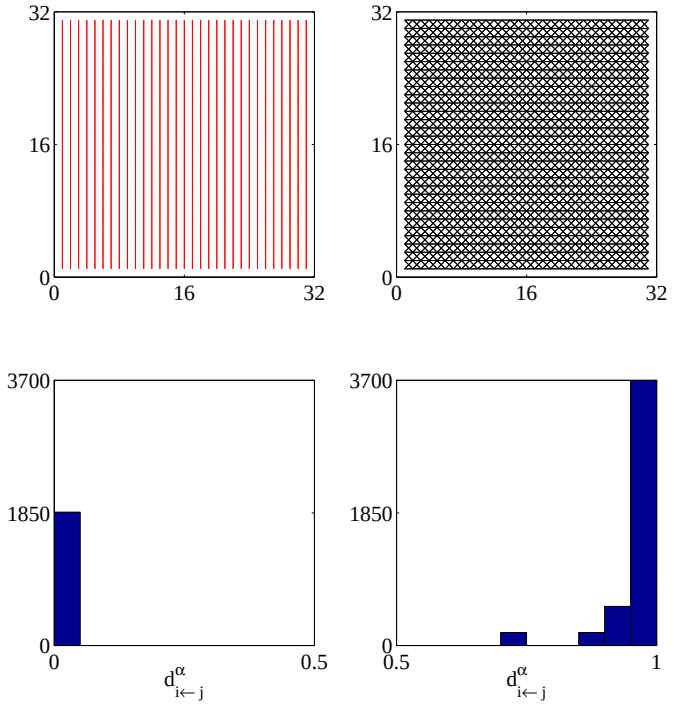


Figure 5.10: Algebraic distance defined by the 16 eigenvectors corresponding to the smallest eigenvalues of the anisotropic Laplace operator with $\epsilon = \frac{1}{100}$, discretized on a 31×31 grid using stretched quadrilateral finite-elements. The threshold is chosen according to the distribution seen in the histograms below, the corresponding connections in the graph of these subsets are depicted above.

Further, motivated by the results presented here, a natural application for algebraic distances is the computation of coarse variables $\mathcal{C}$ in a compatible relaxation algorithm. The notion of algebraic distance could be used to define better candidate sets $\mathcal{N}$ by refining the notion of slow-to-converge by averaging over several test vectors and also to define an improved strong connectivity graph that can be used to choose the right subset of the candidate set $\mathcal{N}$ to be added to $\mathcal{C}$. This modification accelerates Algorithm 2 by simplifying the computation of an independent set of $\mathcal{N}$ and improves its output. In addition this framework allows to incorporate a priori known information about the system into the adaptive coarsening process, but does not require any sense of algebraic smoothness to work.

After this short intermission, we now focus again on the development of an adaptive algebraic multigrid method for problems arising in Quantum

Dynamics.

5.1.3 Gauge Laplace $A(\mathcal{U})$ with $\mathcal{U} \subset U(1)$

Consider

$$(A(\mathcal{U}) + mI)\,\psi = f, \quad \psi, f \in \mathbb{C}^m, \tag{5.4}$$

with $A(\mathcal{U})$ denoting the Gauge Laplace operator on a $N \times N$ lattice with a background gauge configuration $\mathcal{U} \subset U(1)$ and m a mass term, to be specified later.

To illustrate the transition from a discretization of the Laplace Δ to the Gauge Laplace operator, $A(\mathcal{U})$. We recall the formulation of the Gauge Laplace operator for a given background gauge configuration

$$\mathcal{U} = \{U_\mu^x\} \text{ with } U_\mu^x \in U(1),\ \mu = 1, 2,\ x \in \{1, \ldots, N\}^2.$$

Explicitly using the covariant differences of Definition 3.3, we can obtain the discrete equation (5.4) at grid point $x \in \{1, \ldots, N\}^2$ as

$$(4+m)\,\psi_x - \left(U_1^{x-e_1}\right)^H \psi_{x-e_1} - U_1^x \psi_{x+e_1} - \left(U_2^{x-e_2}\right)^H \psi_{x-e_2} - U_2^x \psi_{x+e_1} = \varphi_x,$$

where e_i denotes the i-th canonical basis vector. In what follows, we denote this linear system of equations by

$$A(\mathcal{U})\psi = \varphi.$$

As the elements of $\mathcal{U}$ are complex numbers of modulus one, each element $U_\mu^x \in \mathcal{U}$ can be written as

$$U_\mu^x = e^{i\theta_\mu^x}, \quad \theta_\mu^x \in [0, 2\pi).$$

If the background gauge field is constant 1, i.e., $\theta_\mu^x = 0$ for all $U_\mu^x \in \mathcal{U}$, we obtain the finite difference discretization of the Laplace operator with periodic boundary conditions plus an additional mass term m. We then have the set of equations,

$$(4+m)\,\psi_{ij} - \psi_{i-1,j} - \psi_{i+1,j} - \psi_{i,j-1} - \psi_{i,j+1} = \varphi_{ij},$$

writing $x = (i, j) \in \{1, \ldots, N\}^2$ as usual. Hence, in stencil notation we obtain

$$A(\mathcal{U}_0) = \begin{pmatrix} & -1 & \\ -1 & (4+m) & -1 \\ & -1 & \end{pmatrix}.$$

Now, the transition from $A(\mathcal{U}_0)$ with a special constant gauge configuration that corresponds to the finite difference discretization of the Laplace operator to the physical Gauge Laplace operator $A(\mathcal{U})$ is done via an observation on constant gauge configurations $\mathcal{U}_c \neq \mathcal{U}_0$.

In what follows we fix the mass term m, such that the smallest eigenvalue of $A(\mathcal{U}) + mI$ on a $N \times N$ equidistant lattice is N^{-2}. We have already demonstrated that the bootstrap algebraic multigrid method devised in section 5.1.1 is able to deal with the finite-element discretization of Δ with Dirichlet boundary conditions. In Table 5.9, we present results of the method using a triple-V-setup cycle for the finite-difference discretization of Δ with periodic boundary conditions and a mass-shift $m = N^{-2}$. The coarsest grid is taken to be 8×8. Note that, we can again choose interpolatory points according to Figure 5.3, even though this means that some of the $\mathcal{F}$-variables are not interpolated from direct neighbors in this case. As

N	32	64	128	256
ρ	.082 .060	.082 .060	.110 .058	.147 .057

Table 5.9: Asymptotic convergence of the V(2, 2) multigrid cycle with Gauss-Seidel smoothing for $A(\mathcal{U}_0)$ (Laplace operator). Using LS interpolation with $k = 8$ initially random ($N(0, 1)$ distributed) test vectors and $\eta = 4$ smoothings in a triple-V–cycle setup with $|\mathcal{V}| = k$. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

expected, we obtain a very efficient multigrid solver for this problem.

As a next test we apply this same approach to other choices of constant gauge fields. In order to classify such operators we introduce the notion of a gauge transformation.

Definition 5.1. *A* gauge transformation

$$\begin{array}{rcl} g : \Omega & \rightarrow & SU(n) \\ x & \longmapsto & g_x \end{array}$$

of a gauge configuration $\mathcal{U} = \{U_\mu^x\} \subset SU(n)$ *is defined by*

$$U_\mu^x \longmapsto g_x^* U_\mu^x g_{x+e_\mu}.$$

That is, a gauge transformation can be represented as a block diagonal matrix

$$G = diag(g_x), \quad x \in \Omega.$$

Thus, the action of the gauge transformation on a gauge covariant operator $A(\mathcal{U})$ is given by

$$A(\mathcal{U}) \longmapsto G^H A(\mathcal{U}) G.$$

In the case of $\mathcal{U} \subset U(1)$ we specifically obtain the gauge transformation of $\mathcal{U}$ under the gauge transformation $g : x \longmapsto e^{i\psi_x}$ by

$$U_\mu^x = e^{i\theta_\mu^x} \overset{g}{\longmapsto} e^{-i\psi_x} e^{i\theta_\mu^x} e^{i\psi_{x+e_\mu}}. \tag{5.5}$$

Remark 5.2. *The $U(1)$ gauge transformation G of Definition 5.1 fulfills*

$$G^H G = I \quad \text{and} \quad \|G_{\cdot,i}\| = 1,\ i = 1, \ldots, m.$$

That is, the gauge transformation describes a unitary similarity transformation of the matrix $A(\mathcal{U})$. Hence, if $v_1, \ldots, v_m$ are the eigenvectors of $A(\mathcal{U})$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_m$ then $G^H v_1, \ldots, G^H v_m$ are the eigenvectors of $G^H A(\mathcal{U}) G$ to the same eigenvalues.

Motivated by Remark 5.2 we define an equivalence relation of gauge configurations.

Definition 5.3. *Define the binary relation $\sim$ between two gauge covariant operators $A(\mathcal{U}_1), A(\mathcal{U}_2)$, defined by gauge configurations $\mathcal{U}_1, \mathcal{U}_2$, by*

$$A(\mathcal{U}_1) \sim A(\mathcal{U}_2) \Longleftrightarrow$$
There exists a gauge transformation G s.t. $A(\mathcal{U}_1) = G^H A(\mathcal{U}_2) G$.

The binary relation $\sim$ is an equivalence relation. With the help of the equivalence classes with respect to $\sim$, we are able to classify Gauge Laplace operators $A(\mathcal{U})$ with constant gauge configurations $\mathcal{U}$ that are equivalent to the discretized Laplace operator $A(\mathcal{U}_0)$.

Theorem 5.4. *Consider a constant gauge configuration $\mathcal{U}$ on an $N \times N$ equidistant lattice, where each element $U_\mu^x \in \mathcal{U}$ is given by*

$$U_\mu^x = e^{i\theta}.$$

If θ fulfills

$$\theta = \frac{2\pi k}{N}, \quad \text{for some } k \in \{0, \ldots, N-1\},$$

we have $A(\mathcal{U}) \sim A(\mathcal{U}_0)$.

Proof. We have to find a gauge transformation G, such that

$$G^H A(\mathcal{U})G = A(\mathcal{U}_0).$$

By (5.5) we have for any gauge transformation g with $g_x = e^{i\psi_x}$,

$$(A(\mathcal{U}))_{x,x+\mu} = U^x_\mu = e^{i\theta} \overset{g}{\longmapsto} e^{-i\psi_x} e^{i\theta} e^{i\psi_{x+e_\mu}}$$

Now, we obtain the following system of linear equations by trying to fulfill $e^{i(\theta-(\psi_x-\psi_{x+e_\mu}))} = 1$ for all $x \in \{1,\dots,N\}^2$ and $\mu = 1,2$,

$$(\psi_x - \psi_{x+e_\mu}) = 2\pi\eta_x - \theta, \quad \eta_x \in \mathbb{Z}. \tag{5.6}$$

Writing $x = (x_1, x_2)$ and choosing $\psi_x = -(x_1 + x_2 - 2)\theta$ we obtain

$$\theta + (\psi_{x+e_1} - \psi_x) = 0, \text{ for all } x \text{ with } x_1 \in \{1,\dots,N-1\} \text{ and } x_2 \in \{1,\dots,N\}$$

and analogously,

$$\theta + (\psi_{x+e_2} - \psi_x) = 0, \text{ for all } x \text{ with } x_1 \in \{1,\dots,N\} \text{ and } x_2 \in \{1,\dots,N-1\}.$$

For all links $x = (N, x_2) \to x + e_1 = (1, x_1)$, we have for $x_2 \in \{1,\dots,N\}$

$$\theta + (\psi_{x+e_1} - \psi_x) = (1 - (1 + x_2 - 2) + (N + x_2 - 2))\,\theta = N\theta.$$

Hence, with (5.6) we obtain

$$\frac{\theta N}{2\pi} \in \mathbb{Z} \Rightarrow A(\mathcal{U}) \sim A(\mathcal{U}_0).$$

□

Now, the first interesting test is to check whether the bootstrap algebraic multigrid method devised in section 5.1.1 is gauge invariant, i.e., that it yields comparable results for constant gauge fields that are gauge equivalent to the Laplace operator. By Theorem 5.4 we obtain that on an $N \times N$ grid with N even, the operator given by its stencil notation,

$$A(\mathcal{U}_\pi) = \begin{pmatrix} & 1 & \\ 1 & (4+m) & 1 \\ & 1 & \end{pmatrix},$$

is gauge equivalent to the Laplace operator. In Table 5.10 we show results for this operator. Again we use a triple-V-cycle setup, a mass-shift m, such that $\lambda_{\min}(A(\mathcal{U}_\pi) + mI) = N^{-2}$ and coarsest grids of size 8×8. The results

N	32	64	128	256
ρ	.089 .059	.095 .059	.104 .059	.115 .057

Table 5.10: Asymptotic convergence of the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for $A(\mathcal{U}_\pi)$. Using LS interpolation with $k = 8$ initially random ($N(0,1)$ distributed) test vectors and $\eta = 4$ smoothings in a triple-V–cycle setup with $|\mathcal{V}| = k$. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

show that the adaptive setup is able to treat gauge equivalent problems equally well. As shown in Remark 5.2, the gauge equivalent problem can be interpreted as a diagonally scaled version of the original problem that leaves the relations of eigenvalues unchanged, but scales the eigenvectors such that the eigenvectors to small eigenvalues become highly oscillatory.

In Theorem 5.4 we discovered equivalence classes of Gauge Laplace operators that are potentially non-equivalent to the Laplace operator on a $N \times N$ grid. Next, we give results in Table 5.11 for a representative of a problem not equivalent to $A(\mathcal{U}_0)$ with the constant gauge configuration $\mathcal{U}_{\frac{\pi}{7}}$. The resulting linear operator in stencil notation is given by

$$A(\mathcal{U}_{\frac{\pi}{7}}) = \begin{pmatrix} & -e^{i\frac{\pi}{7}} & \\ -e^{i\frac{\pi}{7}} & (4+m) & -e^{i\frac{\pi}{7}} \\ & -e^{i\frac{\pi}{7}} & \end{pmatrix}.$$

Again, we use a triple-V-cycle setup and define the mass-shift m such that $\lambda_{\min}(A(\mathcal{U}_{\frac{\pi}{7}}) + mI) = N^{-2}$. In the tests, the coarsest grids are of size 8×8. The bootstrap algebraic multigrid setup yields an efficient solver for these

N	32	64	128	256
ρ	.066 .060	.058 .057	.061 .058	.058 .056

Table 5.11: Asymptotic convergence of the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for $A(\mathcal{U}_{\frac{\pi}{7}})$. Using LS interpolation with $k = 8$ initially random (complex, $N(0,1)$ distributed) test vectors and $\eta = 4$ smoothings in a triple-V-cycle setup with $|\mathcal{V}| = k$. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

systems as well.

Finally, we test Gauge Laplace operators $A(\mathcal{U})$ with physical background gauge configurations. The gauge configurations $\mathcal{U}$ of interest in the physical

simulations of quantum dynamics can be thought of as random, but matching a certain probability distribution that is affected by a temperature parameter β. In the limit case $\beta \to \infty$ we obtain $A(\mathcal{U}_0)$ the gauge-less Laplace operator. As mentioned in the introduction of the Gauge Laplace operator in section 3.1.3 and also observed in the discussion of numerical results for the adaptive reduction based approach in section 4.1.2, one of the main difficulties that the Gauge Laplace operator poses, is the local nature of its eigenvectors to small eigenvalues. In Figure 5.11 we present the modulus of the entries of the eigenvectors corresponding to the smallest 4 eigenvalues of $A(\mathcal{U})$ for a gauge configuration on a 64×64 grid at temperature $\beta = 5$. In

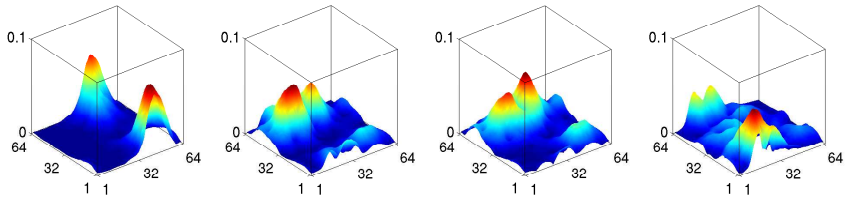


Figure 5.11: Entry-wise modulus of the eigenvectors corresponding to the smallest 4 eigenvalues of the discrete gauge Laplace operator $A(\mathcal{U})$ for a gauge configuration $\mathcal{U}$ on a 64×64 at temperature $\beta = 5$.

order to make sure that the usage of full-coarsening is a sensible choice for the Gauge Laplace operator at hand, we present in Table 5.12 asymptotic convergence rates of compatible relaxation for $\mathcal{C}_0$ matching the coarsening given by full-coarsening as illustrated in Figure 5.2. Compared to the results pre-

β \ N	32	64	128	256
1	.419	.420	.404	.402
5	.397	.398	.398	.400
10	.398	.401	.398	.399

Table 5.12: Asymptotic compatible Gauss-Seidel convergence rates ρ with $\mathcal{C}_0$ matching "full-coarsening" for the discrete Gauge Laplace operators $A(\mathcal{U})$ on an $N \times N$ grid for physical gauge configurations $\mathcal{U}$ at temperatures $\beta = 1, 5, 10$.

sented in Table 5.1 the asymptotic convergence rate of compatible relaxation for the Gauge Laplace operators $A(\mathcal{U})$ for physical gauge configurations $\mathcal{U}$ is worse, but still grid independent implying that full-coarsening is a suitable

grid for our method. The tests are reported in Table 5.13 using a multigrid framework, with a coarsest grid of 8×8. We use Gauge Laplace operators $A(\mathcal{U})$ with physical gauge configurations $\mathcal{U}$ on an $N \times N$ lattice at varying temperatures β and a mass shift m such that, $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. We report asymptotic convergence behavior of a $V(2,2)$-cycle solver and the number of iteration it takes a $V(2,2)$-cycle preconditioned conjugate gradient method to reduce the initial residual by a factor of 10^8.

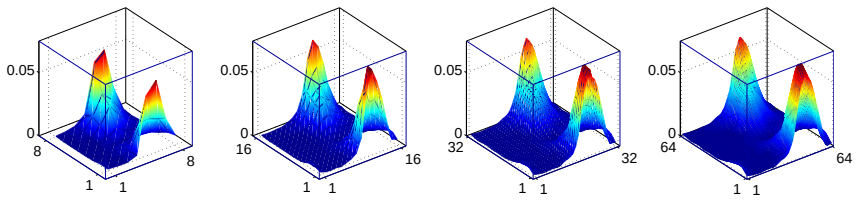
$\beta \backslash N$	32		64		128		256	
1	.284$_W$ 8	.242$_W$ 8	.416$_W$ 9	.286$_W$ 8	.648$_W$ 10	.478$_W$ 9	.658$_{V^3}$ 10	.493$_{V^3}$ 9
5	.275$_W$ 8	.284$_W$ 8	.264$_W$ 8	.225$_W$ 8	.576$_W$ 10	.389$_W$ 10	.672$_{V^3}$ 9	.469$_{V^3}$ 8
10	.137$_W$ 8	.120$_W$ 8	.225$_W$ 7	.223$_W$ 7	.586$_W$ 10	.349$_W$ 9	.433$_{V^3}$ 9	.423$_{V^3}$ 8

Table 5.13: Asymptotic convergence and number of iterations of preconditioned CG to reduce the norm of the residual by a factor of 10^8 for the V$(2,2)$ multigrid cycle with Gauss-Seidel smoothing for the discrete Gauge Laplace operator $A(\mathcal{U})$ for physical gauge configurations $\mathcal{U}$ on an $N \times N$ grid at varying temperatures β and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0,1)$ distributed) test vectors and $\eta = 4$ smoothings in a bootstrap setup that uses the cycling strategy given in green with $|\mathcal{V}| = 16$. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

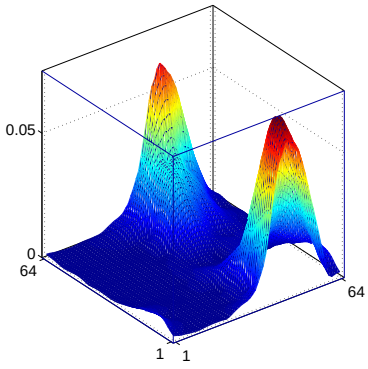
There are a couple of observations to be made. The lack of scaling of the stand-alone method when going to the larger grid-sizes can be explained by the increase of the number of locally supported eigenvectors to small eigenvalues. Due to the fact that any locally supported algebraically smooth vector cannot represent anything outside of its support, a fixed number of such vectors is not able to represent all of them, i.e., one would have to combine several non-overlapping locally supported vectors in order to reduce the number of representatives, though it is unclear how to weight such a linear combination of locally supported vectors in the definition of least squares interpolation. The results for the preconditioned conjugate gradient method suggest, however, that in the overall representation of the system by the multigrid hierarchy only few error components are not well represented. There are three possible ways to treat these error components. First, as we did in the tests, we could try to capture as many of the components in our adaptive setup and get the remaining components by wrapping the multigrid

solver into a Krylov subspace method (e.g., CG). Second, we can treat them as almost zero modes, a technique introduced in section 4.3.2. The last idea is to treat these locally supported vectors by a smoother that efficiently reduces error spanned by these vectors, an idea closer to the mindset of geometric multigrid. We should also remark that it is questionable if it is physically relevant to increase the grid-size while keeping the temperature fixed, as we did in our experiments. Overall, especially compared to the results we obtained with the adaptive AMGr approach, reported in section 4.1.2, we see that the bootstrap algebraic multigrid approach is able to generate an accurate multigrid representation of the Gauge Laplace operators, but sometimes fails to capture the whole subspace of algebraically smooth vectors.

In Figure 5.12 we show again the representation of the coarsest grid eigenvector corresponding to the smallest eigenvalue in the hierarchy as computed in the bootstrap computation of test vectors in Algorithm 4. We clearly see



(a) Multigrid representation of the coarsest grid eigenvector



(b) Finest grid eigenvector

Figure 5.12: Comparing the multigrid representation of the coarsest grid eigenvector corresponding to the smallest eigenvalue of the generalized eigenvalue problem (5.1) in 5.12(a) and the associated exact finest grid eigenvector corresponding to the smallest eigenvalue in 5.12(b).

the coarse grid representation of these local low modes.

With this positive result for the Gauge Laplace operators $A(\mathcal{U})$, we proceed towards more complicated gauged operators that appear in Quantum Dynamics. A natural next step is to consider the Wilson-Schwinger system of equations that we introduced in section 3.1.2 as the model of the interactions of electrons and photons in Quantum Electrodynamics. Recall that the discretization of the Schwinger operator with the Wilson stabilization has the Gauge Laplace operator $A(\mathcal{U})$ with $\mathcal{U} \subset U(1)$ gauge configurations as blocks on its diagonal. We analyze first approaches to solve this system in section 5.3. Once an adaptive algebraic multigrid method for the Wilson-Schwinger operator is found, there are actually two ways to proceed. On one hand, one could take the same route to the Wilson operator of Quantum Chromodynamics D_W as before to the Wilson-Schwinger operator via the Gauge Laplace operator of $SU(3)$ gauge configurations. On the other hand we also could apply the knowledge gathered in the development of an adaptive algebraic multigrid method for the Wilson-Schwinger S_W operator directly to the solution of the Wilson operator D_W.

5.2 Non-symmetric Eigensolver – Markov Chains

In this section we discuss the application of the bootstrap algebraic multigrid methodology to compute the steady-state vector of time independent Markov chains, as introduced in section 3.2.

The Markov chain model is described as a linear operator $A \in \mathbb{R}^{m\times m}$ which is column stochastic, i.e., $\mathbf{1}^T A = \mathbf{1}^T$. The steady-state vector x of this process is then, assuming that A is irreducible, the up to a scalar factor unique eigenvector to the eigenvalue one satisfying

$$Ax = x.$$

In what follows we reformulate this problem as an equivalent homogeneous system of linear equations

$$Bx = (I - A)\, x = 0,$$

instead.

The algorithm we develop in this section consists of two parts. First, we introduce what we call a *Multilevel Eigensolver* that resembles the multigrid eigensolver we presented in section 4.3.1 (cf. Algorithm 4). This step recomputes the multilevel hierarchy in each step, yielding better approximations

to the steady-state vector in each iteration and improving the multilevel hierarchy. The second part of the method makes use of a given multilevel hierarchy as a preconditioner for GMRES to compute the steady-state vector and thereby reduces the computational cost, required to recompute the multilevel hierarchy in each iteration.

Since the operators A and hence B are non-symmetric, we must modify the previously discussed approach for hermitian positive definite systems.

In what follows, we assume that a splitting of variables $\Omega_l = \mathcal{F}_l \cup \mathcal{C}_l$ is given on each level l. As usual, the multilevel hierarchy we use consists of a set of linear operators B_l and T_l, with $B_0 = B$ and $T_0 = I$. In contrast to the hermitian case discussed in section 5.1, we change the definition of these operators. We use the same least squares formulation as before to compute the interpolation operators P_{l+1}^l, but choose to define restriction operators Q_l^{l+1} as averaging operators that have the same sparsity as the corresponding interpolation operator, i.e., the defining properties of Q_l^{l+1} are given by

1. $\text{sparsity}(Q_l^{l+1}) = \text{sparsity}(\left(P_{l+1}^l\right)^T)$,
2. $\left(Q_l^{l+1}\right)_{i,j} = \frac{1}{c_j}$, for $(i,j) \in \text{sparsity}(Q_l^{l+1})$ and c_j the number of non-zeroes in column j.

With this definition of Q_l^{l+1} we have $\mathbf{1}^T Q_l^{l+1} = \mathbf{1}^T$. Then the coarse-grid operators B_l and T_l are canonically given by

$$\begin{aligned} A_l &= Q_{l-1}^l A_{l-1} P_l^{l-1}, \\ T_l &= Q_{l-1}^l T_{l-1} P_l^{l-1}, \\ B_l &= T_l - A_l = Q_{l-1}^l B_{l-1} P_l^{l-1}, \end{aligned}$$

for $l = 1, \ldots, L$. With this choice we have $\mathbf{1}^T B_l = 0$ for all levels l, i.e., thus preserving column stochasticity in the multilevel hierarchy.

5.2.1 Multilevel Eigensolver (MLE)

Our approach for building a multilevel method relies on the bootstrap framework introduced in section 4.3.1. We construct a sequence of increasingly coarser-level descriptions of the given fine-level Markov chain system using the least squares interpolation introduced in section 4.2.

The restriction of fine-level test vectors to the corresponding coarse-grid is done via injection, i.e., given a test vector u, we define its coarse-level representation u_c by $u_c = I_{|\mathcal{C}}^T u$. In Algorithm 6, we outline the multilevel eigensolver, which is very similar to Algorithm 4 that was used in the hermitian case to compute improved test vectors for least squares interpolation.

Algorithm 6 bootamg_mle {Bootstrap AMG MLE scheme}

Input: B_l, $(B_0 = B)$, T_l, $(T_0 = I)$, $\mathcal{U}_l$ (set of test vectors on level l)
Output: $\mathcal{U}_l, (\Lambda_l, \mathcal{V}_l)$, (approximations to the lowest eigenpairs)
if $l = L$ **then**
  Compute k lowest eigenvectors $\mathcal{V}_L$ and eigenvalues Λ_L of $B_L x = \lambda T_L x$.
else
  Relax $B_l x^{(j)} = 0,\ x^{(j)} \in \mathcal{U}_l$
  for $m = 1, \ldots, \mu$ **do**
    Compute P_l {Least squares interpolation}
    Compute Q_l {sparsity(Q_l) = sparsity(P_l), equiweighted}
    Set $B_{l+1} = Q_l B_l P_l$
    Set $T_{l+1} = Q_l T_l P_l$
    $\mathcal{U}_{l+1} = \{I_{|\mathcal{C}}^T x, x \in \mathcal{U}_l\}$ {Restrict test vectors}
    $\mathcal{V}_{l+1} = $ bootamg_mle$(B_{l+1}, T_{l+1}, \mathcal{U}_{l+1})$
    $\mathcal{V}_l = \{P_l x, x \in \mathcal{V}_{l+1}\}$ {Interpolate EV approximations}
    for $i = 1, \ldots, |\mathcal{V}_l|$ **do**
      Relax on $(B_l - \lambda_i T_l) x^{(i)} = 0,\ x^{(i)} \in \mathcal{V}_l$
      Calculate $\lambda_i = \dfrac{\langle B_l x^{(i)}, x^{(i)} \rangle_2}{\langle T_l x^{(i)}, x^{(i)} \rangle_2}$
    end for
  end for
end if

Ultimately, in our multilevel eigensolver approach, the sequence of prolongation operators P_{l+1}^l needs to be accurate for only the smallest eigenvector – the kernel of the finest-level operator $I - A$. The accurate resolution of many other near-kernel components by our least squares based interpolation operator allows for the effective reduction of these "unwanted" error components in the regular algebraic multigrid coarse-grid correction, while the steady-state vector, as the kernel of $I - A$, is preserved in such correction steps. Using this observation the reduction of "unwanted" modes can further be accelerated by the use of a multigrid preconditioned Krylov subspace method (e.g. GMRES).

5.2.2 Multigrid preconditioned GMRES

The multilevel eigensolver steps described above yield increasingly better approximations to the steady-state vector and other vectors that cannot be removed efficiently by the multigrid relaxation method.

Although we are only interested in computing the steady-state vector, the

multigrid hierarchy based on least squares interpolation is able to resolve a larger subspace. This leads to the idea of exploiting this richness of the given hierarchy for use in multigrid correction steps – in addition to the discussed MLE steps.

To illustrate the effect of multigrid correction steps applied to the homogeneous problem $Bx = 0$, we start by analyzing simple relaxation schemes for the steady-state problem, $Ax = x$. As the steady-state solution is the eigenvector corresponding to the eigenvalue with largest absolute magnitude, a power iteration

$$x^{k+1} = Ax^k$$

is guaranteed to converge to the solution if A is irreducible and thus, due to Theorem 3.25 only one eigenvalue with absolute value one exists and all other eigenvalues have modulus less than one. However, convergence can be slow if A has other eigenvalues close to one in absolute value.

Such power iterations applied to the steady-state problem are in turn equivalent to applying a Richardson iteration to the homogeneous system $(I - A)\, x = 0$. A natural modification, using the fact that the field of values of A is contained in the unit circle, is then given by a suitable under-relaxed iteration, yielding the error propagator

$$e^{k+1} = (I - \tau B)\, e^k. \tag{5.7}$$

Rewriting (5.7) we have

$$x^{k+1} = ((1 - \tau)\, I + \tau A)\, x^k,$$

which can be interpreted as a modified power method.

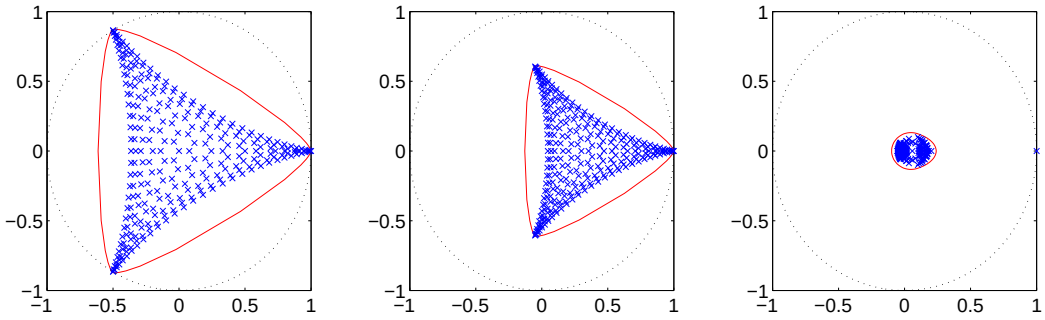
In Figures 5.13(a) and 5.13(b) the spectra of A and of $I - \tau\,(I - A)$ are depicted for a characteristic two-dimensional test problem, along with the field of values. Applying this formalism to the error propagator of the two-grid V$(1,1)$-cycle for B given by

$$E_{tg} = (I - MB)\left(I - PB_c^{\dagger}QB\right)(I - MB)\,, \tag{5.8}$$

yields yet another modified iteration. Herein, $B_c^{\dagger}$ is a suitable pseudo-inverse of the coarse-grid operator B_c. Choosing an appropriate operator $\mathcal{Z}$, we can rewrite (5.8) as

$$I - \mathcal{Z}B.$$

That is, application of the Multigrid V-cycle can be interpreted as a power iteration applied to the preconditioned matrix $I - \mathcal{Z}B = I - \mathcal{Z} + \mathcal{Z}A$. In Figure 5.13(c), the spectrum and field of values (see Definition 5.9 for details)



(a) Spectrum and field of values for Richardson error propagator.

(b) Spectrum and field of values for τ-Richardson, $\tau = .7$.

(c) Spectrum and field of values of Multigrid error propagator.

Figure 5.13: Spectra and field of values of the Richardson, τ-Richardson and Multigrid error propagators for two-dimensional tandem queuing problem on a 33×33 grid.

of the multigrid V(2, 2)-cycle preconditioned matrix is depicted. For this case, it is clear that applying power iterations to this preconditioned operator will converge rapidly to the steady-state vector.

To further accelerate this approach, instead of the straight-forward power method, we consider multigrid preconditioned GMRES steps. As we show below, the convergence of GMRES is guaranteed for our Markov chain systems.

Lemma 5.5. *Let A be a column-stochastic irreducible operator and $B = I - A$. Then we have*

$$\mathcal{R}(B) \cap \mathcal{N}(B) = \{0\}. \tag{5.9}$$

Proof. As A is column-stochastic we know that $\mathbf{1}^T A = \mathbf{1}^T$. Hence, $\mathbf{1}^T B = 0$. Furthermore, we know from the Perron-Frobenius theorem that the null-space of B is one-dimensional, spanned by a strictly positive (component-wise) vector x^*. For all $y \in \mathcal{R}(B)$, $y = Bz$, we have

$$\langle \mathbf{1}, y \rangle = \langle \mathbf{1}, Az \rangle = \langle A^t \mathbf{1}, z \rangle = 0.$$

With this and $\langle \mathbf{1}, x^* \rangle \neq 0$ (as x^* is strictly positive) we get (5.9). □

In [HS09, Theorem 2.8] it is shown that GMRES determines a solution of $Bx = b$ for all $b \in \mathcal{R}(B), x_0 \in \mathbb{R}^m$ iff $\mathcal{R}(B) \cap \mathcal{N}(B) = \{0\}$. Due to Lemma 5.5 the assumptions of this theorem are fulfilled for $B = I - A$, where A is a column-stochastic and irreducible.

Remark 5.6. *Although, we now have theoretical reason to apply GMRES as an iterative solver to the steady-state Markov chain problem, the situation when using a multigrid preconditioner is more complex. In this case, we have to prove that the assumption of Lemma 5.5 are still fulfilled for the preconditioned operator $E_{tg}B$. Unfortunately, we cannot guarantee in general that the bootstrap algebraic multigrid method does fulfill* (5.9)*, though our numerical experience suggests that it is viable to pursue this approach.*

With the theoretical justification for the use of GMRES for these types of problems, we present numerical tests. We limit our discussion to three Markov chain models that can be represented by planar graphs. Each of the models has certain interesting characteristics that pose problems for the solver. We begin our experiments with a very simple model.

The uniform two-dimensional network can be seen as the Markov chain analogue of the Laplace operator. It is defined on an equidistant grid Ω of size $N \times N$. We denote this grid in graph notation as $G_\Omega = (V_\Omega, E_\Omega)$. The entries of A are then given as

$$a_{i,j} = \begin{cases} \frac{1}{d_{out}(j)}, & \text{if } (i,j) \in E_\Omega \\ 0, & \text{else,} \end{cases}$$

where $d_{out}(j)$ is the number of outgoing edges of $j \in \Omega$. In Figure 5.14, we illustrate the two-dimensional uniform network problem. In the tests we conduct, we use again full-coarsening, i.e., we choose $\mathcal{C}$ to be every other grid point in both spatial dimensions, as illustrated in Figure 5.2. In order to keep the overall invested work small, we consider to take only up to two

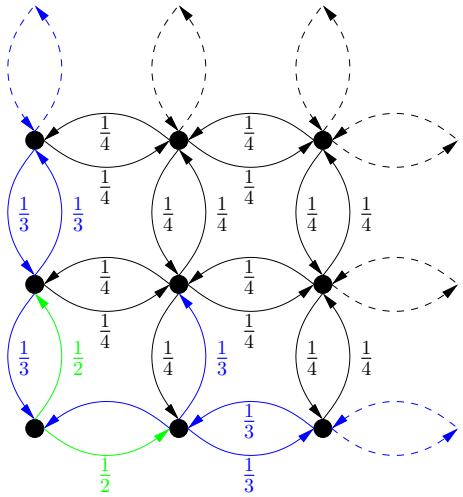


Figure 5.14: Uniform network model on a two-dimensional equidistant grid.

interpolatory points $\mathcal{C}_i$ per point $i \in \mathcal{F}$ determined by the greedy strategy introduced in section 4.2.2. As usual we define $\mathcal{F} = \Omega \setminus \mathcal{C}$. Hence, we use fewer test vectors as well. As we know that the steady-state vector we want to compute is strictly positive, we choose to use initially random, but positive test vectors for this first problem, and also in all following tests. More precisely, we choose vectors with entries uniformly distributed in $[1, 2]$.

In Table 5.14, we present results that use 6 test vectors and a $V(2,2)$-cycle MLE step with ω-Jacobi smoother, $\omega = .7$. We report the number of iterations needed to compute the steady-state vector x, such that

$$\|Bx\|_2 \leq 10^{-8}, \|x\|_2 = 1.$$

In addition the number of preconditioned GMRES iterations needed to achieve the same accuracy, where we report the initial MLE setup cycle in the subscript. The coarsest grid in the experiments is always 5×5. The operator

N	17		33		65		129	
MLE	9	8	10	8	11	9	11	9
pGMRES	10_V	10_V	12_V	11_V	14_V	12_V	18_{V^2}	16_V

Table 5.14: Multilevel results for the two-dimensional uniform network model on an $N \times N$ grid. We report results to compute the steady-state vector x to an accuracy of 10^{-8}, using a $V(2,2)$-MLE cycle with ω-Jacobi smoothing, $\omega = .7$. In addition we also report the number of iterations pGMRES needs to achieve the same accuracy, where we denote the initial bootstrap setup in the subscript. The sets $\mathcal{U}$ and $\mathcal{V}$ consist of 6 initially positive random vectors and coarsest grid eigenvectors, respectively. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

complexity in these tests is bounded by 1.6. The test shows that the MLE and the preconditioned GMRES approach both yield methods that scale when increasing the problem-size. Note, that one step of pGMRES is much cheaper than one step of MLE. We consider restarting pGMRES if the solution is not found in a reasonable number of iterations (e.g., smaller than 32).

The next Markov chain model we consider in our tests is a tandem queuing network, illustrated in Figure 5.15. The spectrum of this operator is complex, as we show in Figure 5.13(a). Again, we use full-coarsening and a coarsest grid of 5×5. We present our results in Table 5.15. In the test we use the following probabilities,

$$\mu = \frac{11}{31}, \quad \mu_x = \frac{10}{31}, \quad \mu_y = \frac{10}{31}.$$

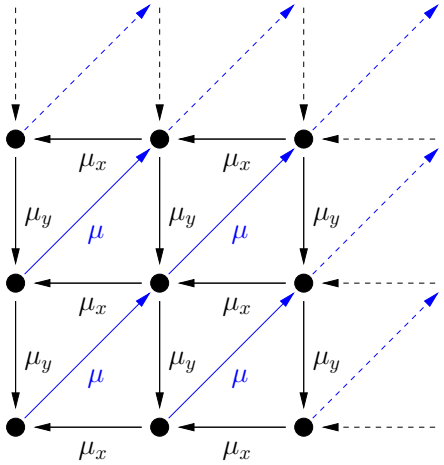


Figure 5.15: Tandem-Queuing Network with probability to advance μ and queuing probabilities μ_x and μ_y on a two-dimensional equidistant grid.

N	17		33		65		129	
MLE	8	6	8	6	8	6	8	6
pGMRES	8_{V^2}	8_{V^2}	8_{V^2}	8_{V^2}	8_{V^2}	8_{V^2}	8_{V^2}	9_{V^2}

Table 5.15: Multilevel results for the tandem queuing network model on an $N \times N$ grid. We report results to compute the steady-state vector x to an accuracy of 10^{-8}, using a $V(2,2)$-MLE cycle with ω-Jacobi smoothing, $\omega = .7$. In addition we also report the number of iterations pGMRES needs to achieve the same accuracy, where we denote the initial bootstrap setup in the subscript. The sets $\mathcal{U}$ and $\mathcal{V}$ consist of 6 initially positive random vectors and coarsest grid eigenvectors, respectively. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

Again, we see that the MLE method converges rapidly to the steady-state vector and also yields a very efficient preconditioner for the GMRES method. We observe that the number of iterations does not depend on the size of the problem. Note that the MLE method also yields accurate approximations to the eigenvectors corresponding to all k smallest eigenvalues on the finest grid. In Figure 5.16, we show the computed approximations and report in Table 5.16 their accuracy upon convergence of the steady-state vector. One should keep in mind that the results are intended to show the promise of our approach rather than presenting an optimized method in the bootstrap framework for this type of problems. We limit our analysis to the statement that with minimal effort spent in adjusting the parameters of the setup of

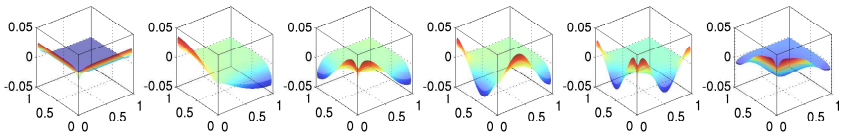


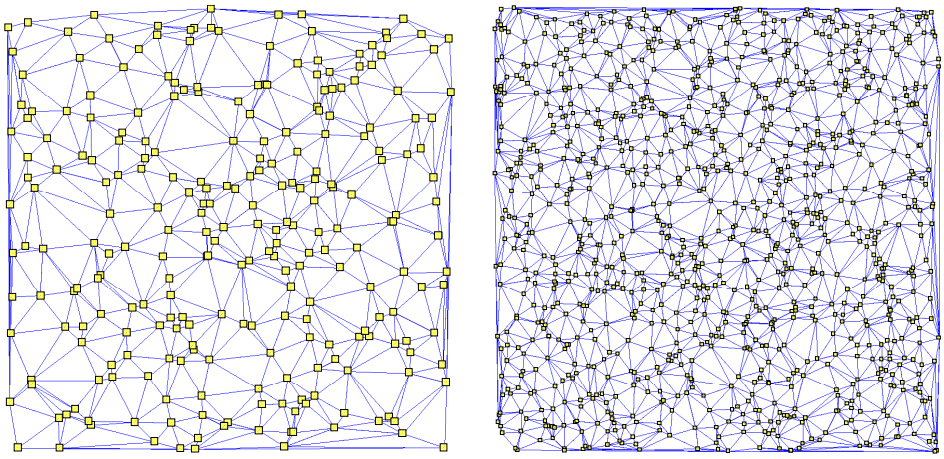
Figure 5.16: Approximations to the eigenvectors corresponding to the smallest 6 eigenvalues of the tandem queuing network problem on a 129×129 grid with 6 levels upon convergence of the steady-state solution of the MLE method.

i	1	2	3	4	5	6
$\|v_i^L - v_i\|_2$	$1.02E{-}8$	$9.04E{-}3$	$2.68E{-}2$	$2.84E{-}2$	$1.42E{-}1$	$3.77E{-}1$

Table 5.16: Accuracy of the eigenvectors corresponding to the smallest 6 eigenvalues of the tandem queuing network problem on a 129×129 grid with 6 levels upon convergence of the steady-state solution of the MLE method.

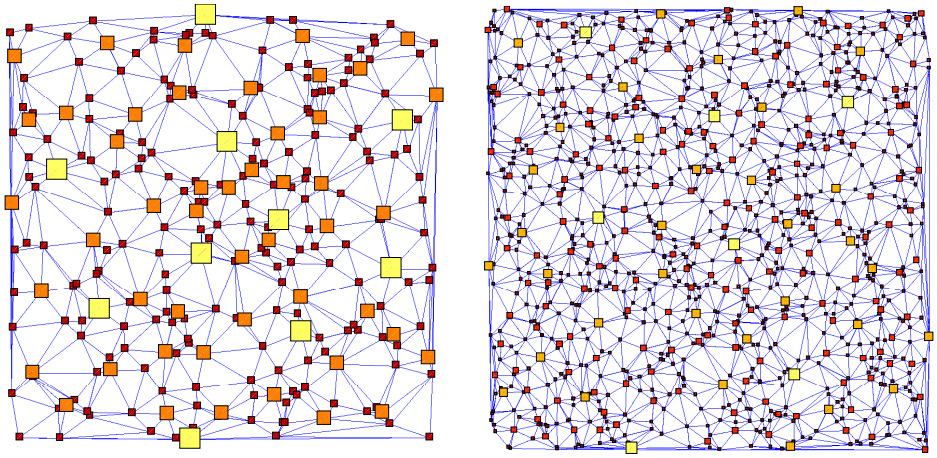
the bootstrap approach we obtain scalable solvers. The optimization of the method by tuning the parameters involved, e.g., relaxation parameter of the smoother, coarsening, caliber, number of test vectors, weighting in the least squares interpolation, number of relaxations in the setup and solution process, is part of future research.

The last test we consider corresponds to a triangulation of N randomly chosen points in $[0,1] \times [0,1]$. The transition probabilities in the network are then given by the number of outgoing edges at each point, similar to the uniform network. In Figure 5.17, we show two examples of such networks, one with $N = 256$ and one with $N = 1024$ points. Due to the fact that the corresponding graphs of this model are planar, we call this model *unstructured planar graph* model. As there is no natural way to define the set of coarse variables $\mathcal{C}$ for this problem we use compatible relaxation, introduced in section 2.7 to define the splitting of variables $\Omega = \mathcal{F} \cup \mathcal{C}$. In Figure 5.18, a resulting coarsening is presented. The size of each individual point represents how many grids it appears on. In Table 5.17, we present results of our methods to deal with the unstructured planar graph model for a selection of graphs. Even for this unstructured graph network we obtain a fast converging method with our MLE approach. The variation in the results when increasing the problem-size might be caused by the fact that by increasing the problem-size the nature of the problem changes in the sense that the average number of outgoing edges of grid points increases. That is,



(a) Random planar graph with $N = 256$ (b) Random planar graph with $N = 1024$

Figure 5.17: Delaunay-triangulations of N randomly chosen points in the unit square $[0, 1] \times [0, 1]$.



(a) Random planar graph with $N = 256$ (b) Random planar graph with $N = 1024$

Figure 5.18: Coarsening of Delaunay-triangulations of N points randomly scattered in the unit square $[0, 1] \times [0, 1]$ shown in Figure 5.17 using compatible relaxation (cf. Algorithm 2).

it is not necessarily clear whether two unstructured graphs of different sizes are comparable.

The proposed approach for the solution of the steady-state Markov chain problems can be viewed as a combination of several ideas. On one hand, we applied the bootstrap algebraic multigrid framework with least squares

N	256		512		1024		2048	
MLE	13	12	19	16	20	20	20	20
pGMRES	11_V	10_V	13_V	13_V	14_V	14_V	17_V	16_V

Table 5.17: Multilevel results for the unstructured planar graph model on an $N \times N$ grid. We report results to compute the steady-state vector x to an accuracy of 10^{-8}, using a $V(2,2)$-MLE cycle with ω-Jacobi smoothing, $\omega = .7$. In addition we also report the number of iterations pGMRES needs to achieve the same accuracy, where we denote the initial bootstrap setup in the subscript. The sets $\mathcal{U}$ and $\mathcal{V}$ consist of 6 initially positive random vectors and coarsest grid eigenvectors, respectively. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

interpolation and bootstrap multilevel eigensolver. Similar multilevel computations of steady-state vectors for Markov chain problems were introduced in an aggregation-based framework of aggregation-disaggregation approaches (cf. [Sch91]) and further developed in [DSMM+09]. Yet these approaches focused only on the steady-state vector and did not re-use the multigrid hierarchy. On the other hand, we have the use of algebraic multigrid as a preconditioner for a suitable Krylov subspace method (e.g., GMRES). This idea was pursued earlier in [Vir07] using a reduction-based approach to define an appropriate algebraic multigrid preconditioner. In this sense, our approach in this context can also be seen as a mix of these ideas, adding an adaptive and robust way to define the algebraic preconditioner to the composition with the bootstrap algebraic multigrid framework and least squares interpolation.

5.3 Non-Hermitian Systems of PDEs – The Wilson-Schwinger System

In the last section of numerical experiments we consider applying the bootstrap algebraic multigrid method to the Wilson-Schwinger model of Quantum Electrodynamics, introduced in section 3.1.2. There are several changes to the algorithm that are needed when applying this approach to the Wilson-Schwinger problem.

First, we recall properties of the Wilson-Schwinger operator that guide our choices for these modifications. The action of the discrete Schwinger operator S_W with Wilson stabilization term on a spinor ψ with its two spin-

components ψ_1 and ψ_2 is given by

$$S_W\psi = \begin{pmatrix} \hat{\partial}_x\hat{\partial}^x + \hat{\partial}_y\hat{\partial}^y & \left(\hat{\partial}_x + i\hat{\partial}_y\right) \\ -\left(\hat{\partial}_x - i\hat{\partial}_y\right) & \hat{\partial}_x\hat{\partial}^x + \hat{\partial}_y\hat{\partial}^y \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix},$$

where $\hat{\partial}_\mu$ denotes the centralized covariant finite difference in spatial direction μ, as introduced in section 3.1. In what follows, we interpret the Wilson-Schwinger operator as a coupled system of two discretized partial differential equations on a $N \times N$ lattice, i.e., $S_W \in \mathbb{C}^{2N^2\times 2N^2}$ and we write $m = N^2$.

The most important property of S_W that we exploit in the method is the Σ_3-hermiticity,

$$(\Sigma_3 S_W)^H = \Sigma_3 S_W, \quad \Sigma_3 = \begin{pmatrix} I & \\ & -I \end{pmatrix} \in \mathbb{C}^{2m\times 2m}.$$

From Lemma 3.12 we have that the eigenvalues λ of S_W are either real or appear in complex conjugate pairs. Given an eigenvalue λ we denote $\left(\lambda, v_\lambda^r, v_\lambda^l\right)$ the triplet with the right eigenvector v_λ^r and the left eigenvector v_λ^l of λ. Recall that in Lemma 3.13, we showed that the two eigentriples associated with eigenvalue λ and its complex conjugate $\bar{\lambda}$ satisfy

$$v_\lambda^l = \Sigma_3 v_\lambda^r \quad \text{and} \quad v_{\bar{\lambda}}^l = \Sigma_3 v_{\bar{\lambda}}^r.$$

Note further that in the case $\lambda \in \mathbb{R}$, this relationship connects the right and left eigenvector of λ. This result implies that once we know the right eigenvectors of S_W, we also know the left eigenvectors of S_W.

Defining the coarse-grid operator S_W^c for non-symmetric operators A in a Petrov-Galerkin fashion by

$$S_W^c = RS_WP,$$

with restriction R and interpolation P, we can use the right eigenvectors of S_W to choose R in terms of P. It is common practice (e.g. [Not09]) in algebraic multigrid methods for non-symmetric problems to base the definition of interpolation P on the right eigenvectors corresponding to the small eigenvalues and accordingly, the definition of restriction R on the left eigenvectors corresponding to the small eigenvalues. Hence, a natural choice in our setting would be to take

$$R = (\Sigma_3 P)^H \quad \text{and} \quad S_W^c = P^H\Sigma_3 S_W P.$$

In this way, we can think of this approach as building interpolation P for the hermitian, but indefinite system $\Sigma_3 S_W$ and use the Galerkin coarse-grid operator for the hermitian indefinite system. Another observation that leads us to further simplifications of the definition of restriction R is as follows.

Lemma 5.7. *Assume that interpolation P is defined such that*

$$\Sigma_3 P = P\Sigma_3^c, \tag{5.10}$$

with Σ_3, Σ_3^c defined as

$$\Sigma_3 = \begin{pmatrix} I & \\ & -I \end{pmatrix} \in \mathbb{C}^{2m\times 2m}, \quad \Sigma_3^c = \begin{pmatrix} I & \\ & -I \end{pmatrix} \in \mathbb{C}^{2m_c\times 2m_c}.$$

Then the coarse-grid correction error propagator (2.8) *with $R = (\Sigma_3 P)^H$ fulfills,*

$$I - P\left(RAP\right)^{-1} RA = I - P\left(P^H AP\right)^{-1} P^H A.$$

Proof. Direct computation yields,

$$\begin{aligned} I - P\left(RAP\right)^{-1} RA &= I - P\left(P^H\Sigma_3 AP\right)^{-1} P^H\Sigma_3 A \\ &= I - P\left(\Sigma_3^c P^H AP\right)^{-1} \Sigma_3^c P^H A \\ &= I - P\left(P^H AP\right)^{-1} P^H A. \end{aligned}$$

□

Thus, assuming that interpolation commutes with Σ_3, we actually can choose $R = P^H$, as we do in the hermitian case, without changing the resulting coarse-grid correction.

Lemma 5.8. *Assume that interpolation P is defined such that* (5.10) *is satisfied and the coarse-grid operator S_W^c is defined by*

$$S_W^c = P^H S_W P.$$

Then the coarse-grid operator is Σ_3^c-hermitian, i.e., it satisfies

$$\left(\Sigma_3^c S_W^c\right)^H = \Sigma_3^c S_W^c.$$

Proof. We have,

$$\begin{aligned} \left(\Sigma_3^c A_c\right)^H &= A_c^H \Sigma_3^c = P^H A^H P\Sigma_3^c = P^H A^H \Sigma_3 P \\ &= P^H \left(\Sigma_3 A\right)^H \Sigma_3 P = P^H\Sigma_3 AP = \Sigma_3^c P^H AP = \Sigma_3^c A_c. \end{aligned}$$

□

Now, a careful look at Σ_3 reveals that the set of interpolation operators that fulfill (5.10) is given by

$$P = \begin{pmatrix} P_1 & \\ & P_2 \end{pmatrix}. \tag{5.11}$$

Thus, such an interpolation operator is spin-decoupling, i.e., it only interpolates spin s_1-variables to s_1-variables and spin s_2-variables to spin s_2-variables.

In what follows, we assume that interpolation P is spin-decoupling, i.e., of the form (5.11). As already mentioned in section 3.1.2, we are interested in solving systems

$$(S_W + mI)\,\psi = \varphi,$$

with a mass-shift m that has a physical interpretation. In order to get a better understanding of the mass-shift and the consequences for the design of an algebraic multigrid method for these linear systems, we recall an important definition.

Definition 5.9. *Given a matrix $A \in \mathbb{C}^{m\times m}$, its* field of values, $\mathbf{F}(A)$, *is defined by*

$$\mathbf{F}(A) = \{z \in \mathbb{C} \mid z = \langle Ax, x\rangle_2, x \in \mathbb{C}^m, \|x\|_2 = 1\}.$$

We define also a coarse-grid version of $\mathbf{F}(A)$.

Definition 5.10. *Let be given a matrix A and a set of interpolation operators $P_{l+1}^l, l = 0, \ldots, L-1$ and multigrid operators defined by*

$$A_l = \left(P_l^{l-1}\right)^H A_{l-1} P_l^{l-1}, A_0 = A \quad \textit{and} \quad T_l = \left(P_l^{l-1}\right)^H T_{l-1} P_l^{l-1}, T_0 = I.$$

We define the field of values $\mathbf{F}_l(A_l, T_l)$ on level l by

$$\mathbf{F}_l(A_l, T_l) = \{z \in \mathbb{C} \mid z = \langle A_l x, x\rangle_2, x \in \mathbb{C}^{m_l}, \langle T_l x, x\rangle_2 = 1\}.$$

In order to make use of the properties of $\mathbf{F}(S_W)$, we present an observation on the field of values. An interesting relation between the field of values $\mathbf{F}(A)$ and its hermitian part $H_A = \frac{1}{2}\left(A + A^H\right)$ is as follows.

Lemma 5.11. *For any $z \in \mathbf{F}(A)$ we have*

$$\lambda_{\min}(H_A) \le \Re(z) \le \lambda_{\max}(H_A).$$

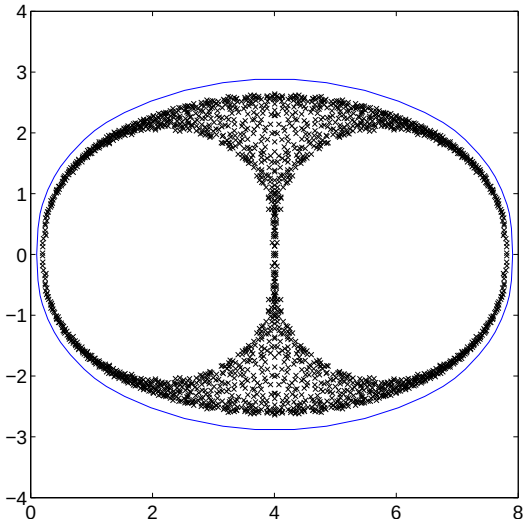


Figure 5.19: Eigenvalues and approximated boundary of the field of values $\mathbf{F}(S_W)$ of a Wilson-Schwinger operator S_W on a 32×32 grid with a gauge configuration $\mathcal{U}$ at temperature $\beta = 5$.

Proof. Due to Definition 5.9, there exists $x \in \mathbb{C}^m$, $\|x\|_2 = 1$ for any $z \in \mathbf{F}(A)$, such that $z = \langle Ax, x\rangle_2$. Then we also have $\bar{z} = \langle A^H x, x\rangle_2$ for the complex conjugate of z. Hence, combining these two equations we finally get

$$\Re(z) = \frac{1}{2}(z + \bar{z}) = \langle H_A x, x\rangle_2.$$

□

With Lemma 5.11 and the observation that

$$H_{(S_W+mI)} = \begin{pmatrix} A(\mathcal{U}) + mI & \\ & A(\mathcal{U}) + mI \end{pmatrix},$$

we obtain a connection between the mass-shift m for S_W and a corresponding mass-shift of the Gauge Laplace operators $A(\mathcal{U})$ on its diagonal. This allows us to distinguish two major classes of mass-shifts: In Figure 5.19 we show an example of the location of eigenvalues and an approximated boundary of the field of values $\mathbf{F}(S_W)$ for a Wilson-Schwinger operator on a 32×32 grid with a gauge configuration at temperature $\beta = 5$. The first class of shifts consists of mass-shifts m that keep the field of values $\mathbf{F}(S_W + mI)$ in the right half-plane of the complex plane and thus the diagonal blocks of $S_W + mI$ positive definite. That is, we have hermitian positive definite Gauge Laplace operators $A(\mathcal{U}) + mI$ on the diagonal of $S_W + mI$. The second class of shifts,

and unfortunately the ones that in general are considered physically relevant, consists of mass-shifts that move the smallest eigenvalues λ of S_W close to the imaginary axis. These shifts still keep the spectrum of $S_W + mI$ in the right half-plane, but the field of values now can have a non-empty intersection with the left half-plane.

In what follows we discuss these two classes of mass shifts, starting with the class that keeps the former.

Mass shift m such that $\mathbf{F}(S_W + mI) \subset \mathbb{C}^+$

If the mass-shift m leaves the field of values in the right half-plane, the Gauge Laplace operator on the block-diagonal of $S_W + mI$ is positive definite. We know that the bootstrap algebraic multigrid method yields an efficient multigrid hierarchy for the Gauge Laplace operator. Our first idea is therefore to use this knowledge to build interpolation for the system $S_W + mI$. We choose interpolation P for S_W fulfilling (5.11), in a way that, given interpolation $\hat{P}$ for the associated Gauge Laplace operator $A(\mathcal{U}) + mI$, we have $P_1 = P_2 = \hat{P}$. The only differences to the method presented in section 5.1.3 are then the use of Kaczmarz relaxation instead of Gauss-Seidel in the solution process and the use of the generalized minimal residual (GMRES) method instead of conjugate gradients (CG) as a Krylov subspace wrapper for the multigrid preconditioner. In Table 5.18 we report the results found with this approach for Wilson-Schwinger operators for the same gauge configurations used in the tests for the Gauge Laplace operator in Table 5.13, but with Kaczmarz relaxation instead of Gauss-Seidel. The multigrid method we get by defining interpolation in terms of the interpolation for the associated Gauge Laplace operator, shows similar behavior for the Wilson-Schwinger problem as the results we presented for the Gauge Laplace operator in Table 5.13. These results indicate that this approach is a viable way to get an efficient multigrid preconditioner for GMRES for the Wilson-Schwinger operators as long as the associated Gauge Laplace operators $A(\mathcal{U})$ are positive definite.

In Figure 5.20, we illustrate the multigrid spectra and field of values according to Definition 5.10 for a Wilson-Schwinger operator for a gauge configuration $\mathcal{U}$ at temperature $\beta = 5$ using the setup of the multigrid hierarchy based on the Gauge Laplace $A(\mathcal{U})$ as explained above. The multigrid hierarchy we get with this approach does match the lower part of the field of values on coarser grids. A feature one would expect in order to get an efficient multigrid method.

Although the results presented in Table 5.18 suggest that it is viable to build interpolation using the hermitian part of the Wilson-Schwinger operator, still an interesting question is if we can also apply the bootstrap algebraic

β \ N	32		64		128	
1	.326$_W$ 13	.325$_W$ 13	.504$_W$ 15	.513$_W$ 14	.614$_W$ 13	.599$_W$ 13
5	.330$_W$ 9	.340$_W$ 9	.486$_W$ 14	.486$_W$ 13	.680$_W$ 17	.695$_W$ 17
10	.167$_W$ 8	.171$_W$ 8	.549$_W$ 14	.549$_W$ 14	.690$_W$ 19	.674$_W$ 18

Table 5.18: Asymptotic convergence and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(2, 2) multigrid cycle with Kaczmarz relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on an $N \times N$ grid at varying temperatures β and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors and $\eta = 4$ smoothings in a bootstrap setup that uses the cycling strategy given in green with $|\mathcal{V}| = 16$. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

multigrid framework directly to the Wilson-Schwinger operator S_W. This is especially interesting if we keep in mind that we cannot straight-forwardly apply the approach based on the hermitian part of S_W to the case where we use "physical" shifts, as for these shifts the hermitian part of S_W is in general no longer positive definite due to Lemma 5.11. Although our analysis for the bootstrap algebraic multigrid framework and least squares interpolation was based on a hermitian positive definite operator A, we consider its application to the non-hermitian Wilson-Schwinger operator. We modify some components accordingly for this non-hermitian setting. While we were successful applying Kaczmarz for the approach that was based on the hermitian part of S_W, we also investigate weighted Richardson iterations, given by the error propagator

$$e^{k+1} = e^k - \omega S_W e^k, \quad \omega = \frac{1}{\|A\|_2}.$$

The choice of ω-Richardson is particularly interesting, as with this choice the eigenvectors of S_W with small eigenvalues coincide with the eigenvectors with small eigenvalues of the smoother. In this sense, ω-Richardson is perhaps a more intuitive choice for a setup that aims at computing eigenvectors to small eigenvalues within the bootstrap process. In what follows, we present results for both, Kaczmarz and ω-Richardson.

We also use the spin-decoupling interpolation P that fulfills (5.11) and

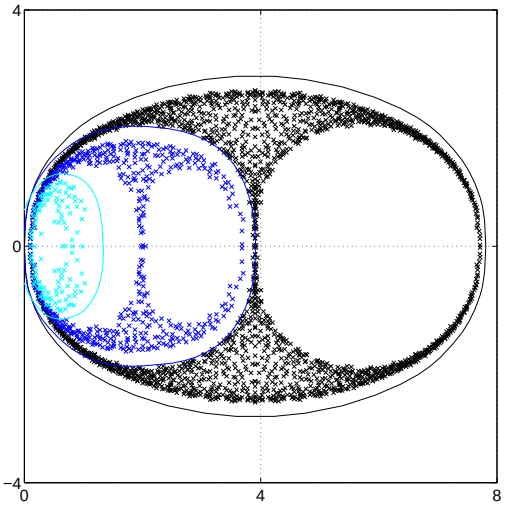


Figure 5.20: Spectra and approximated boundaries of the field of values of the multigrid hierarchy defined by interpolation for the Gauge Laplace operator $A(\mathcal{U})$, for a Wilson-Schwinger operator on a 32×32 grid for a gauge configuration $\mathcal{U}$ at temperature $\beta = 5$.

coarse-grid matrices defined by $A_c = P^H A P$ that preserve the Σ_3-hermiticity. Again, we choose to use full-coarsening for both spin-grids, as illustrated in Figure 5.2.

In the definition of least squares interpolation, we use weights defined by

$$\omega_l = \left(\frac{\|u^{(l)}\|_2}{\|Au^l\|_2} \right)^2$$

as S_W does not define a sesquilinear inner-product, which renders the use of (4.11) not viable.

In order to use the bootstrap algebraic multigrid setup scheme, we have to introduce several modifications to the multilevel eigensolver as given in Algorithm 4. Assume in a two-grid setting that we consider fine-grid eigentriples $(\lambda, v_\lambda^r, v_\lambda^l)$ and computed coarse-grid eigentriples $(\lambda^c, v_{\lambda^c}^r, v_{\lambda^c}^l)$ satisfying

$$S_W^c v_{\lambda^c}^r = \lambda^c T^c v_{\lambda^c}^r \quad \text{and} \quad \left(v_{\lambda^c}^l\right)^H S_W = \lambda^c \left(v_{\lambda^c}^l\right)^H,$$

we interpolate both right and left eigenvectors from the coarse grid to get a first approximation on the fine grid eigenvectors $v_\lambda^r = P v_{\lambda^c}^r$ and $v_\lambda^l = P v_{\lambda^c}^l$, with an associated approximation to the corresponding eigenvalue $\lambda =$

λ^c. We improve this approximations by application of relaxation to the homogeneous systems

$$(S_W - \lambda I)\, v_\lambda^r = 0 \quad \text{and} \quad (S_W - \lambda I)^H\, v_\lambda^l = 0. \tag{5.12}$$

After each iteration we update the approximation to the eigenvalue by

$$\lambda = \frac{\langle A v_\lambda^r, v_\lambda^l \rangle_2}{\langle v_\lambda^r, v_\lambda^l \rangle_2}. \tag{5.13}$$

At a first glance, this approach seems to require double the amount of relaxations than the procedure applied in the hermitian case, but again, we can use the Σ_3-hermiticity in order to remove this obstacle. Assuming that for each eigentriplet $\left(\lambda^c, v_{\lambda^c}^r, v_{\lambda^c}^l\right)$ we also computed the eigentriplet corresponding to $\bar{\lambda}^c$, we know due to Lemma 3.13 that

$$v_{\lambda^c}^l = \Sigma_3 v_{\bar{\lambda}^c} \quad \text{and} \quad v_{\bar{\lambda}^c}^l = \Sigma_3 v_{\lambda^c}.$$

Hence, we can omit the storage and processing of left eigenvectors all together by using the relation between left and right eigenvectors of complex conjugate pairs of eigenvalues λ and $\bar{\lambda}$. Instead of computing both right and left eigenvectors as in (5.12), we simply relax the eigenvalue equations for λ, v_λ^r and $\bar{\lambda}, v_{\bar{\lambda}}^r$ and compute a new approximation to the eigenvalue λ according to (5.13) using the Σ_3 relation between the right and left eigenvectors to a pair of complex conjugate eigenvalues.

In Table 5.19, we first report the asymptotic convergence rates of the resulting multigrid V(4, 4) cycle using ω-Richardson smoothing, and the number of iterations pGMRES needed to reduce the initial residual by a factor of 10^8 when preconditioned with the multigrid V(4, 4)-cycle. We limit this test to the lattice-size 64×64 and a three level method, varying only the setup cycling strategy (V, V^2) and the temperature of the gauge configuration β. In the V^2-setup we compute 16 coarsest grid eigenvectors and apply $\eta = 8$ iterations to the test vectors, both to improve the vectors prior to computing the least squares interpolation and to relax the generalized eigenvalue problem on the intermediate levels in the multigrid hierarchy on the way back to the finest grid. Using the same parameter setting – except for the use of a V(2, 2)- instead of a V(4, 4)-cycle – we report results for Kaczmarz relaxation in Table 5.20. For both relaxations we observe that using only one V-cycle setup already yields quite good results, yet still some fluctuation when changing β. By using a V^2-cycle setup, we see that the results for pGMRES with the multigrid preconditioner become more stable while the stand-alone asymptotic convergence slightly degrades. The fact that the

setup \ β	1		5		10	
V	.377 13	.410 16	.450 17	.626 28	.576 22	.715 36
V^2	.636 13	.768 13	.800 12	.740 12	.822 12	.735 11

Table 5.19: Asymptotic convergence and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(4, 4) multigrid cycle with ω-Richardson relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on a 64×64 lattice at varying temperatures β and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors and $\eta = 8$ iterations in the setup. In the V^2 setup we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

setup \ β	1		5		10	
V	.356 11	.494 14	.364 11	.478 15	.390 12	.551 17
V^2	.422 12	.405 12	.376 11	.387 12	.319 11	.338 11

Table 5.20: Asymptotic convergence and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(2, 2) multigrid cycle with Kaczmarz relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on a 64×64 lattice at varying temperatures β and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors and $\eta = 8$ iterations in the setup. In the V^2 setup we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

preconditioned GMRES method still converges fast suggests that the deterioration of the stand-alone multigrid solver is caused by only a few error components that are not treated by the multigrid method.

We discuss an observation that might cause this behavior later in the case of "physical" shifts, as it emerges with more severe consequences there. Another remark concerning these tests is that the bootstrap V-cycle setup

together with the sub-sequential solution of the linear system for one right-hand-side with the preconditioned pGMRES iteration in our MATLAB implementation takes about the same time than as direct application of unpreconditioned GMRES.

With a first impression of the performance of bootstrap algebraic multigrid directly applied to the Wilson-Schwinger operator S_W for a fixed lattice-size, we present in Tables 5.21 and 5.22 results for varying lattice-sizes at temperature $\beta = 5$ for ω-Richardson and Kaczmarz, respectively. In these tests we compute a three-grid hierarchy using a V^2-cycle setup with $k = 8$ initial test vectors and $|\mathcal{V}| = 16$ approximate eigenvectors computed in the setup using $\eta = 8$ relaxations. We observe that with ω-Richardson the

β \ N	32		64		128	
5	.586 10	.455 10	.800 12	.740 12	$\star$ 16	.951 15

Table 5.21: Asymptotic convergence (divergence marked by a "$\star$") and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(4, 4) multigrid cycle with ω-Richardson relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on a $N \times N$ lattice at fixed temperatures $\beta = 5$ and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors, $\eta = 8$ iterations in the V^2-setup and we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

stand-alone performance of the bootstrap multigrid method degrades with increasing problem-size, but when used as a preconditioner this tendency is much less pronounced and again suggests that only a few error components are not well enough represented in the multigrid hierarchy. A similar observation is true for the approach that uses Kaczmarz relaxation in the multigrid setup and solver, though, in this case the performance of the stand-alone solver is much more robust compared to the use of ω-Richardson when increasing the problem-size, but still not quite scalable. Here we have to keep in mind that due to the randomness contained in the "physical" gauge configurations $\mathcal{U}$, the transition in problem-size does not necessarily imply that the systems for the same β are directly comparable to each other. The performance of the Kaczmarz-based bootstrap approach used as a preconditioner shows the same behavior as the stand-alone method, for the same reasons as just explained.

β \ N	32		64		128	
5	.273 9	.259 9	.376 11	.387 12	.530 13	.459 13

Table 5.22: Asymptotic convergence and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(2, 2) multigrid cycle with Kaczmarz relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on a $N \times N$ lattice at fixed temperatures $\beta = 5$ and a mass shift m, s.t., $\lambda_{\min}(A(\mathcal{U}) + mI) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors, $\eta = 8$ iterations in the V^2-setup and we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

Mass shift m such that $\min(\Re(\sigma(S_W + mI))) = m_0$

We now consider the situation for physical shifts, where m is chosen such that

$$\min(\Re(\sigma(S_W + mI))) = m_0,$$

with m_0 small enough to yield an indefinite diagonal block $A(\mathcal{U}) + mI$ in general. This limits the approach of using interpolation for $A(\mathcal{U}) + mI$ and results in a non-empty intersection of the field of values $\mathbf{F}(S_W + mI)$ with the left half-plane of the complex plane. First, this presents difficulties for Krylov subspace methods (e.g., GMRES) for these problems. For most of the gauge configurations available restarted GMRES, with a restart every 64 iterations, fails to converge in less than 1024, iterations even for the smallest lattice-sizes. On the other hand, as we follow the heuristic that the multigrid hierarchy should accurately represent error components in the part of the field of values close to the origin, we not only have to preserve the intersection with the left half-plane, but also may find eigenvalues on the coarse-grid that lie in the left half-plane.

In our tests we again use both ω-Richardson and Kaczmarz relaxation in the setup and solver and report asymptotic convergence (or divergence, marked by a "$\star$") of the stand-alone solver and the iteration count of the preconditioned pGMRES method to reduce the initial residual by a factor of 10^8. In Table 5.23, we first report the results for a $V(4, 4)$-cycle multigrid method, using ω-Richardson, followed by results in Table 5.24 for Kaczmarz relaxation and a $V(2, 2)$-cycle. As before, we use $k = 8$ initially random test vectors and $\eta = 8$ relaxations throughout the V^2-cycle bootstrap setup computing $|\mathcal{V}| = 16$ eigenvectors for small eigenvalues of the generalized

eigenvalue problem on the coarsest grid. For all lattice-sizes we limit our experiments to a three-level method.

β \ N	32		64		128	
1	⋆ 21	⋆ 22	⋆ 27	⋆ 27	⋆ 33	⋆ 64
5	.743 13	.805 13	⋆ 15	⋆ 17	⋆ 26	⋆ 26
10	.44 10	.299 12	.997 13	.863 13	⋆ 21	⋆ 24

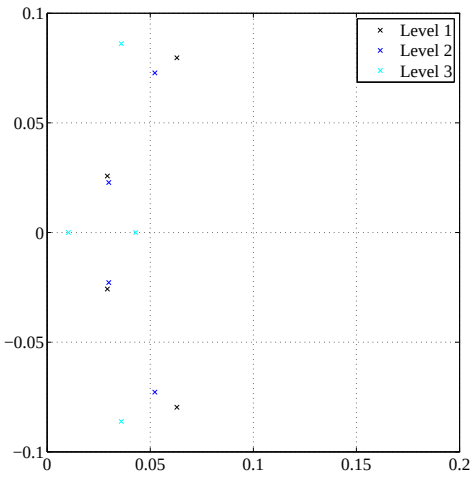
Table 5.23: Asymptotic convergence (divergence marked by a "⋆") of the stand-alone multigrid V(4, 4)-cycle and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(4, 4) multigrid cycle with ω-Richardson relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on an $N \times N$ grid at varying temperatures β and a mass shift m, s.t., $\min(\Re(\sigma(S_W + mI))) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors, $\eta = 8$ iterations in the V^2-setup and we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).

We observe that the multigrid hierarchy computed using ω-Richardson in the bootstrap setup does rarely yield a converging $V(4, 4)$-cycle, but again when used as a preconditioner the number of iterations of GMRES to reduce the initial residual by a factor of 10^8 is small, especially compared to the unpreconditioned restarted GMRES(64) method that did not converge at all. Although the number of iterations grows when increasing the lattice-size, the growth is at least only weakly coupled to N^2 as it roughly doubles when quadrupling the number of unknowns. Similar observations are true for the bootstrap algebraic multigrid method that uses Kaczmarz relaxation, though, again the tendencies are much less pronounced.

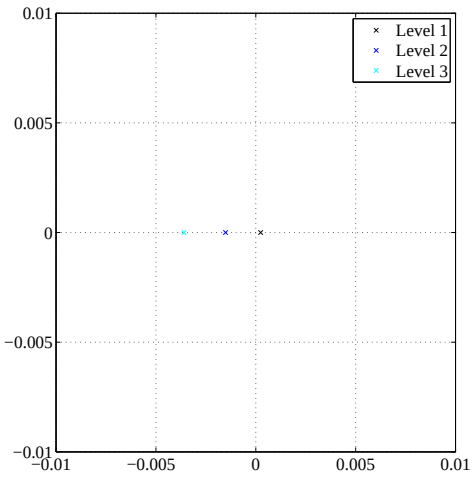
In Figure 5.21, we illustrate the location of the smallest eigenvalues in the multigrid hierarchy in order to study the divergence of the stand-alone multigrid method in case of Kaczmarz relaxation for lattice-sizes 32×32 and $\beta = 1$, using the residual corrected least squares interpolation and for the 64×64 lattice with a gauge configuration at temperature $\beta = 1$. We observe that in Figure 5.21(a), the smallest (in magnitude) eigenvalues of the system is a pair of complex conjugate eigenvalues. While they are nicely represented on the second grid, we see that on the third, and coarsest grid, they are mapped to the real axis. Further tests reveal that the interpolated eigen-

β \ N	32		64		128	
1	.892 14	⋆ 15	⋆ 19	⋆ 20	⋆ 22	⋆ 22
5	.324 11	.259 11	.435 13	.434 14	⋆ 18	⋆ 17
10	.276 10	.252 10	.326 12	.327 12	⋆ 16	⋆ 16

Table 5.24: Asymptotic convergence (divergence marked by a "⋆") of the stand-alone multigrid V(2, 2)-cycle and number of iterations of preconditioned GMRES to reduce the norm of the residual by a factor of 10^8 for the V(2, 2) multigrid cycle with Kaczmarz relaxation for the Wilson-Schwinger operator S_W for "physical" gauge configurations $\mathcal{U}$ on an $N \times N$ grid at varying temperatures β and a mass shift m, s.t., $\min(\Re(\sigma(S_W + mI))) = N^{-2}$. Using LS interpolation with $k = 8$ initially random (complex, $N(0, 1)$ distributed) test vectors, $\eta = 8$ iterations in the V^2-setup and we compute $|\mathcal{V}| = 16$ eigenvector approximations. Black uses the original LS interpolation (4.10); red includes the residual correction (4.19).



(a) Multigrid EWs, top. charge zero



(b) Multigrid EWs, top. charge one

Figure 5.21: Lower end of the spectra of the operators in the three-level hierarchy computed using a V^2-bootstrap setup with Kaczmarz relaxation. In 5.21(a) we present the situation for lattice-size 32×32 at temperature $\beta = 1$ with topological charge zero. In 5.21(b) we illustrate the multigrid eigenvalues for a 64×64 Wilson-Schwinger operator with a gauge configuration of temperature $\beta = 1$, a system with topological charge one.

vectors to the eigenvalues on the real axis indeed are dominated by the pair of complex conjugate eigenvalues. The fact that one of the two eigenvalues on the real axis is much smaller in magnitude than its complex conjugate counterparts on the next finer grid results in a coarse-grid correction that over-corrects the error component in this direction. If the eigenvalue is close enough to the origin, this situation leads to divergence of the stand-alone multigrid method. Further evidence of this situation causing divergence is the analysis of the "divergent" error, i.e., the dominant error component of the iteration, which again turns out to be dominated by a linear combination of the pair of eigenvalues that get mapped to the real axis. This "mis-map" of eigenvalues might be caused by a violation of symmetries of the operator, but unfortunately, to date we have not been able to verify this conjecture.

Figure 5.21(b) represents another class of gauge configurations. These gauge configurations lead to a in magnitude smallest eigenvalue on the real axis and are said to have nonzero topological charge. This results in a maximal shift of the operator, based on the smallest real-part of any eigenvalue of the operator. What we see in the depicted situation is a slight error in the eigenvalue on the coarse-grids, causing the method to diverge as the eigenvalue crosses the origin, flipping its sign in the coarse-grid correction.

The observations we made here may be reasons for divergence of the stand-alone multigrid method and the rapid convergence of the preconditioned Krylov subspace method, as these separated outliers are treated in just one or two additional steps.

Gaining a theoretical understanding of the observed behavior of the multigrid hierarchy is part of future research. In the case where a pair of complex conjugate eigenvalues is mapped to the real axis, we suspect that our way of defining interpolation might violate some hidden symmetry of the problem that we have not found, yet. In the topological charge case, the situation is more intricate and requires intense focus in the future. The derivation of analytical tools and methods for non-hermitian problems, which has been a field of only minor progress in terms of multigrid methods in the last 20 years, is a challenging topic of future work.

Chapter 6

Conclusion and Outlook

The key-point of this thesis is the development of adaptive algebraic multigrid methods. We chose several challenging applications that motivated most of the proposed techniques and posed difficulties that today's algebraic multigrid methods could hardly overcome. Overall we introduced and analyzed two ways to adaptively define interpolation in algebraic multigrid method.

First, we developed a modified version of the adaptive reduction-based approach [MMM06] that allows us to apply the method and corresponding theory to complex-valued systems. From the results we achieved with the adaptive reduction-based approach for the Gauge Laplace operator, which is the simplest operator in Lattice Gauge Theory, it became apparent that an efficient adaptive algebraic multigrid method has to incorporate more than one representation of algebraically smooth error. That is, due to the randomness inherent in the gauge configurations used to define the operators in Lattice Gauge Theory, algebraically smooth error tends to be highly oscillatory and only locally supported. Due to this fact, approaches that heavily rely on the fact that all algebraically smooth error can be represented by only one representation do not yield very efficient algebraic multigrid methods. This observation lead to the development and analysis of a novel framework for adaptive algebraic multigrid, based on the bootstrap algebraic multigrid idea proposed by Brandt and Livne [BL04].

As the key-feature of the bootstrap framework we introduced and analyzed least squares interpolation. In addition we defined and discussed several bootstrap techniques that are mutually beneficial to the least squares definition of interpolation, but can also be used to analyze and adaptively improve algebraic multigrid methods in general. Moreover, we applied methods based on the bootstrap algebraic multigrid framework to the challenging problems arising in Lattice Gauge Theory and demonstrated their efficiency for discretizations of scalar partial differential equations ranging from the

Laplace operator with Dirichlet boundary conditions to its gauged analogue, the Gauge Laplace operator. We further discussed several bootstrap techniques in detail, such as algebraic distance using an anisotropic test problem and the bootstrap multigrid eigensolver in the framework of the computation of steady-state solutions of Markov chain problems. In these experiments the bootstrap multigrid framework not only overcame the difficulties posed by the non-symmetric operators arising in these applications, but also yielded a very efficient stand-alone solver and preconditioner for a Krylov-subspace method. Finally, we presented results for the non-hermitian Wilson-Schwinger operator of Lattice Gauge Theory for Quantum Electrodynamics as a test problem for the systems arising in Lattice Quantum Chromodynamics. Although the proposed method did not always achieve text-book multigrid performance, we made interesting observations for the development of adaptive multigrid methods for non-hermitian operators arising in this context and observed that the proposed methods yield very efficient preconditioners for Krylov-subspace methods.

Our future research aims at gaining additional insight into the bootstrap algebraic multigrid framework proposed in this thesis. In this thesis we were able to formulate some mathematical rigorous results for this approach, but many details in the bootstrap algebraic multigrid framework are still based on heuristics. Part of our future research is to find mathematical justification of these heuristics, including the discussion of appropriate weights in least squares interpolation both in the hermitian and non-hermitian case and the appropriate choice of test vectors and smoothers. An additional question of interest in this context is the debate between the use of eigenvectors or singular vectors to define interpolation for non-hermitian problems and their bootstrap computation. Interesting applications for these investigations include elasticity and bi-harmonics in the hermitian case, whereas we consider the Stokes equation, the Cauchy-Riemann equation and the systems of Lattice Gauge Theory as model problems for the non-hermitian case.

Another topic of our research is the analysis of the notion of algebraic distance in the context of adaptive coarsening and to gain insight in the use of the least squares functional in least squares interpolation to effectively control accuracy and sparsity of least squares interpolation and the bootstrap setup.

We further plan to investigate the use of symmetries in the development of algebraic multigrid methods, e.g., γ_5-hermiticity of the Dirac operator. In this, we also want to consider the problem of eigenvalue "mismatches" in the multigrid hierarchy found in the application of algebraic multigrid for such systems, which we suspect is due to the violation of such symmetries. Possible applications for this investigation include the systems arising in Lattice

Gauge Theory, but also Stokes equation and saddle-point problems in general. We are confident that the proposed research leads to a deeper understanding of adaptive algebraic multigrid in both the hermitian and non-hermitian case and in the end also yields the techniques to define an efficient multigrid solver for the Wilson-Dirac operator of Quantum Chromodynamics.

Bibliography

[Axe94] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, Cambridge, 1994.

[BBC+08] J. Brannick, J. Brower, M. Clark, J. Osborn, and C. Rebbi. Adaptive Multigrid algorithm for lattice QCD. *Phys. Rev. Lett.*, 100(4):041601, January 2008.

[BBK+07] J. Brannick, M. Brezina, D. Keyes, O. Livne, I. Livshits, S. MacLachlan, T. Manteuffel, S. McCormick, J. Ruge, and L. Zikatanov. *Adaptive smoothed aggregation in Lattice QCD*, volume 55, chapter Lecture Notes in Computational Science and Engineering, pages 505–512. Springer, 2007.

[BCF+00] M. Brezina, A. J. Cleary, R. D. Falgout, V. E. Henson, and J. E. Jones. Algebraic multigrid based on element interpolation (AMGe). *SIAM J. Sci. Comput.*, 22(5):1570–1592, 2000.

[BF09] J. Brannick and R. D. Falgout. Compatible relaxation and coarsening in algebraic Multigrid. In preparation, 2009.

[BFKM09] J. Brannick, A. Frommer, K. Kahl, and S. MacLachlan. Adaptive multigrid methods for nearly singular matrices arising from highly disordered physical systems. In preparation, 2009.

[BFM+04] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA). *SIAM J. Sci. Comput.*, 25(6):1896–1920, 2004.

[BFM+05] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA) multigrid. *SIAM Rev.*, 47(2):317–346, 2005.

[BFM+06] M. Brezina, R. Falgout, T. A. Manteuffel, S. MacLachlan, S. McCormick, and J. Ruge. Adaptive algebraic multigrid. *SIAM J. Sci. Comput.*, 27(4):1261–1286, 2006.

[BHM00] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A multigrid tutorial*. SIAM, Second edition, 2000.

[BL04] A. Brandt and O. Livne. Bootstrap Algebraic Multigrid for CFD Applications. Research Report, 2004. Available from authors.

[BL09] A. Brandt and O. E. Livne. Personal communication, 2009.

[Bla88] R Blaheta. A multilevel method with overcorrection by aggregation for solving discrete elliptic problems. *Journal of Computational and Applied Mathematics*, 24(1-2):227, 1988.

[BMR84] A. Brandt, S. F. McCormick, and J. W. Ruge. Algebraic Multigrid (AMG) for sparse matrix equations. In D. J. Evans, editor, *Sparsity and Its Applications*, Cambridge, 1984. Cambridge University Press.

[BP94] A. Berman and P. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics Appl. Math. SIAM, Philadelphia, Ninth edition, 1994.

[BPWX91] J. H. Bramble, J. E. Pasciak, J. Wang, and J. Xu. Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comput.*, 57(195):23–45, 1991.

[BR99] A. Brandt and D. Ron. Satistically optimal renormalization group flow and coarse-to-fine Monte-Carlo acceleration. Technical report, The Weizmann Institute of Science, Rehovot, Israel, May 1999. Gauss Center Report WI/GC-11.

[BR01] A. Brandt and D. Ron. Renormalization Multigrid (RMG): Statistically Optimal Renormalization Group Flow and Coarse-to-Fine Monte Carlo Acceleration. *Journal of Statistical Physics*, 102(1/2):231–257, 2001.

[Bra86] A. Brandt. Algebraic multigrid theory: The symmetric case. *Appl. Math. Comput.*, 19:23–56, 1986.

[Bra95] D. Braess. Towards algebraic multigrid for elliptic problems of second order. *Computing*, 55(4):379–393, 1995.

[Bra00] A. Brandt. General highly accurate algebraic coarsening. *ETNA, Electron. Trans. Numer. Anal.*, 10:1–20, 2000.

[Bra02] A. Brandt. *Multiscale scientific computation: Review 2001*, volume 20 of *Lecture Notes in Computational Science and Engineering*, pages 1–96. Springer, 2002.

[Bra05] J. Brannick. *Adaptive Algebraic Multigrid Coarsening Strategies*. PhD thesis, University of Colorado, Boulder, Boulder, CO, 2005.

[DD06] T. DeGrand and C. DeTar. *Lattice Methods For Quantum Chromodynamics*. World Scientific Publishing Co., 2006.

[DSMM+09] H. De Sterck, T. A. Manteuffel, S. F. McCormick, K. Miller, J. Pearson, J. W. Ruge, and G. Sanders. Smoothed aggregation multigrid for Markov chains. *SIAM J. Sci. Comput.*, accepted, 2009.

[Fal02] R. Falgout. An Adaptive Multigrid Algorithm Based on AMGe. March 2002. Copper Mountain Conference on Iterative Methods.

[FV04] R. D. Falgout and P. S. Vassilevski. On generalizing the algebraic multigrid framework. *SIAM J. Numer. Anal.*, 42(4):1669–1693, 2004.

[FVZ05] R. D. Falgout, Panayot S. V., and L. T. Zikatanov. On two–grid convergence estimates. *Numer. Linear Algebra Appl.*, 12(5–6):471–494, 2005.

[GVL89] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. The John Hopkins University Press, Baltimore, 2nd edition, 1989.

[Hac91] W. Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. Leitfäden der angewandten Mathematik und Mechanik LAMM. Teubner, Stuttgart, 1991.

[HS09] K. Hayami and M. Sugihara. A Geometric View of Krylov Subspace Methods on Singular Systems. Technical report, NII [http://research.nii.ac.jp/TechReports/09-007E.html], 2009.

[HV01] V. E. Henson and P. S. Vassilevski. Element-free AMGe: General algorithms for computing interpolation weights in AMG. *SIAM J. Sci. Comput.*, 23(2):629–650, 2001.

[Kah06] K. Kahl. An algebraic multi-level approach for a model problem for disordered physical systems. Master's thesis, Bergische Universität Wuppertal, November 2006. Diplomarbeit.

[Liv04] O. E. Livne. Coarsening by compatible relaxation. *Numer. Linear Algebra Appl.*, 11(2-3):205–227, 2004.

[Mac04] S. P. MacLachlan. *Improving Robustness in Multiscale Methods*. PhD thesis, University of Colorado, Boulder, Boulder, Colorado, 2004.

[McC85] S. F. McCormick. Multigrid methods for variational problems: General theory for the V- cycle. *SIAM J. Numer. Anal.* , 22:634–643, 1985.

[McC87] S. F. McCormick, editor. *Multigrid Methods*. Frontiers in Applied Mathematics. 1987.

[MMM06] S. P. MacLachlan, T. A. Manteuffel, and S. F. McCormick. Adaptive reduction-based AMG. *Numer. Linear Algebra Appl.*, 13(599–620), 2006.

[MR82] S. F. McCormick and J. W. Ruge. Multigrid methods for variational problems. *SIAM J. Numer. Anal.*, 19:924–929, 1982.

[MS07] S. MacLachlan and Y. Saad. A greedy strategy for coarse-grid selection. *SIAM J. Sci. Comput.*, 29(5):1825–1853, 2007.

[MV92] S. Míka and P. Vaněk. A modification of the two-level algorithm with overcorrection. *Appl. Math.*, 37(1):13–28, 1992.

[Not09] Yvan Notay. Algebraic analysis of two-grid methods: The nonsymmetric case. *Numerical Linear Algebra with Applications*, 2009.

[RS86] J. Ruge and K. Stüben. Algebraic multigrid (AMG). Technical report, Fraunhofer Publications [http://publica.fraunhofer.de/oai.har], 1986.

[RS87] J. W. Ruge and K. Stüben. *Algebraic multigrid*, volume 3, chapter Frontiers Appl. Math., pages 73–130. SIAM, Philadelphia, 1987.

[RTW83] M. Ries, U. Trottenberg, and G. Winter. A note on MGR methods. *Linear Algebra Appl.*, 49:1–26, 1983.

[Sch91] P. J. Schweitzer. *A Survey of Aggregation-Disaggregation in Large Markov Chains*, chapter 4. Probability: Pure and Applied. CRC, 1991.

[SF73] G. Strang and G. Fix. *An analysis of the finite element method.* Prentice-Hall series in Automatic Computation. Prentice-Hall, 1973.

[Stü99] K. Stüben. Algebraic Multigrid (AMG): An Introduction with Applications. GMD Report 53, GMD - Forschungszentrum Informationstechnologie GmbH, Sankt Augustin, Germany, March 1999.

[Stü01] K. Stüben. *An Introduction to Algebraic Multigrid*, pages 413–532. Academic Press, London, UK, 2001.

[TOS01] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid. With guest contributions by A. Brandt, P. Oswald, K. Stüben.* Academic Press, Orlando, FL, 2001.

[Van95] P. Vaněk. Fast multigrid solver. *Appl. Math., Praha* , 40(1):1–20, 1995.

[VBM01] P. Vaněk, M. Brezina, and J. Mandel. Convergence of algebraic multigrid based on smoothed aggregation. *Numer. Math.*, 88(3):559–579, 2001.

[Vir07] E. Virnik. An algebraic multigrid preconditioner for a class of singular M-matrices. *SIAM J. Sci. Comput.*, 29(5):1982–1991, 2007.

[VMB96] P. Vaněk, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56(3):179–196, 1996.

[Wil65] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford University Press, Oxford, UK, 1965.

[Wil74] K. G. Wilson. Confinement of Quarks. *Phys. Rev. D*, 10(8):2445–2459, 1974.

[XZ02] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *Journal of the American Mathematical Society*, 15(03):573, 2002.