

# An Introduction to Algebraic Multigrid

## MAT 228C Project Report

Chuanbin LI  
Wenhao WU

June 10, 2013

## 1 Introduction

The basic idea for the algebraic multigrid (AMG) algorithm is to apply multigrid techniques to problems without a geometric grid, or ones with highly irregular geometric structures [1]. Since it is common to discretize a PDE to a linear system in many numerical methods, AMG directly deal with the discretized linear equations. In fact, AMG is a black-box solver for a linear system  $Au = f$ , where  $A$  is a non-singular matrix. In the following discussion, we require  $A$  to be a symmetric M-matrix, i.e.  $A$  is symmetric, positive definite with diagonal entries positive and off-diagonal entries non-positive. These properties are not necessary for AMG to work, but if the matrix is far from being a symmetric M-matrix, it is less likely that standard AMG works well to it.

In this project, we studied the idea and implementation of the standard AMG, and performed several numerical experiments.

## 2 Grids, Neighbors, and Dependence

In a multigrid algorithm, the following concepts are essential:

- Grids and Neighbors;
- Smoothness;
- Interpolation and Restriction;
- Coarse Grid Selection.

In AMG we also need these components. First we define grids and neighbors.

**Definition:** The grid points are identified with the indices of the unknown quantities.

For example, if the linear system to be solved is  $Au = f$ , where

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

then  $\{1, 2, \dots, n\}$  are the fine-grid points, and a coarse grid means some subset of it.

The connection between each pair of grid points are determined by the matrix  $A = (a_{ij})$ , so we have the following definition for neighbors.

**Definition:** If  $a_{ij} \neq 0$  or  $a_{ji} \neq 0$ , then the  $i$ th and  $j$ th grid points are neighbors. The amplitude of  $a_{ij}$  decides the strength of dependence of the  $i$ th grid point on the  $j$ th point.

**Definition:** Given a threshold value  $0 < \theta \leq 1$ , the variable  $u_i$  *strongly depends* on  $u_j$  if  $-a_{ij} \geq \theta \max_{k \neq i} \{-a_{ik}\}$ . Similarly,  $u_i$  *weakly depends* on  $u_j$  if  $-a_{ij} < \theta \max_{k \neq i} \{-a_{ik}\}$ .

### 3 Algebraic Smoothness

In the geometric case, we define smoothness to be the elimination of high-frequency Fourier modes. However, in AMG, we cannot have a similar definition since it is not easy to apply Fourier analysis to a problem without any geometric information.

On the other hand, a smooth error in the geometric case has an important property that it cannot be reduced effectively by a relaxation scheme. We simply make use of this property to define the algebraic smoothness.

**Definition:** A smooth error is defined to be any error that is not reduced effectively by relaxation.

Notice that this definition depends on the choice of relaxation schemes. Fortunately for a symmetric M-matrix, common relaxation schemes such as weighted Jacobi, Gauss-Seidel, and Gauss-Seidel Red-black schemes all work fine.

#### 3.1 The Relation between Geometric and Algebraic Smoothness

By the definition, it is obvious that a geometrically smooth error must also be an algebraically smooth one. However, the following numerical example indicates that the converse is not necessarily true.

**Example:**

$$-au_{xx} - cu_{yy} + bu_{xy} = 0, \quad (x, y) \in (-1, 1) \times (-1, 1),$$

with homogeneous Dirichlet boundary conditions. We apply the following stencils:

$$D_{xx}^h = \frac{1}{h^2} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}, D_{yy}^h = \frac{1}{h^2} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

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

Coefficients  $a$ ,  $b$ , and  $c$  are locally constant but have different values in different quadrants.

$a = 1$	$a = 1$
$c = 1000$	$c = 1$
$b = 0$	$b = 2$
$a = 1$	$a = 1000$
$c = 1$	$c = 1$
$b = 0$	$b = 0$

The initial value was chosen randomly. After 10 sweeps of G-S relaxation, the error doesn't change essentially, so we reach algebraic smoothness, but not geometric smoothness. To see this point, we present an example similar to one in [2] in Figure 1.

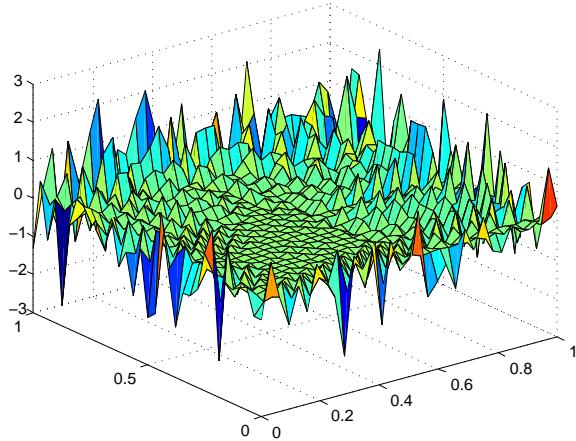


Figure 1: Error after 10 sweeps of G-S relaxation.

### 3.2 Interpreting Algebraic Smoothness

Although algebraic smoothness does not imply geometric smoothness, it has two good properties which play a significant role in the implement of the AMG algorithm.

**Property 1:** If  $A = D - L - L^T$  is a symmetric M-matrix and  $\mathbf{e}$  is a smooth error, then it can be shown that

$$\|\mathbf{r}\|_{D^{-1}} \ll \|\mathbf{e}\|_A, \quad \text{or} \quad (D^{-1} A \mathbf{e}) \ll (\mathbf{e}, A \mathbf{e}).$$

This property means that algebraic smoothness infers small residual. Writing in components yields  $\sum_i \frac{r_i^2}{a_{ii}} \ll a_{ii} |\mathbf{e}_i|$ . At least on average we have  $|r_i| \ll a_{ii} |\mathbf{e}_i|$ .

We write it loosely as  $A\mathbf{e} \approx 0$ , or

$$a_{ii} \mathbf{e}_i \approx - \sum_{j \neq i} a_{ij} \mathbf{e}_j. \quad (1)$$

This equation indicates that  $\mathbf{e}_i$  can be approximately expressed as a linear combination of its neighbors, which lays a foundation for the interpolation operator, which will be discussed in the next section.

**Property 2:** If  $\mathbf{e}$  is a smooth error, then for most  $i$ ,

$$\sum_{j \neq i} \left( \frac{|a_{ij}|}{a_{ii}} \right) \left( \frac{\mathbf{e}_i - \mathbf{e}_j}{\mathbf{e}_i} \right)^2 \ll 1, \quad 1 \leq i \leq n. \quad (2)$$

Observing (2), we find that the left side is a sum of products of non-negative terms. The inequality implies that at least one of the factor in each product must be small. If  $\mathbf{e}_i$  depends strongly on  $\mathbf{e}_j$ ,  $-a_{ij}$  can be compared to  $a_{ii}$ . Therefore, for those terms,  $\mathbf{e}_i - \mathbf{e}_j$  must be small, or in other words  $\mathbf{e}_i \approx \mathbf{e}_j$ . We describe this property as a smooth error varies slowly among strong dependent grids. It justifies the idea that we can interpolate a fine grid point from its strong dependent neighbors on the coarse grid.

## 4 Interpolation

Now assuming that we have already had a coarse grid, we want to define an interpolation operator from the coarse grid to the fine grid in a way that smooth error can be transferred accurately. Our goal is to find a operator  $I_{2h}^h$  s.t.

$$(I_{2h}^h \mathbf{e})_i = \begin{cases} \mathbf{e}_i, & \text{if } i \in C; \\ \sum_{j \in C_i} \omega_{ij} \mathbf{e}_j, & \text{if } i \in F, \end{cases}$$

where  $C$  stands for the coarse grid, and  $F$  for the set of grid points that fall into the fine grid only.

First we introduce some denotations. We denote the neighbors of  $i$  to be  $N_i$ , and

$$N_i = C_i \cup D_i^s \cup D_i^w,$$

where

- $C_i$  stands for the neighboring coarse grid points that strongly influence  $i$ , which is the coarse interpolatory set for  $i$ ;
- $D_i^s$  stands for neighboring fine grid points that strongly influence  $i$ ;
- $D_i^w$  stands for neighboring grid points that weakly influence  $i$ ; it may contain both coarse and fine grid points.

Recall (1). We have

$$a_{ii}e_i \approx - \sum_{j \neq i} a_{ij}e_j = \sum_{j \in N_i} a_{ij}e_j = - \sum_{j \in C_i} a_{ij}e_j - \sum_{j \in D_i^s} a_{ij}e_j - \sum_{j \in D_i^w} a_{ij}e_j.$$

Our task is to represent  $e_j$  in the last two sums of the right hand side with  $e_i$  and/or  $e_j$  in the coarse grid.

First we deal with  $D_i^w$  simply by attributing it to diagonal terms, i.e.

$$(a_{ii} + \sum_{j \in D_i^w} a_{ij})e_i \approx - \sum_{j \in C_i} a_{ij}e_j - \sum_{j \in D_i^s} a_{ij}e_j.$$

We justify this treatment from two aspects:

- If the dependence is underestimated, then  $e_i$  strongly depends on  $e_j \in D_i^w$ . Since smooth error varies slowly among them by Property 2 of algebraic smoothness, it is justified to replace  $e_j$  with  $e_i$ ;
- If the dependence is weak, then  $a_{ij} \ll a_{ii}$ , making the error insignificant.

Now we want to treat  $D_i^s$ . Experience shows that it is better to attribute them to points in  $C_i$ .

$$e_j \approx \frac{\sum_{k \in C_i} a_{jk}e_k}{\sum_{k \in C_i} a_{jk}}, \quad j \in D_i^s.$$

This approximation computes  $e_j \in D_i^s$  as a weighted average of its neighbors. The weight is set to be  $a_{ij}$ , since  $a_{ij}$  determines the dependence of  $e_j$  on its neighbors. Notice that this approximation requires that if  $e_i$  and  $e_j$  are strongly connected, then  $C_i \cap C_j \neq \emptyset$ .

After Combining the above and calculation, we get

$$(I_{2h}^h \mathbf{e})_i = \begin{cases} e_i, & \text{if } i \in C; \\ \sum_{j \in C_i} \omega_{ij}e_j, & \text{if } i \in F; \end{cases}$$

where  $\omega_{ij} = -\frac{a_{ij} + \sum_{m \in D_i^s} \frac{a_{im}a_{mj}}{\sum_{k \in C_i} a_{mk}}}{a_{ii} + \sum_{n \in D_i^w} a_{in}}$ .

## 5 Selecting the Coarse Grid and Coarse-Grid Operators

### 5.1 Preliminaries

Without geometric structure, the selection of coarse grid point must be based solely on the matrix  $\mathbf{A}$ . As in the geometric multigrid method, the coarse grid must be [3]

- on which smooth error can be approximated accurately,
- from which smooth functions can be interpolated accurately,
- that has substantially fewer points than the fine grid, so that the residual problem may be solved with relatively little expense.

The selection of coarse grid points, i.e. the partition of  $C$ -points and  $F$ -points, relies on the following 2 heuristic criteria:

- **H-1:** For each  $F$ -point  $i$ , every point  $j \in S_i$  that strongly influences  $i$  either should be in the coarse interpolatory set  $C_i$  or should strongly depend on at least one point in  $C_i$ .
- **H-2:** The set of coarse points  $C$  should be a maximal subset of all points with the property that no  $C$ -point strongly depends on another  $C$ -point.

The essential idea of **H-1** is that each  $F$ -point can be interpolated by either  $C$ -points directly or via other  $F$ -points which can be interpolated by  $C$ -points, and there is no 2  $F$ -points that one strongly influences the other but they do not strongly depend on a common  $C$ -point. This criterion ensures that the interpolation introduced previously can be carried out. **H-2** is designed to limit the total number of  $C$ -point to guaranteed the efficiency of the computation on the coarse grid. In the  $C$ -points selection algorithm we are going to introduce in the next subsection, **H-1** is enforced rigorously while H-2 is used as a guide.

### 5.2 The Coloring Scheme

The  $C$ -points selection algorithm named *coloring scheme* consists of two passes. In the first pass, a set of  $C$ -points that have good approximation property and tend to satisfy H-2 are selected; then a second pass is carried out to ensure that **H-1** is satisfied.

#### 5.2.1 The First Pass

In the first pass, a measurement  $\lambda_i$  is defined for each point  $i$  to indicate its potential to be a  $C$  point. This measurement is initialized as the  $|S_i^T|$ , since the more points that point- $i$  can influence, the more economic it is to define  $i$  as a  $C$  point.

Each time we look among all unassigned points for the point with the maximal  $\lambda_i$ , and assign it to be a  $C$ -point. For all the unassigned points  $j \in S_i^T$ , since they now can be interpolated with the  $C$ -point  $i$ , they are assigned to be  $F$ -points. For all unassigned points  $k \in S_j$ , since if they are assigned to be  $C$ -points, we will have more  $C$ -points to interpolate the  $F$ -points  $j$  so that the interpolation will be more accurate, so we tend to assign points  $k$  to be  $C$ -points. Therefore, we increase the measurement  $\lambda_k$  by 1. The above procedures are repeated until all points has been assigned.

The first pass is summarized as follows

- Define  $\lambda_i$ : a measure of potential quality of point  $i$  as a  $C$ -point.
- Initialize  $\lambda_i = |S_i^T|$
- While there are points unassigned to  $C$  or  $F$ 
  - Assign  $\hat{i} = \arg \max_{i \text{ unassigned}} \{\lambda_i\}$  to  $C$ .
  - Assign all unassigned points  $j \in S_{\hat{i}}^T$  to  $F$ .
  - For each  $j \in S_{\hat{i}}^T$  in the above step,  $\lambda_k = \lambda_k + 1$  for all unassigned points  $k \in S_j$ .

### 5.2.2 The Second Pass

After the first pass, we will go through a second pass to ensure that H-1 is satisfied: there are no two  $F$ -points that one strongly depends on another but they dont depend on a common  $C$ -point. We simply check each pair of  $F$ -points to see whether one strongly depends on the other. If so, one of them is assigned to be  $C$ -point. This procedure is repeated until all pairs of  $F$ -points have been checked. The second pass can be summarized as follows

- (Following Pass 1) While there are  $F$ - $F$  pairs unchecked
  - Check the next pair of  $F$ -points
  - If 1 of the 2  $F$ -points strongly depends on the other, but they don't strongly depend on a common  $C$ -point, change one of them to be  $C$ -point.

### 5.2.3 Demonstration of Coloring Scheme

We present a demonstration based on the Numerical example in [] for the coloring scheme algorithm (the same example will be used to simulate the AMG algorithm in the next section). We have a Poisson equation

$$-\nabla \cdot (a(x, y)\nabla u) = 1,$$

on a plane with homogeneous Dirichlet boundary condition and a hole in the middle. The coefficient  $a$  is different in these three areas. This equation is

discretized with an unstructured grid. Now we begin the pass 1 of the coloring scheme. The assignment after the first and second round of the first pass are shown in Figure 2 and Figure 3. The assignment after the first and second pass of the coloring scheme are plotted in Figure 4 and Figure 5. Finally, we demonstrate the hierarchy of the coarse grid in Figure 6.

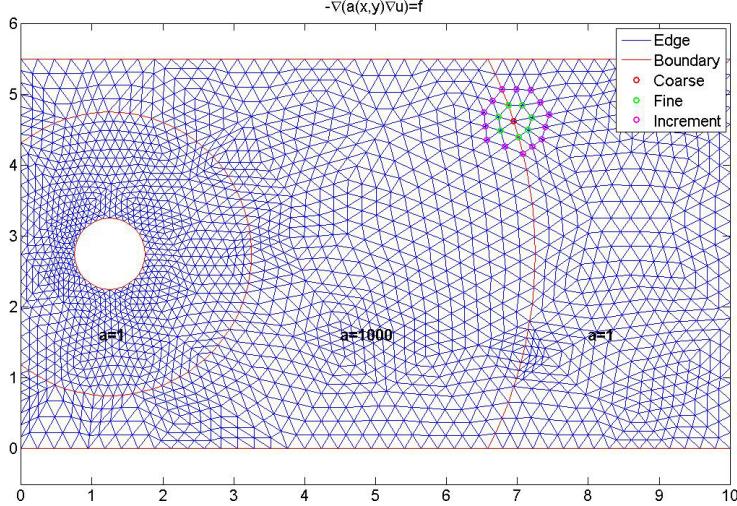


Figure 2: The 1st round of the first pass of the coloring scheme.

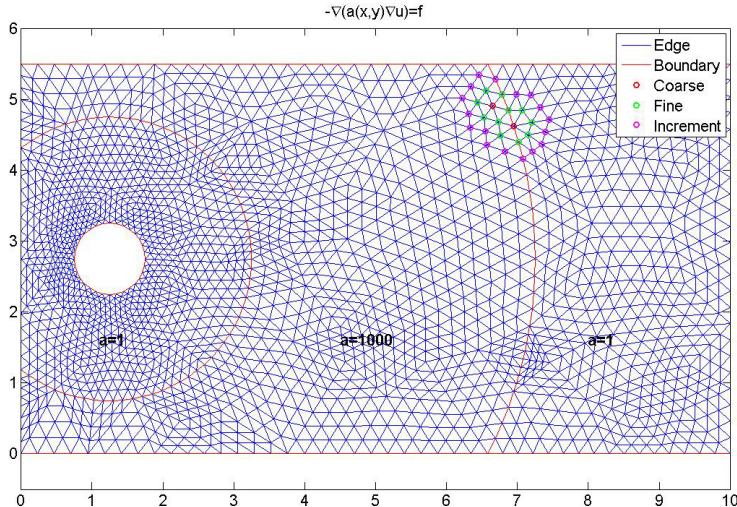


Figure 3: The 2nd round of the first pass of the coloring scheme.

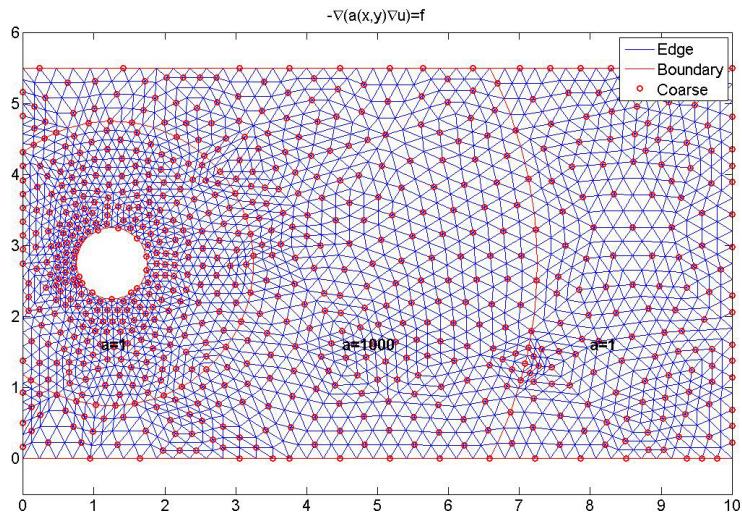


Figure 4: The first pass of the coloring scheme.

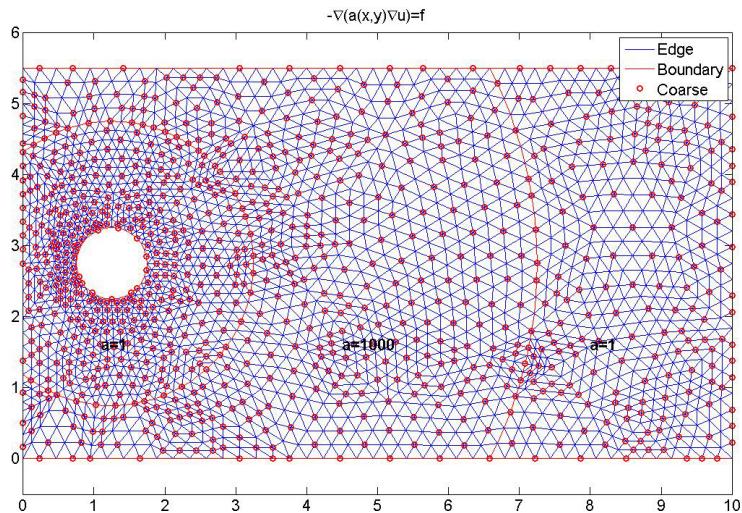


Figure 5: The 2nd pass of the coloring scheme.

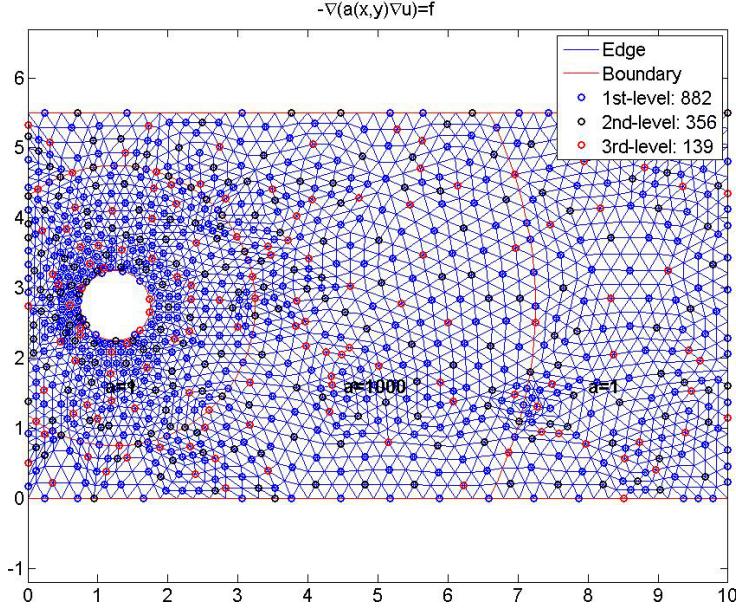


Figure 6: The hierarchy of the coarse grid.

### 5.3 Coarse-Grid Operators

Based on previous coarse grid selection method and the interpolation method in the previous section, the interpolation operator  $I_{2h}^h$  can be constructed. Then the restriction operator, by using the usual variational property, can be defined as

$$I_h^{2h} = (I_{2h}^h).$$

The coarse grid operator can be constructed using the Galerkin condition

$$A^{2h} = I_h^{2h} A^h I_{2h}^h.$$

## 6 V-Cycle AMG

### 6.1 Algorithm Description

So far we have defined all the elements for AMG just as in the case of geometric MG, including a relaxation scheme, interpolation and restriction operator  $I_{2h}^h$  and  $I_h^{2h}$ , coarse grid operator  $A^{2h}$ . We can define the V-Cycle AMG algorithm as follows:

$$\mathbf{v}^h \leftarrow \text{AMG}(A^h, \mathbf{v}^h, \mathbf{f}^h).$$

- If  $\Omega^h$  =coarsest grid, then solve  $A^h \mathbf{v}^h = \mathbf{f}^h$  directly; Else

- Relax  $\nu_1$  times on  $A^h \mathbf{v}^h = \mathbf{f}^h$  with a given initial guess  $\mathbf{v}^h$

$$\mathbf{r}^{2h} \leftarrow I_h^{2h}(\mathbf{f}^h - A^h \mathbf{v}^h),$$

$$A^{2h} = I_h^{2h} A^h I_{2h}^h,$$

$$\mathbf{e}^{2h} \leftarrow \text{AMG}(A^{2h}, \mathbf{0}, \mathbf{r}^{2h})$$

- Interpolate  $\mathbf{e}^h = I_{2h}^h \mathbf{e}^{2h}$
- Correct  $\mathbf{v}^h \leftarrow \mathbf{v}^h + I_{2h}^h \mathbf{e}^{2h}$
- Relax  $\nu_2$  times on  $A^h \mathbf{v}^h = \mathbf{f}^h$  with a given initial guess  $\mathbf{v}^h$

## 6.2 Simulation

Now we present our simulation on V-Cycle AMG based on the same examples as in section 5.2.3. The threshold value to determine strong dependence is  $\theta = 0.25$ . In the V-Cycle of depth 4, we use a weighted Jacobi relaxation scheme with the weighting coefficient  $\omega = 2/3$  and relax  $\nu_1 = 2$  times before going down to the coarser grid and  $\nu_2 = 2$  times before going up to the finer grid. Firstly, the solution of this V-Cycle AMG method is compared with the solution given by MATLAB PDE tool box, which is assumed to be the actual solution, as in Figure 7.

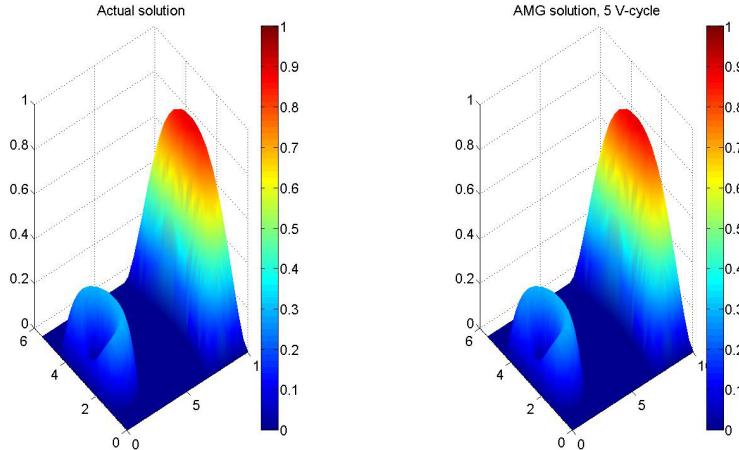


Figure 7: Comparison between the V-Cycle AMG-method after 5 V-Cycle with the solution given by MATLAB pdetool toolbox.

Next we compare the convergence rate of the residual of the V-Cycle AMG method with that of pure weighted Jacobi relaxation method. In Figure 8, we can see that the convergence rate of weighted Jacobi scheme is very slow after merely a short period of time, indicating  $a$ -smoothness. On the contrary, in

Figure 9 the convergence of residual for V-Cycle AMG method is very fast, confirming the effective ness of AMG method.

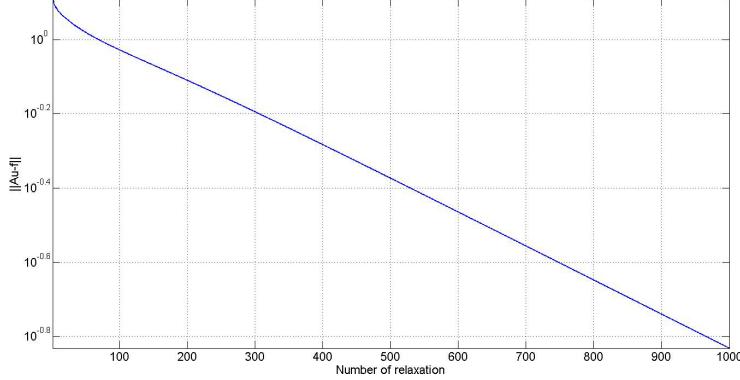


Figure 8: Convergence rate of the residual for pure weighted Jacobi scheme.

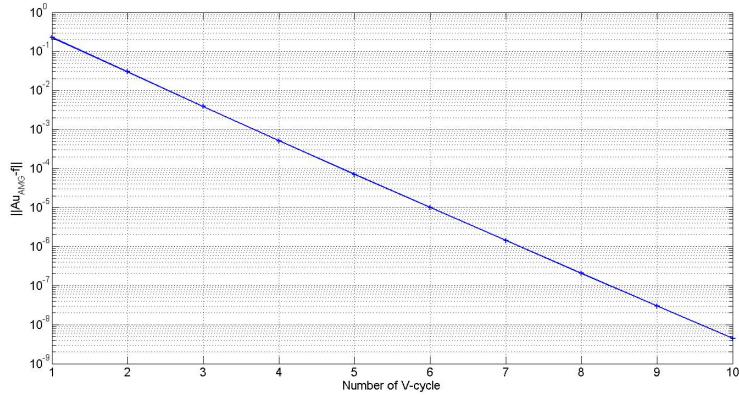


Figure 9: Convergence rate of the residual for V-Cycle AMG method.

## 7 Conclusion

In this project, we present a brief introduction to Algebraic Multigrid method. We first introduced the concept of grids, neighbors, strong dependence and algebraic smoothness which bridged AMG with geometric MG method. Then we definedd the fundamental elements of AMG method including interpolation and restriction operator, coarse grid operator, which are the counterpart to those of MG method. based on these definition, we implemented the V-Cycle AMG method and presented a few numerical examples.

## References

- [1] Ulrich Trottenberg, Cornelius W Oosterlee, and Anton Schuller. *Multigrid*. Academic press, 2000.
- [2] Klaus Stüben. *Algebraic multigrid (AMG): an introduction with applications*. GMD-Forschungszentrum Informationstechnik, 1999.
- [3] William L Briggs, Steve F McCormick, et al. *A multigrid tutorial*, volume 72. Siam, 2000.