Chapter 7

Algebraic Multigrid Methods

Remark 7.1 *Motivation.* The (geometric) multigrid methods described so far need a hierarchy of (geometric) grids, from the coarsest one ($l = 0$) to the finest one. On all levels but the coarsest one, the smoother will be applied and on the coarsest level, the system is usually solved exactly. However, the following question arises:

- What should be done if the available coarsest grid possesses already that many degrees of freedom that the use of a direct solver takes too much time ?

This situation will happen frequently if the problem is given in a complicated domain in $\mathbb{R}^d$, $d \in \{2, 3\}$, see Figure 7.1 for an (academic) example. Complicated domains are very likely to be given in applications. Then, the application of a grid generator will often lead to (coarsest) grids that are so fine that a refinement would lead to so many degrees of freedom that an efficient simulation of the problem is not possible. Altogether, there is just one grid.

To handle the situation of a coarsest grid with many degrees of freedom, there are at least two possibilities.

- *One level iterative scheme.* In the case that there is a geometric grid hierarchy but the coarse grid is already fine, one can use a simple iterative method, e.g., the smoother, to solve the system on the coarsest grid approximately. Then, the smooth error modes on this grid are not damped. However, experience shows that this approach works in practice sometimes quite well.
  If there is just one grid available, a Krylov subspace method can be used for solving the arising linear systems of equations.
- *Iterative scheme with multilevel ideas.* Construct a more complicated iterative method which uses a kind of multigrid idea for the solution of the system on the coarsest geometric grid. The realization of this multigrid idea should be based only on information which can be obtained from the matrix on the coarsest grid. This type of solver is called Algebraic Multigrid Method (AMG).

□

7.1 Components of an AMG and Definitions

Remark 7.2 *Components.* An AMG possesses the same components as a geometric multigrid method:

- a hierarchy of levels,
- a smoother,
- a prolongation,
- a restriction,
- coarse grid operators.

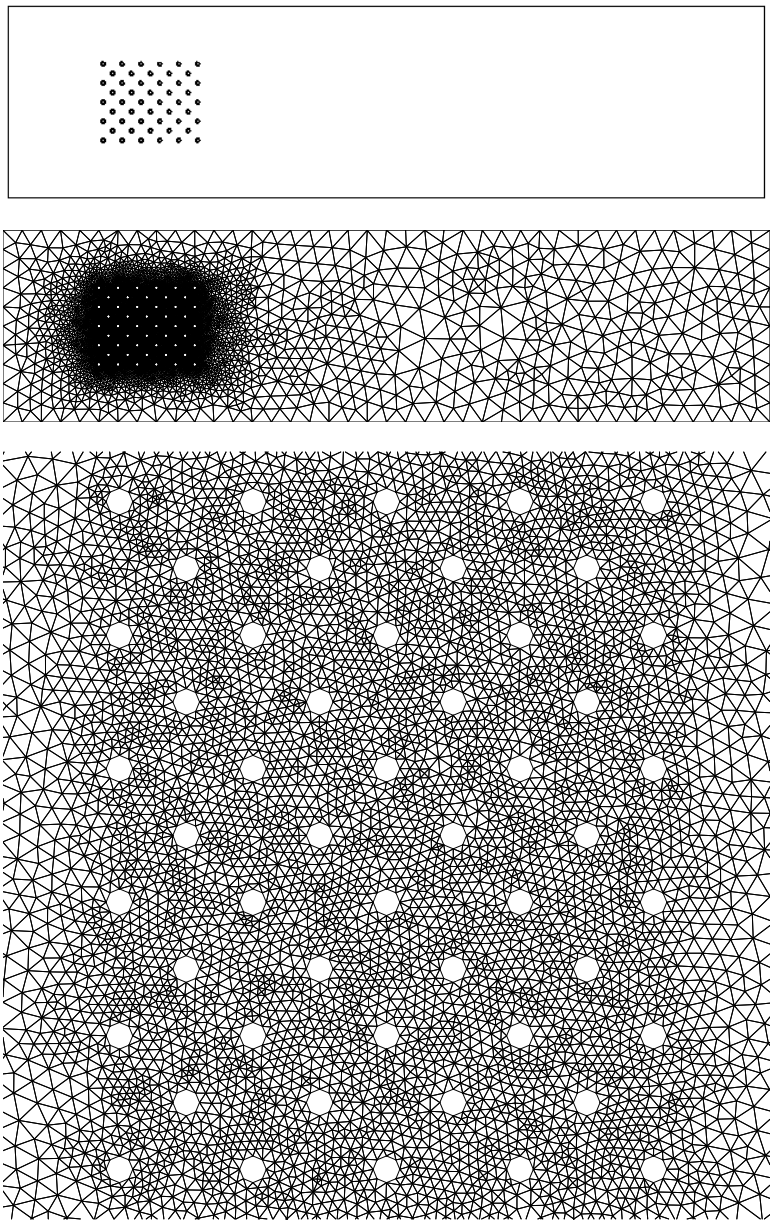

Figure 7.1: Top: domain with many holes (like the stars in the flag of the United Stars); middle: triangular grid from a grid generator; bottom: zoom into the region with the holes.

A level or a grid is a set of unknowns or degrees of freedom. In contrast to geometric multigrid methods, the hierarchy of levels is obtained by starting from a finest level and reducing the number of unknowns to get the coarser levels.

AMGs restrict themselves on using only simple smoothers, e.g., the damped Jacobi method. This approach is in contrast to geometric multigrid methods, whose efficiency can be enhanced by using appropriate smoothers.

In this course, only the case of symmetric positive definite matrices will be considered. Then, the restriction is always defined as the transpose of the prolongation, i.e.,

$$I_f^c = \left(I_c^f\right)^T,$$

where "f" refers to the fine grid and "c" to the next coarser grid.

The coarse grid operator is defined by the Galerkin projection

$$A^c = I_f^c A^f I_c^f .$$

□

Remark 7.3 *Main tasks in the construction of AMGs.* There remain two main tasks in the construction of an AMG:

- An appropriate hierarchy of levels has to be constructed fully automatically, using only information from the matrix on the current grid to construct the next coarser grid.
- One has to define an appropriate prolongation operator.

These two components will determine the efficiency of the AMG.

In contrast to geometric multigrid methods, an AMG constructs from a given grid a coarser grid. Since the final number of coarser grids is not known a priori, it makes sense to denote the starting grid by level 0, the next coarser grid by level 1 and so on.

The coarsening process of an AMG should work automatically, based only on information from the matrix on the current level. To describe this process, some notation is needed. AMGs are set up in an algebraic environment. However, it is often convenient to use a grid terminology by introducing fictitious grids with grid points being the nodes of a graph which is associated with the given matrix $A = (a_{ij})$. □

Definition 7.4 Graph of a matrix, set of neighbor vertices, coupled vertices. Let A be a sparse $n \times n$ matrix with a symmetric sparsity pattern, i.e., a_{ij} is allowed to be non-zero if and only if a_{ji} is allowed to be non-zero. Let $\Omega = G_A(V, E)$ be the graph of the matrix consisting of a set

$$V = \{v_1, \ldots, v_n\}$$

of n ordered vertices (nodes, unknowns, degrees of freedom) and a set of edges E such that the edge e_{ij}, which connects v_i and v_j for $i \neq j$, belongs to E if and only if a_{ij} is allowed to be non-zero.

For a vertex v_i, the set of its neighbor vertices N_i is defined by

$$N_i = \{v_j \in V \ : \ e_{ij} \in E\} .$$

The number of elements in N_i is denoted by $|N_i|$.

If $e_{ij} \in E$, then the vertices v_i and v_j are called coupled or connected. □

Example 7.5 *Graph of a matrix.* Consider the matrix

$$A = \begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix} . \tag{7.1}$$

Let the vertex v_i correspond to the i-th unknown, i.e., to the degree of freedom that corresponds to the i-th row of the matrix. Then the graph of A has the form as given in Figure 7.2. It is

$$E = \{e_{12}, e_{21}, e_{13}, e_{31}, e_{24}, e_{42}, e_{34}, e_{43}\} .$$

□

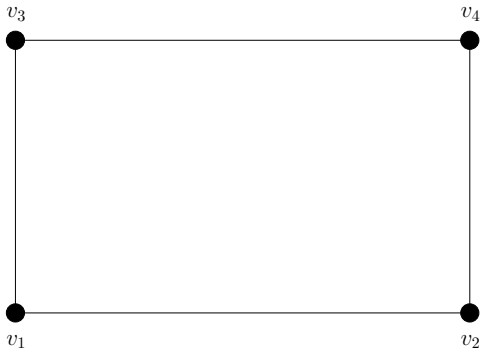

Figure 7.2: Graph $\Omega = G_A(V, E)$ of the matrix (7.1).

7.2 Algebraic Smoothness

Remark 7.6 *Notations.* In geometric multigrid methods, an error is called smooth if it can be approximated well on some pre-defined coarser level. In AMGs there are no pre-defined grids. Let S be the smoother on Ω, then an error is said to be algebraically smooth if the convergence of the fixed point iteration with the matrix S is slow, i.e., $S\mathbf{e} \approx \mathbf{e}$.

To define the property of algebraic smoothness precisely, some inner products and norms of vectors have to be introduced. Let D be the diagonal matrix corresponding to $A \in \mathbb{R}^{n\times n}$ and let $(\cdot,\cdot)$ be the Euclidean inner product of two vectors

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} u_i v_i.$$

Then, the following inner products and norms are defined

$$\begin{aligned}
(\mathbf{u},\mathbf{v})_0 &= (D\mathbf{u},\mathbf{v}), & \|\mathbf{u}\|_0 &= (\mathbf{u},\mathbf{u})_0^{1/2},\\
(\mathbf{u},\mathbf{v})_1 &= (A\mathbf{u},\mathbf{v}), & \|\mathbf{u}\|_1 &= (\mathbf{u},\mathbf{u})_1^{1/2},\\
(\mathbf{u},\mathbf{v})_2 &= \left(D^{-1}A\mathbf{u}, A\mathbf{v}\right), & \|\mathbf{u}\|_2 &= (\mathbf{u},\mathbf{u})_2^{1/2}.
\end{aligned}$$

The norm $\|\mathbf{u}\|_1$ is sometimes called energy norm.

In this course, only classes of matrices will be considered where $\rho\left(D^{-1}A\right)$ is uniformly bounded, i.e., the spectral radius is bounded independently of the grid. This property holds for many classes of matrices which occur in applications. □

Lemma 7.7 Properties of the norms. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. Then the following inequalities hold for all $\mathbf{v} \in \mathbb{R}^n$:*

$$\|\mathbf{v}\|_1^2 \leq \|\mathbf{v}\|_0 \|\mathbf{v}\|_2, \tag{7.2}$$

$$\|\mathbf{v}\|_2^2 \leq \rho\left(D^{-1}A\right) \|\mathbf{v}\|_1^2, \tag{7.3}$$

$$\|\mathbf{v}\|_1^2 \leq \rho\left(D^{-1}A\right) \|\mathbf{v}\|_0^2. \tag{7.4}$$

Proof: (7.2). This estimate follows from the Cauchy–Schwarz inequality and the

symmetry of D

$$\begin{aligned}
\|\mathbf{v}\|_1^2 &= (A\mathbf{v},\mathbf{v}) = \mathbf{v}^T A\mathbf{v} = \mathbf{v}^T A D^{-1/2} D^{1/2}\mathbf{v} = \left(D^{-1/2}A\mathbf{v}, D^{1/2}\mathbf{v}\right) \\
&\leq \left\|D^{-1/2}A\mathbf{v}\right\| \left\|D^{1/2}\mathbf{v}\right\| \\
&= \left(D^{-1/2}A\mathbf{v}, D^{-1/2}A\mathbf{v}\right)^{1/2} \left(D^{1/2}\mathbf{v}, D^{1/2}\mathbf{v}\right)^{1/2} \\
&= \left(A\mathbf{v}, D^{-1/2}D^{-1/2}A\mathbf{v}\right)^{1/2} \left(\mathbf{v}, D^{1/2}D^{1/2}\mathbf{v}\right)^{1/2} \\
&= \left(A\mathbf{v}, D^{-1}A\mathbf{v}\right)^{1/2} (\mathbf{v}, D\mathbf{v})^{1/2} \\
&= \|\mathbf{v}\|_2 \|\mathbf{v}\|_0 ,
\end{aligned}$$

where $\|\cdot\|$ is here the Euclidean vector norm.

(7.3). The matrix $D^{-1}A$ is in general not a symmetric matrix. However, it has the same eigenvalues as the symmetric matrix $A^{1/2}D^{-1}A^{1/2}$, since from

$$D^{-1}A\mathbf{x} = \lambda\mathbf{x}$$

one obtains with $\mathbf{x} = A^{-1/2}\mathbf{y}$

$$D^{-1}AA^{-1/2}\mathbf{y} = \lambda A^{-1/2}\mathbf{y} \quad \Longleftrightarrow \quad A^{1/2}D^{-1}A^{1/2}\mathbf{y} = \lambda\mathbf{y}.$$

In particular, the spectral radii of both matrices are the same. Using the definition of the positive definiteness, one sees that $A^{1/2}D^{-1}A^{1/2}$ is positive definite since the diagonal of a positive definite matrix is a positive definite matrix. Hence, one gets, using a well known property of the spectral radius for symmetric positive definite matrices (Rayleigh quotient)

$$\rho\left(D^{-1}A\right) = \rho\left(A^{1/2}D^{-1}A^{1/2}\right) = \lambda_{\max}\left(A^{1/2}D^{-1}A^{1/2}\right) = \sup_{\mathbf{x}\in\mathbb{R}^n} \frac{\left(A^{1/2}D^{-1}A^{1/2}\mathbf{x},\mathbf{x}\right)}{(\mathbf{x},\mathbf{x})}.$$

Setting now $\mathbf{x} = A^{1/2}\mathbf{v}$ gives an estimate of the spectral radius from below

$$\rho\left(D^{-1}A\right) \geq \frac{\left(A^{1/2}D^{-1}A^{1/2}A^{1/2}\mathbf{v}, A^{1/2}\mathbf{v}\right)}{(A^{1/2}\mathbf{v}, A^{1/2}\mathbf{v})} = \frac{(D^{-1}A\mathbf{v}, A\mathbf{v})}{(A\mathbf{v},\mathbf{v})} = \frac{\|\mathbf{v}\|_2^2}{\|\mathbf{v}\|_1^2},$$

where the symmetry of A was also used.

(7.4). The matrix $D^{-1}A$ has also the same eigenvalues as the matrix $D^{-1/2}AD^{-1/2}$, since from

$$D^{-1}A\mathbf{x} = \lambda\mathbf{x}$$

it follows with $\mathbf{x} = D^{-1/2}\mathbf{y}$ that

$$D^{-1}AD^{-1/2}\mathbf{y} = \lambda D^{-1/2}\mathbf{y} \quad \Longleftrightarrow \quad D^{-1/2}AD^{-1/2}\mathbf{y} = \lambda\mathbf{y}.$$

Hence, $\rho\left(D^{-1}A\right) = \rho\left(D^{-1/2}AD^{-1/2}\right)$. The matrix $D^{-1/2}AD^{-1/2}$ is symmetric and positive definite, which follows by the definition of the positive definiteness and the assumed positive definiteness of A. Using the Rayleigh quotient yields

$$\rho\left(D^{-1}A\right) = \rho\left(D^{-1/2}AD^{-1/2}\right) = \lambda_{\max}\left(D^{-1/2}AD^{-1/2}\right) = \sup_{\mathbf{x}\in\mathbb{R}^n} \frac{\left(D^{-1/2}AD^{-1/2}\mathbf{x},\mathbf{x}\right)}{(\mathbf{x},\mathbf{x})}.$$

Setting $\mathbf{x} = D^{1/2}\mathbf{v}$, it follows that

$$\rho\left(D^{-1}A\right) \geq \frac{\left(D^{-1/2}AD^{-1/2}D^{1/2}\mathbf{v}, D^{1/2}\mathbf{v}\right)}{(D^{1/2}\mathbf{v}, D^{1/2}\mathbf{v})} = \frac{(A\mathbf{v},\mathbf{v})}{(D\mathbf{v},\mathbf{v})} = \frac{\|\mathbf{v}\|_1^2}{\|\mathbf{v}\|_0^2}.$$

■

Lemma 7.8 On the eigenvectors of $D^{-1}A$. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and $\boldsymbol{\phi}$ be an eigenvector of $D^{-1}A$ with the eigenvalue λ, i.e.,*

$$D^{-1}A\boldsymbol{\phi} = \lambda\boldsymbol{\phi}.$$

Then it is

$$\|\boldsymbol{\phi}\|_2^2 = \lambda\,\|\boldsymbol{\phi}\|_1^2\,, \quad \|\boldsymbol{\phi}\|_1^2 = \lambda\,\|\boldsymbol{\phi}\|_0^2\,.$$

Proof: The first statement is obtained by multiplying the eigenvalue problem from the left with $\boldsymbol{\phi}^T A$ giving

$$\left(A\boldsymbol{\phi}, D^{-1}A\boldsymbol{\phi}\right) = \lambda\,(A\boldsymbol{\phi}, \boldsymbol{\phi})\,.$$

The second equality follows from multiplying the eigenvalue problem from left with $\boldsymbol{\phi}^T D$

$$\left(\boldsymbol{\phi}, DD^{-1}A\boldsymbol{\phi}\right) = \lambda\,(\boldsymbol{\phi}, D\boldsymbol{\phi})\,.$$

■

Definition 7.9 Smoothing property of an operator. A smoothing operator S is said to satisfy the smoothing property with respect to a symmetric positive definite matrix A if

$$\|S\mathbf{v}\|_1^2 \leq \|\mathbf{v}\|_1^2 - \sigma\,\|\mathbf{v}\|_2^2 \tag{7.5}$$

with $\sigma > 0$ independent of $\mathbf{v}$.

Let $\mathcal{A}$ be a class of matrices. If the smoothing property (7.5) is satisfied for all $A \in \mathcal{A}$ for a smoothing operator S with the same σ, then S is said to satisfy the smoothing property uniformly with respect to $\mathcal{A}$. □

Remark 7.10 *On the smoothing property.* The definition of the smoothing property implies that S reduces the error efficiently as long as $\|\mathbf{v}\|_2$ is relatively large compared with $\|\mathbf{v}\|_1$. However, the smoothing operator will become very inefficient if $\|\mathbf{v}\|_2 \ll \|\mathbf{v}\|_1$. □

Definition 7.11 Algebraically smooth error. An error $\mathbf{v}$ is called algebraically smooth if $\|\mathbf{v}\|_2 \ll \|\mathbf{v}\|_1$. □

Remark 7.12 *Algebraically smooth error.* An algebraically smooth error is a vector for which an iteration with S converges slowly. The term "smooth" for this property is used for historical reasons.

It will be shown now that the damped Jacobi iteration satisfies the smoothing property (7.5) uniformly for symmetric positive definite matrices. □

Lemma 7.13 Equivalent formulation of the smoothing property. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and let the smoothing operator be of the form*

$$S = I - Q^{-1}A$$

with some non-singular matrix Q. Then the smoothing property (7.5) *is equivalent to*

$$\sigma\left(Q^T D^{-1} Q\mathbf{v}, \mathbf{v}\right) \leq \left(\left(Q + Q^T - A\right)\mathbf{v}, \mathbf{v}\right) \quad \forall\ \mathbf{v} \in \mathbb{R}^n. \tag{7.6}$$

Proof: It is

$$\begin{aligned}
\|S\mathbf{v}\|_1^2 &= (AS\mathbf{v}, S\mathbf{v}) = \left(A\left(I - Q^{-1}A\right)\mathbf{v}, \left(I - Q^{-1}A\right)\mathbf{v}\right) \\
&= (A\mathbf{v}, \mathbf{v}) - \left(AQ^{-1}A\mathbf{v}, \mathbf{v}\right) - \left(A\mathbf{v}, Q^{-1}A\mathbf{v}\right) + \left(AQ^{-1}A\mathbf{v}, Q^{-1}A\mathbf{v}\right) \\
&= \|\mathbf{v}\|_1^2 - \left(Q^T Q^{-1}A\mathbf{v}, Q^{-1}A\mathbf{v}\right) - \left(QQ^{-1}A\mathbf{v}, Q^{-1}A\mathbf{v}\right) + \left(AQ^{-1}A\mathbf{v}, Q^{-1}A\mathbf{v}\right) \\
&= \|\mathbf{v}\|_1^2 - \left(\left(Q^T + Q - A\right)Q^{-1}A\mathbf{v}, Q^{-1}A\mathbf{v}\right).
\end{aligned}$$

Hence, the algebraic smoothing property (7.5) is equivalent to the condition that for all $\mathbf{v} \in \mathbb{R}^n$:

$$\begin{aligned} \sigma \|\mathbf{v}\|_2^2 &\leq \left(\left(Q^T + Q - A\right) Q^{-1} A\mathbf{v}, Q^{-1} A\mathbf{v}\right) \quad \Longleftrightarrow \\ \sigma \left(D^{-1} A\mathbf{v}, A\mathbf{v}\right) &\leq \left(\left(Q^T + Q - A\right) Q^{-1} A\mathbf{v}, Q^{-1} A\mathbf{v}\right) \quad \Longleftrightarrow \\ \sigma \left(D^{-1} Q\mathbf{y}, Q\mathbf{y}\right) &\leq \left(\left(Q^T + Q - A\right) \mathbf{y}, \mathbf{y}\right), \end{aligned}$$

with $\mathbf{y} = Q^{-1} A\mathbf{v}$. Since the matrices A and Q are non-singular, $\mathbf{y}$ is an arbitrary vector from $\mathbb{R}^n$. Hence, the statement of the lemma is proved. ■

Theorem 7.14 Algebraic smoothing property of the damped Jacobi method. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix and let $\eta > \rho\left(D^{-1}A\right)$. Then, the damped Jacobi iteration with the damping parameter $\omega \in (0, 2/\eta)$ satisfies the algebraic smoothing property (7.5) uniformly with $\sigma = \omega(2 - \omega\eta)$.*

Proof: The damped Jacobi iteration satisfies the assumptions of Lemma 7.13 with $Q = \omega^{-1} D$. Hence, the algebraic smoothing property (7.5) is eqivalent to (7.6), which gives

$$\begin{aligned} \sigma \left(\frac{D}{\omega^2}\mathbf{v}, \mathbf{v}\right) \leq \left(\frac{2D}{\omega}\mathbf{v}, \mathbf{v}\right) - (A\mathbf{v}, \mathbf{v}) \qquad &\Longleftrightarrow \qquad (A\mathbf{v}, \mathbf{v}) \leq \left(\left(\frac{2}{\omega} - \frac{\sigma}{\omega^2}\right) D\mathbf{v}, \mathbf{v}\right) \\ &\Longleftrightarrow \qquad \|\mathbf{v}\|_1^2 \leq \left(\frac{2}{\omega} - \frac{\sigma}{\omega^2}\right) \|\mathbf{v}\|_0^2 . \end{aligned} \tag{7.7}$$

From inequality (7.4) and the assumption on η it follows for all $\mathbf{v} \in \mathbb{R}^n$ that

$$\|\mathbf{v}\|_1^2 \leq \rho\left(D^{-1}A\right) \|\mathbf{v}\|_0^2 < \eta \|\mathbf{v}\|_0^2 .$$

Thus, if

$$\eta \leq \left(\frac{2}{\omega} - \frac{\sigma}{\omega^2}\right), \tag{7.8}$$

then (7.7) is satisfied (sufficient condition) and the damped Jacobi iteration fulfills the algebraic smoothing property. One obtains from (7.8)

$$\sigma \leq 2\omega - \eta\omega^2 = \omega\left(2 - \omega\eta\right).$$

Obviously it is $\sigma > 0$ if $\omega \in (0, 2/\eta)$. ■

Remark 7.15 *On the algebraic smoothing property.*

- The optimal value of ω, which gives the largest σ is $\omega^* = 1/\eta$, such that $\sigma = 1/\eta$. This statement can be proved by standard calculus, *exercise*.
- The algebraic smoothing property can be proved also for the Gauss–Seidel iteration.

□

Remark 7.16 *The algebraic smooth error for M-matrices.* The meaning of "$\mathbf{v}$ being an algebraic smooth error" will be studied in some more detail for symmetric positive definite M-matrices. This class of matrices was introduced in the course on numerical methods for convection-dominated problems.

An algebraic smooth error satisfies $\|\mathbf{v}\|_2 \ll \|\mathbf{v}\|_1$. By (7.2), this property implies

$$\|\mathbf{v}\|_1 \ll \|\mathbf{v}\|_0 . \tag{7.9}$$

For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, it is, *exercise*,

$$\|\mathbf{v}\|_1 = \frac{1}{2} \sum_{i,j=1}^{n} (-a_{ij}) (v_i - v_j)^2 + \sum_{i=1}^{n} s_i v_i^2, \quad \text{with } s_i = \sum_{j=1}^{n} a_{ij}$$

being the i-th row sum of A. It follows from (7.9) that

$$\frac{1}{2}\sum_{i,j=1}^{n}(-a_{ij})\,(v_i - v_j)^2 + \sum_{i=1}^{n} s_i v_i^2 \ll \sum_{i=1}^{n} a_{ii} v_i^2. \tag{7.10}$$

Let A be an M-matrix. Then $a_{ij} \leq 0$, i.e., $|a_{ij}| = -a_{ij}$ for $i \neq j$. In many applications, it is $s_i = 0$. Then, from (7.10) it follows on the average for each i (consider just a fixed i)

$$\sum_{j=1}^{n} \frac{|a_{ij}|}{a_{ii}} \frac{(v_i - v_j)^2}{v_i^2} \ll 1.$$

In the sum, there are only nonnegative terms. Thus, if $|a_{ij}|/a_{ii}$ is large, then $(v_i - v_j)^2/v_i^2$ has to be small such that the sum becomes small. One says, schemes which satisfy the smoothing property (7.5) smooth the error along the so-called strong connections, i.e., where $|a_{ij}|/a_{ii}$ is large, since for these connections a good smoothing can be expected on the given grid. This property implies that the corresponding nodes i and j do not need to be both on the coarse grid. □

7.3 Coarsening

Remark 7.17 *Goal.* Based on the matrix information only, one has to choose in the graph of the matrix nodes which become coarse nodes and nodes which stay on the fine grid. There are several strategies for coarsening. In this course, a standard way will be described. It will be restricted to the case that $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite M-matrix. □

Definition 7.18 Strong coupling. A variable (node) i is said to be strongly coupled to another variable j if

$$-a_{ij} \geq \varepsilon_{\text{str}} \max_{a_{ik}<0} |a_{ik}|$$

with fixed $\varepsilon_{\text{str}} \in (0,1)$. The set of all strong couplings of i is denoted by

$$S_i = \{j \in N_i \ : \ i \text{ is strongly coupled to } j\}.$$

The set S_i^T of strong transposed couplings of i consists of all variables j which are strongly coupled to i

$$S_i^T = \{j \in N_i \ : \ i \in S_j\}.$$

□

Remark 7.19 *On strong couplings.*

- Even for symmetric matrices, the relation of being strongly coupled is in general not symmetric. Consider, e.g.,

$$A = \begin{pmatrix} 5 & -1 & -0.1 \\ -1 & 3 & -0.1 \\ -0.1 & -0.1 & 3 \end{pmatrix}, \quad \varepsilon_{\text{str}} = 0.25.$$

  Then, one gets $S_1 = \{2\}, S_2 = \{1\}, S_3 = \{1,2\}$, such that $S_1 = \{2,3\}, S_2 = \{1,3\}, S_3 = \emptyset$.
- The actual choice of ε_{str} is in practical computations not very critical. Values of around 0.25 are often used.

□

Remark 7.20 *Aspects of the coarsening process.* In the coarsening process, one has to pay attention to several aspects.

- The number of coarse nodes (C-nodes) should not be too large, such that the dimension of the coarse system is considerably smaller than the dimension of the fine system.
- Nodes i and j, which are strongly coupled, have a small relative error

$$(e_i - e_j)^2 / e_i^2$$

  such that a coarse grid correction of this error is not necessary. That means, it will be inefficient to define both nodes as coarse nodes.
- All fine nodes (F-nodes) should have a substantial coupling to neighboring C-nodes. In this way, the F-nodes obtain sufficient information from the C-nodes.
- The distribution of the C-nodes and F-nodes in the graph should be reasonably uniform.

□

Remark 7.21 *A standard coarsening procedure.* A standard coarsening procedure starts by defining some first variable i to become a C-node. Then, all variables j that are strongly coupled with i, i.e., all $j \in S_i^T$, become F-nodes. Next, from the remaining undecided variables, another one is defined to become a C-node and all variables which are strongly coupled to it and are still undecided become F-nodes, and so on. This process stops if all variables are either C-nodes or F-nodes.

To obtain a uniform distribution of the C-nodes and F-nodes, the process of choosing C-nodes has to be done in a certain order. The idea consists in starting with some variable and to continue from this variable until all variables are covered. Therefore, an empirical "measure of importance" λ_i for any undecided variable to become a C-node is introduced

$$\lambda_i = \left|S_i^T \cap U\right| + 2\left|S_i^T \cap F\right|, \quad i \in U, \tag{7.11}$$

where U is the current set of undecided variables, F the current set of F-nodes and $|\cdot|$ is the number of elements in a set. One of the undecided variables with the largest value of λ_i will become the next C-node. After this choice, all variables which are strongly coupled to the new C-node become F-nodes and for the remaining undecided variables, one has to update their measure of importance.

With the measure of importance (7.11), there is initially the tendency to pick variables which are strongly coupled with many other variables to become C-nodes, because $|U|$ is large and $|F|$ is small, such that the first term dominates. Later, the tendency is to pick as C-nodes those variables on which many F-nodes depend strongly, since $|F|$ is large and $|U|$ is small such that the second term in λ_i becomes dominant. Thus, the third point of Remark 7.20 is taken into account. □

Example 7.22 *A standard coarsening procedure.* Consider a finite difference discretization of the Laplacian in the unit square with the five point stencil. Assuming that the values at the boundary are known, the finite difference scheme gives for the interior nodes i the following matrix entries, apart of a constant factor,

$$a_{ij} = \begin{cases} 4 & \text{if } i = j, \\ -1 & \text{if } j \text{ is left, right, upper, or lower neighbor of } i, \\ 0 & \text{else.} \end{cases}$$

Taking an arbitrary ε_{str}, then each node i is strongly coupled to its left, right, upper, and lower neighbor. Consider a 5×5 patch and choose some node as C-node. In

the first step, one obtains

$$\begin{array}{ccccc} U & U & U & U & U \\ U & U & U & U & U \\ U & U & F & U & U, \\ U & F & C & F & U \\ U & U & F & U & U \end{array}$$

where for U it is $\lambda_i = 2 + 2 \cdot 2 = 6$ and for U it is either $\lambda_i = 4 + 2 \cdot 0 = 4$ or $\lambda_i = 3 + 2 \cdot 1 = 5$. The next step gives, e.g.,

$$\begin{array}{ccccc} U & U & U & U & U \\ U & U & U & F & U \\ U & U & F & C & F, \\ U & F & C & F & U \\ U & U & F & U & U \end{array}$$

with $\lambda_i = 2 + 2 \cdot 2 = 6$ for U and $\lambda_i \leq 5$ else. Continuing this process leads to

$$\begin{array}{ccccc} U & U & F & U & U \\ U & F & C & F & U \\ F & C & F & C & F, \\ U & F & C & F & U \\ U & U & F & U & U \end{array}$$

and so on.

In this particular example, one obtains a similar coarse grid as given by a geometric multigrid method. However, in general, especially with non-symmetric matrices, the coarse grid of the AMG looks considerably different than the coarse grid of a geometric multigrid method. □

Remark 7.23 *On coarsening strategies.*

- In the standard coarsening scheme, none of the C-nodes is strongly coupled to any of the C-nodes created prior itself. However, since the relation of being strongly coupled might be non-symmetric, in particular for non-symmetric matrices, this property may not be true the other way around. Numerical experience shows that in any case the resulting set of C-nodes is close to a maximal set of variables which are not strongly coupled among each other.
- Other ways of coarsening can be found, e.g., in K. Stüben "Algebraic Multigrid (AMG): An introduction with applications", which is part of Trottenberg et al. (2001).

□

7.4 Prolongation

Remark 7.24 *Prolongation.* The last component of an AMG, which has to be defined, is the prolongation. It will be matrix-depend, in contrast to geometric multigrid methods. □

Remark 7.25 *Construction of an prolongation operator.* To motivate the construction of an prolongation operator, once more the meaning of an error to be algebraically smooth will be discussed. From the geometric multigrid methods, it is known that the prolongation has to work well for smooth functions, see Remark 4.11. By definition, an algebraic smooth error is characterized by

$$S\mathbf{e} \approx \mathbf{e}$$

or

$$\|\mathbf{e}\|_2 \ll \|\mathbf{e}\|_1 .$$

In terms of the residual

$$\mathbf{r} = \mathbf{f} - A\mathbf{v} = A\mathbf{u} - A\mathbf{v} = A(\mathbf{u} - \mathbf{v}) = A\mathbf{e},$$

this inequality means that

$$\left(D^{-1}A\mathbf{e}, A\mathbf{e}\right) \ll (A\mathbf{e}, \mathbf{e}) \quad \Longleftrightarrow \quad \left(D^{-1}\mathbf{r}, \mathbf{r}\right) \ll (\mathbf{r}, \mathbf{e}) .$$

One term in both inner products is the same. One can interprete this inequality in the way that on the average, algebraic smooth errors are characterized by a scaled residual (first argument on the left-hand side) to be much smaller than the error (second argument on the right-hand side). On the average, it follows that

$$a_{ii}^{-1} r_i^2 \ll |r_i e_i| \quad \Longleftrightarrow \quad |r_i| \ll a_{ii} |e_i| .$$

Thus, $|r_i|$ is close to zero and one can use the approximation

$$0 \approx r_i = a_{ii} e_i + \sum_{j \in N_i} a_{ij} e_j. \tag{7.12}$$

Let i be a F-node and $P_i \subset C_{\text{nod}}$ a subset of the C-nodes, where the set of C-nodes is denoted by C_{nod}, the so-called interpolatory points. The goal of the prolongation consists in obtaining a good approximation of e_i using information from the coarse grid, i.e., from the C-nodes contained in P_i. Therefore, one likes to compute prolongation weights ω_{ij} such that

$$e_i = \sum_{j \in P_i} \omega_{ij} e_j \tag{7.13}$$

and e_i is a good approximation for any algebraic smooth error which satisfies (7.12). □

Remark 7.26 *Direct prolongation.* Here, only the so-called direct prolongation in the case of A being an M-matrix will be considered. Direct prolongation means that $P_i \subset C_{\text{nod}} \cap N_i$, i.e., the interpolatory nodes are a subset of all coarse nodes which are coupled to i. Inserting the ansatz (7.13) into (7.12) gives

$$e_i = \sum_{j \in P_i} \omega_{ij} e_j = -\frac{1}{a_{ii}} \sum_{j \in N_i} a_{ij} e_j. \tag{7.14}$$

If $P_i = N_i$, then the choice $\omega_{ij} = -a_{ij}/a_{ii}$ will satisfy this equation. But in general, $P_i \subsetneq N_i$. If there are sufficiently many nodes which are strongly connected to i contained in P_i, then for the averages it holds

$$\frac{1}{\sum_{j \in P_i} a_{ij}} \sum_{j \in P_i} a_{ij} e_j \approx \frac{1}{\sum_{j \in N_i} a_{ij}} \sum_{j \in N_i} a_{ij} e_j.$$

Inserting this relation into (7.14) leads to the proposal for using matrix-dependent prolongation weights

$$\omega_{ij} = -\left(\frac{\sum_{k \in N_i} a_{ik}}{\sum_{k \in P_i} a_{ik}} \right) \frac{a_{ij}}{a_{ii}} > 0, \quad i \in F, j \in P_i.$$

Summation of the weights gives

$$\sum_{j \in P_i} \omega_{ij} = -\left(\frac{\sum_{k \in N_i} a_{ik}}{\sum_{k \in P_i} a_{ik}} \right) \frac{\sum_{j \in P_i} a_{ij}}{a_{ii}} = \frac{a_{ii} - s_i}{a_{ii}} = 1 - \frac{s_i}{a_{ii}},$$

where s_i is the sum of the i-th row of A. Thus, if $s_i = 0$, then $\sum_{j \in P_i} \omega_{ij} = 1$ such that constants are prolongated exactly. □

7.5 Concluding Remarks

Example 7.27 *Behavior of an AMG for the Poisson equation.* The same situation as in Example 2.5 will be considered. In the code MooNMD, a simple AMG is implemented. The number of iterations and computing times for applying this method as solver or as preconditioner in a flexible GMRES method are presented in Table 7.1.

Table 7.1: Example 7.27. Number of iterations and computing times (14/01/24 on a HP BL460c Gen8 2xXeon, Eight-Core 2700MHz). The number of degrees of freedom (d.o.f.) includes the Dirichlet values. The time for the setup of the AMG is included into the total solution time.

level	h	d.o.f.	AMG		FGMRES+AMG		setup time
			ite	time	ite	time	(FGMRES+AMG)
1	1/4	25	1	0	1	0	0
2	1/8	81	1	0	1	0	0
3	1/16	289	34	0.01	18	0.01	0
4	1/32	1089	41	0.02	19	0.01	0.01
5	1/64	4225	45	0.13	21	0.08	0.03
6	1/128	16641	47	0.69	22	0.43	0.15
7	1/256	66049	51	3.81	23	2.66	1.32
8	1/512	263169	49	25.08	24	14.82	7.28
9	1/1024	1050625	50	157.14	24	119.96	84.95
10	1/2048	4198401	50	1500.75	24	1333.09	1103.40

It can be seen, that using AMG as preconditioner is more efficient than using it as solver. The number of iterations for both applications of AMG is constant independently of the level. However, the solution time does not scale with the number of degrees of freedom. The reason is that in the used implementation, the time for constructing the AMG does not scale in this way but much worse. Comparing the results with Table 2.2, one can see that AMG is not competitive with a geometric multigrid method, if the geometric multigrid method works well. □

Remark 7.28 *Concluding remarks.*

- A number of algebraic results for AMGs is available, see the survey paper of K. Stüben. But there are still many open questions, even more than for the geometric multigrid method.
- As seen in Example 7.27, in problems for which a geometric multigrid method can be applied efficiently, the geometric multigrid method will in general outperform AMG. But there are classes of problems for which AMG is as efficient or even more efficient than a geometric multigrid method. One of the most important fields of application for AMG are problems for which a geometric multigrid method cannot be performed.

□