Méthodes Itératives

Carola Kruse, Alena Kopanicakova et Ronan Guivarch

Cours 4, 18/03/2025
Méthodes Multigrilles

Algebraic Multigrid Methods (AMG)

Problem setting

What do we do, if no mesh information is available?

- For example, in industrial software, not possible to ‘touch’ the discretization of the problem, but the matrix has to be solved efficiently.

What if the mesh is highly unstructured or irregular?

- How do we then define a coarse grid?

We address these problems by a method called Algebraic Multigrid.

Components of algebraic multigrid

In algebraic multigrid, we take information of the matrix itself, not of the grid.

The components are however principally the same as for geometric multigrid, these are

- A hierarchy of levels,
- A smoother,
- A prolongation operator,
- A restriction operator,
- Coarse grid operators.

But if no grid is available, how do we define the coarse grid, and thus prolongation/ restriction / coarse grid solutions?

Components of algebraic multigrid

- A level or grid is a set of unknowns of degree of freedoms.
- We start from the finest level and ‘remove’ unknowns to obtain a coarse grid.

Main tasks

- A hierarchy of levels has to be defined fully automatically. This is done by using only information from the matrix on the current grid.
- We have to define an appropriate prolongation operator.
- The restriction operator is defined as the transposed of the proplongation, i.e. $I_f^c = \left(I_c^f\right)^T$.

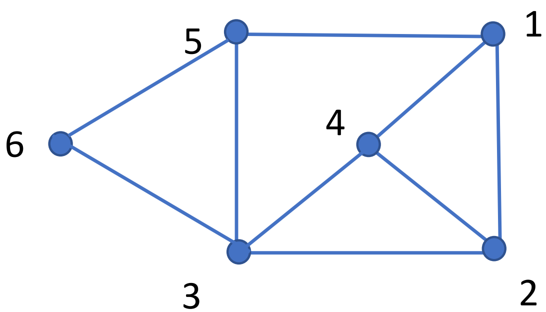
In the following, we will consider only symmetric M-matrices: symmetric, positive definite ($u^T A u > 0$) and positive diagonal entries and nonpositive off-diagonal ones.

The grid

Graph of a matrix

Let a_{ij} be the entries of A. We associate the vertices of the matrix and draw an edge between the i-th and j-th vertice if $a_{ij} \neq 0$.

$$A = \begin{pmatrix} * & * & & * & * & \\ * & * & * & * & & \\ & * & * & * & * & * \\ * & * & * & * & & \\ * & & * & & * & * \\ & & * & & * & * \end{pmatrix}$$



We have:

- $V = \{v_1, v_2, \dots, v_n\}$ set of n ordered vertices
- E a set of egdes, such that edge $\mathrm{e}_{i,j}$ connects v_i and v_j
- $\Omega = G_A(V, E)$ be the graph of the matrix
- For a vertex v_i, the set of its neighbor vertices N_i is defined by

$$\mathrm{N_i} = \{\mathrm{v_j} \in V : \mathrm{e}_{i,j} \in E\}$$

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

Smoothness/Influence/Dependence

Algebraic Smoothness

Recall in the weighted Jacobi method, the error propagation can be written as

$$\boldsymbol{e}_{i+1} = (I - \omega D^{-1}A)\boldsymbol{e}_i$$

Remember that the weighted Jacobi relaxation made great progress towards convergence in the first few steps, but then stalls and only little improvement is made after. We define this point as algebraically smooth.

By our definition, algebraic smoothness means that e_{i+1} is not significantly less than e_i. Thus it is characterized by

$$\|(I - \omega D^{-1}A)\boldsymbol{e}\|_A \approx \|\boldsymbol{e}\|_A$$

This translates into

$$(D^{-1}A\boldsymbol{e}, A\boldsymbol{e}) \ll (\boldsymbol{e}, A\boldsymbol{e})$$

and also

$$(I - \omega D^{-1}A)\boldsymbol{e} \approx \boldsymbol{e} \quad \Rightarrow \omega D^{-1}A\boldsymbol{e} \approx 0 \quad \Rightarrow \boldsymbol{r} \approx 0$$

That is, smooth error has relatively small residuals.

Influence and dependence

- Of course it takes all of the equations to determine any value precisely, but …
- … due to the diagonal dominance of A (is an M-matrix), we can thus say that the job of the i-th equation is to determine the value of u_i.

$$\begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -0.5 & 0 \\ 0 & -0.5 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix}$$

- To obtain a more precise estimate of u_i, which other variables are most important in the i-th equation?

- If the coefficient a_{ij}, which multiplies u_j in the *i*-th equation, is large relative to the other coefficients in the *i*-th equation, then a small change in the value of u_j has more effect on u_i than a small change in other variables in the *i*-th equation.

Influence and dependence

Definition 1

Given a threshold value $0 < \theta \leq 1$, the variable (point) u_i strongly depends on the variable (point) u_j, if

$$-a_{ij} \geq \theta \max_{k \neq i}\{-a_{ik}\}$$

Definition 2

If the variable u_i strongly depends on the variable u_j, then the variable u_j strongly influences the variable u_i.

Next steps…

As in geometric multigrid,

- Select a coarse grid so that the smooth components can be represented accurately,

- Select an interpolation operator, so that the smooth components can be accurately transferred from the coarse grid to the fine grid,

- Define a restriction operator and a coarse grid version of A using the variational (Galerkin) properties.

Interpolation operator

Interpolation operator

We now assume that we have already found the coarse grid points and fine grid points.

- We want a partitioning of the indices $\{1, 2, \dots, n\} = C \cup F$.
- The variables $i \in C$ are the coarse grid variables.
- Of course the $i \in C$ are also fine grid variables.
- However, we define $i \in F$ as those variables that are *only* fine grid variables.

Next assume, that $\mathrm{e}_i, i \in C$, is a set of values on the coarse grid representing the smooth error

What do we know about e_i that allows us to build an interpolation operator that is accurate?

Algebraic smooth error

- Let a C-point j strongly influence an F-point i.
- One can show that for smooth error holds on average

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

- Since there are only non-negative terms in the sum and sum << 1, each term has to be small.
- If $\frac{|a_{ij}|}{a_{ii}}$ is not small (if e_i strongly depends on e_j), then $(e_i - e_j)^2$ must be small.

Smooth error varies slowly in the direction of strong connection.

Interpolation operator

- Since the error varies slowly in the direction of the strong connections, thus the fine-grid quantity u_i could be interpolated from the coarse grid quantity u_j.

Definition

For each fine-grid point i, we define N_i as the neighborhood of i, whenever $j \neq i$ and $a_{ij} \neq 0$. These can be divided into

- The set C_i : neighboring coarse-grid points that strongly influence i, i.e. the coarse-grid interpolatory set for i.
- The set D_i^s : neighboring fine-grid points that strongly influence i.
- The set D_i^w: points that do not strongly influence i, possibly both coarse- and fine-grid points. It is called the set of weakly connected neighbors.

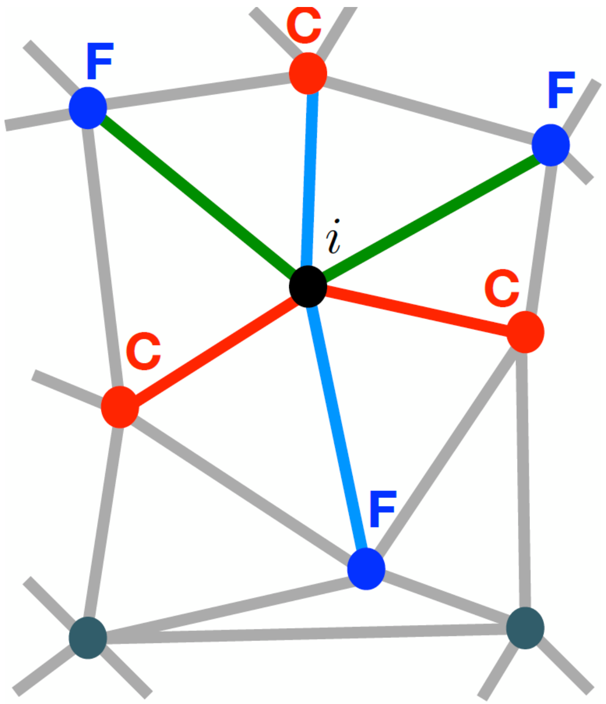


Figure taken of Copper Mountain AMG tutorial 2021

Interpolation operator

We want to define an interpolation operator for the *i*-th component of $I_c^f \boldsymbol{e}$, that is of the form

$$\left(I_c^f \boldsymbol{e}\right)_i = \begin{cases} e_i & if\ i \in C, \\ \sum_{j \in C_i} w_{ij} e_j, & if\ i \in F, \end{cases}$$

with weights w_{ij}, that must be determined.

Recall that smooth error is characterized by $r \approx 0$. We can write the i-th component of this condition as

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

We want to express the second and third term on the right in terms of e_i or e_j of strongly connected coarse grid points.

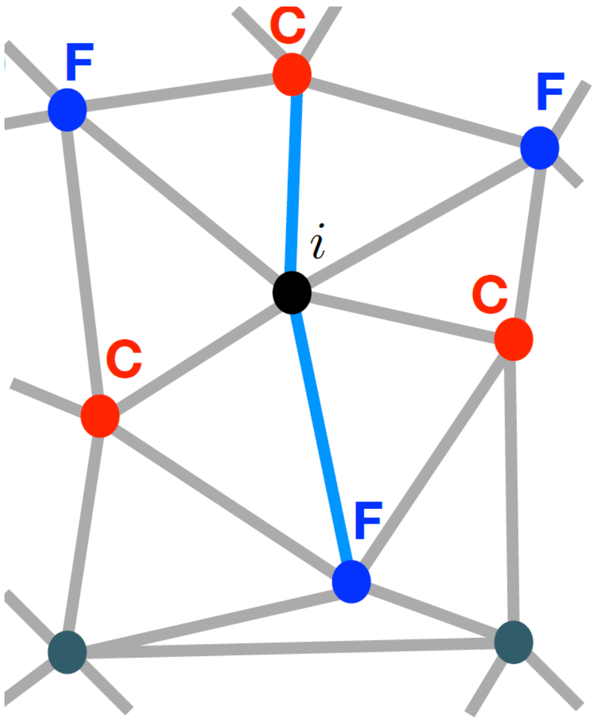
Interpolation operator

For the sum over the weakly connected neighbors D_i^w, we approximate

$$\sum_{j \in D_i^w} a_{ij} e_j \approx \sum_{j \in D_i^w} a_{ij} e_i$$

Justification:

- If we have underestimated the dependence, so that e_i actually depends strongly on the value of some of the points in D_i^w, then $e_i \approx e_j$ (the smooth error varies slowly).
- If e_i indeed does not depend strongly on the points in D_i^w, then the corresponding value of a_{ij} will be small and any error done in this assignement will be rather insignificant.



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

Interpolation operator

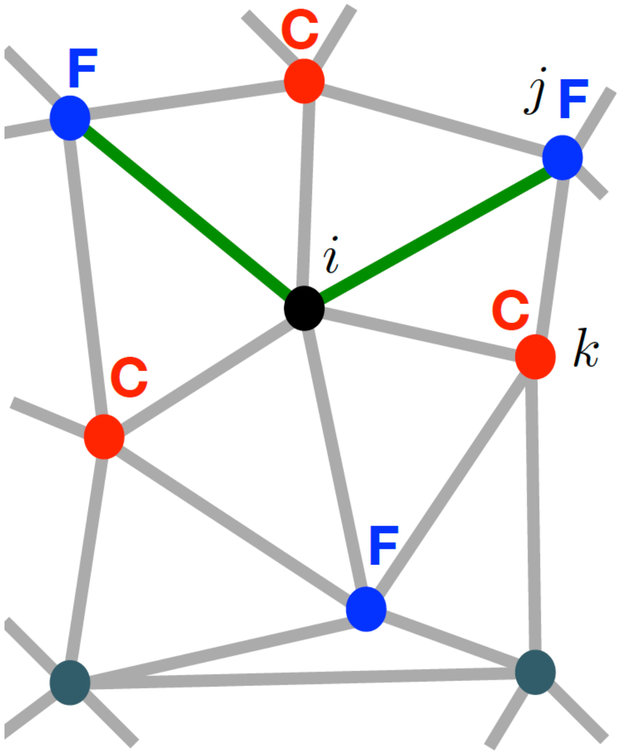
- We could use the same argumentation for the sum over the strongly connected fine-grid points D_i^s, that $e_i \approx e_j$ and distribute the values on the diagonal.
- However experience has shown that the interpolation results are better when the values are distributed over the strongly influencing coarse-grid values C_i.
- We thus want to express e_j as a linear combination over the coarse interpolatory set, i.e. $e_k \in C_i$.

We do this for each fixed $j \in D_i^s$, by making the approximation

$$e_j \approx \frac{\sum_{k\in C_i} a_{jk} e_k}{\sum_{k\in C_i} a_{jk}}$$

After substitution into the previous equation and some computations, we obtain

$$w_{ij} = \frac{a_{ij} + \sum_{m\in D_i^s}\left(\frac{a_{im}a_{mj}}{\sum_{k\in C_i} a_{mk}}\right)}{a_{ii} + \sum_{n\in D_i^w} a_{in}}$$

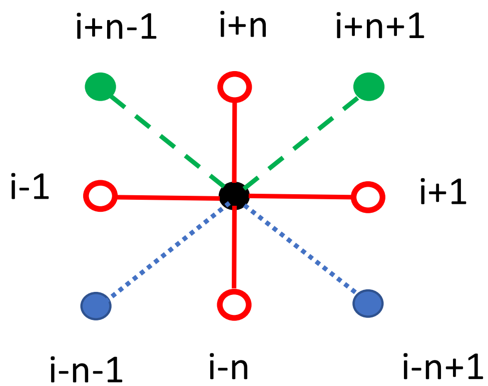


Example

We take an operator A defined, on a uniform nxn grid, by the stencil

$$\begin{bmatrix} -\frac{1}{2} & -2 & -\frac{1}{2} \\ -1 & \frac{29}{4} & -1 \\ -\frac{1}{8} & -2 & -\frac{1}{8} \end{bmatrix}$$

With $\theta = 0.2$, we have the connections:



C_i, D_i^s, D_i^w

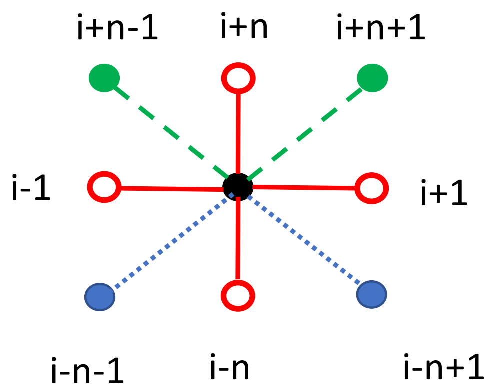
We then have: $\frac{29}{4} e_i = 2e_{i+n} + 2e_{i-n} + e_{i-1} + e_{i+1} + \frac{1}{2} e_{i+n-1} + \frac{1}{2} e_{i+n+1} + \frac{1}{8} e_{i-n-1} + \frac{1}{8} e_{i-n+1}$

Move D_i^w points to the diagonal: $\left(\frac{29}{4} - \frac{1}{8} - \frac{1}{8}\right) e_i = 2e_{i+n} + 2e_{i-n} + e_{i-1} + e_{i+1} + \frac{1}{2} e_{i+n-1} + \frac{1}{2} e_{i+n+1}$

Example

$$\begin{bmatrix} -\frac{1}{2} & -2 & -\frac{1}{2} \\ -1 & \frac{29}{4} & -1 \\ -\frac{1}{8} & -2 & -\frac{1}{8} \end{bmatrix}$$

With $\theta = 0.2$, we have the connections:



C_i, D_i^s, D_i^w

See that the points in D_i^s are by 'their stencil' strongly connected to neighboring points contained in C_i. We thus approximate them in terms of these points by

$$e_{i+n-1} \approx \frac{-2e_{i-1} - e_{i+n}}{-(2+1)} \qquad e_{i+n+1} \approx \frac{-2e_{i+1} - e_{i+n}}{-(2+1)}$$

We finally get

$$e_i = \frac{7}{21} e_{i+n} + \frac{6}{21} e_{i-n} + \frac{4}{21} e_{i+1} + \frac{4}{21} e_{i-1}$$

Next steps…

As in geometric multigrid,

- Select a coarse grid so that the smooth components can be represented accurately,

- Select an interpolation operator, so that the smooth components can be accurately transferred from the coarse grid to the fine grid,

- Define a restriction operator and a coarse grid version of A using the variational (Galerkin) properties.

Coarse Grid

Selection of the coarse grid

Goal: Select a coarse grid,
- Such that smooth error is well represented,
- From which smoth functions can be interpolated accurately
- That has substanctially fewer points than the fine grid.

- We want a partitioning of the indices $\{1, 2, \dots, n\} = C \cup F$.
- The variables $i \in C$ are the coarse grid variables.
- Of course the $i \in C$ are also fine grid variables.
- However, we define $i \in F$ as those variables that are *only* fine grid variables.

Selection of the coarse grid

We need the sets:

- $S_i = \{j \ : -a_{ij} \geq \theta \max_{k \neq i}(-a_{ik})\}$
- $S_i^T = \{j : i \in S_j\}$

- **Two heuristic criteria**
- **H-1**: For each F-point i, every point j in S_i that strongly influences i either should be in the set of coarse grid nodes C or should strongly depend on at least one point in C .

- **H-2**: The set of coarse points C should be a maximal subset of all points with the property that no C-point depends on another C-point.

The coloring scheme

1. Each point is assigned a measure of its potential quality as coarse grid point. Therefore, we count for each i the number of strongly influenced points (this is the set S_i^T) and call it λ_i .

2. We select a point with maximum λ_i value as first coarse-grid point.

3. The selected coarse point strongly influences several of the other points and should appear in the interpolation formula for each of them ⇒ Points that depend strongly on i become F-points, thus all S_i^T gets assigned to F.

4. We look at other points that strongly influence these new F-points as potential C-points. Their value could be useful for accurate interpolation.

5. Therefore, for each new F-point j in S_i^T, we increment the measure λ_k of each unassigned point k that strongly influences j, this is each unassigned member of $k \in S_i$

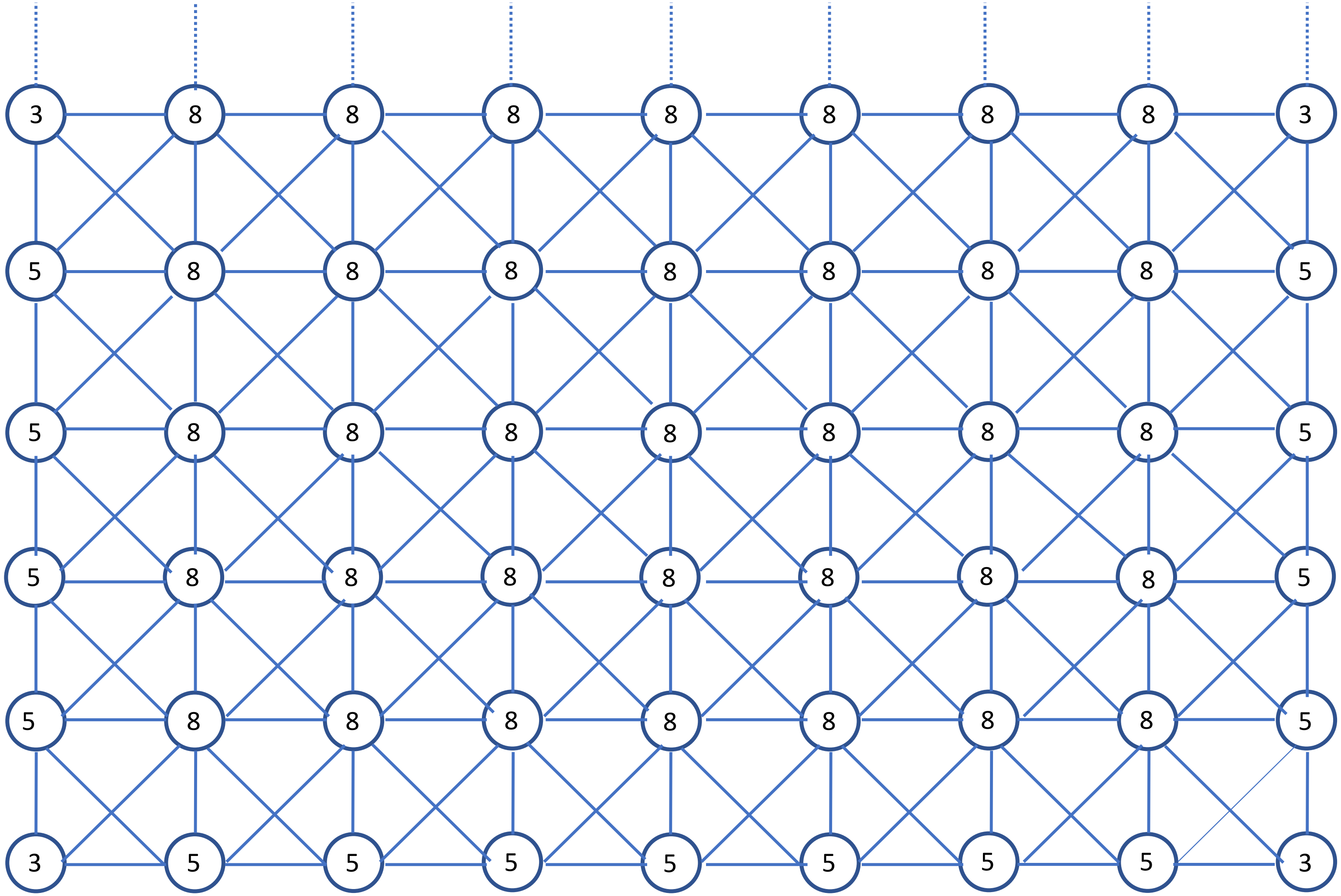
Best visualized by an example…

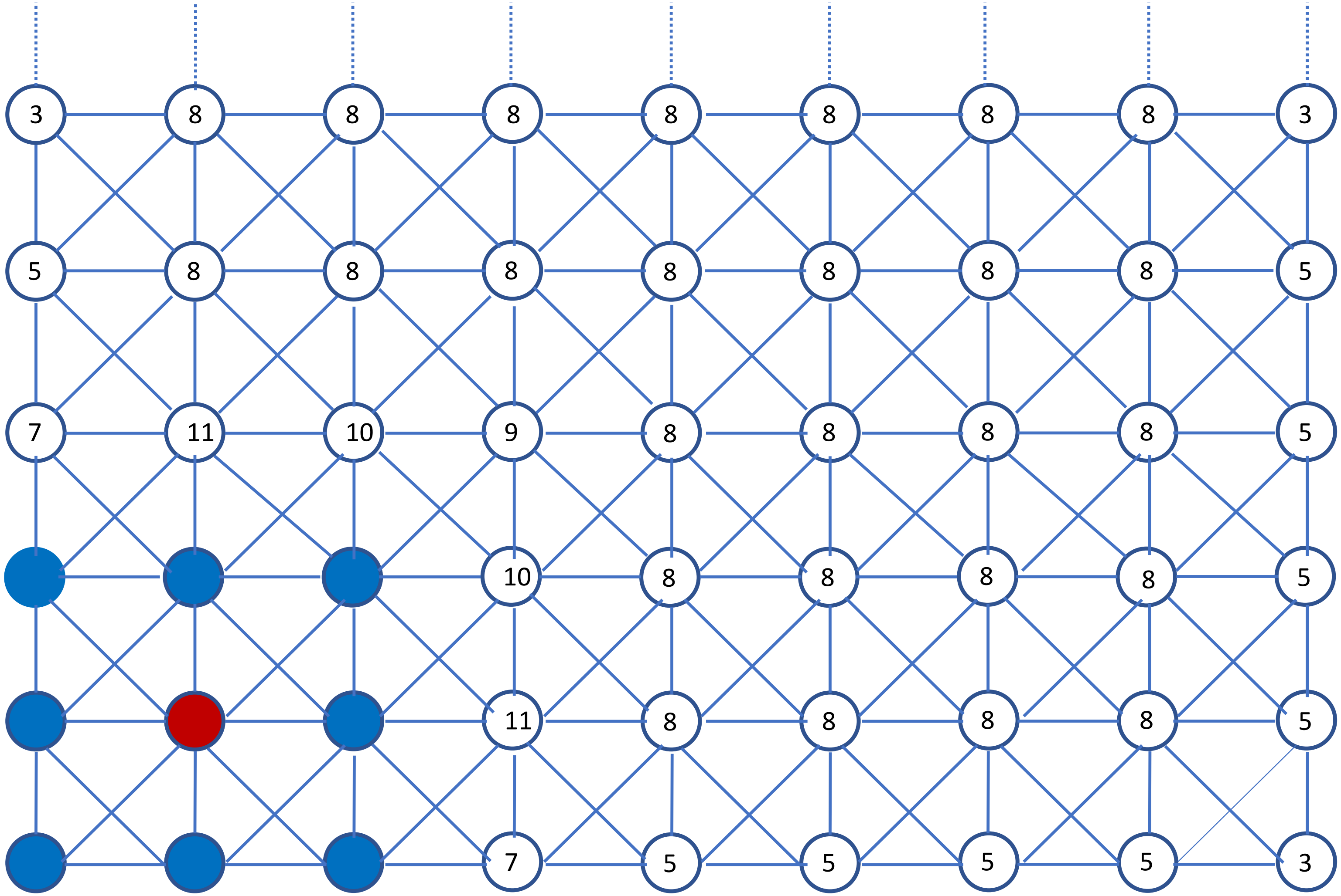
Example

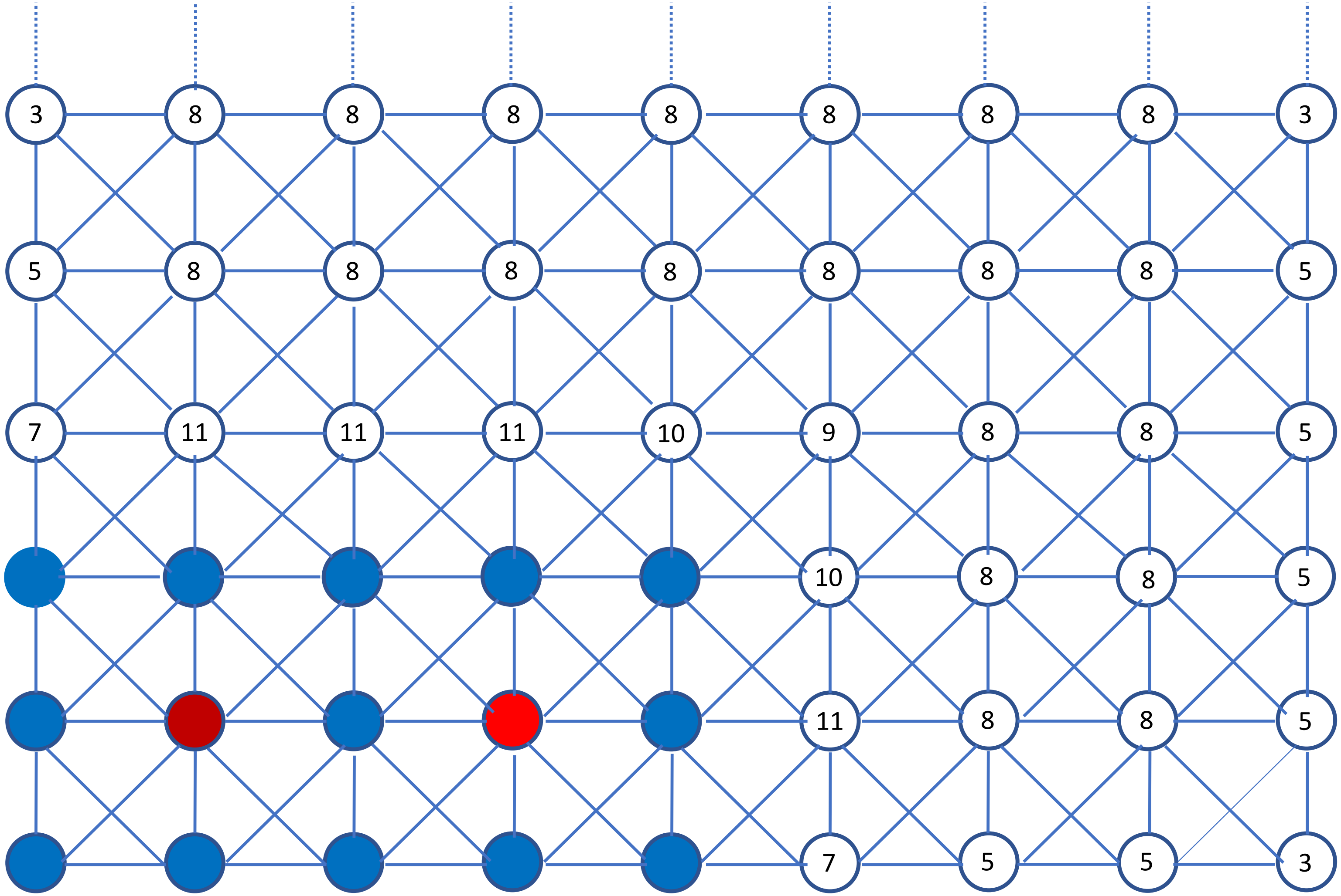
We use a nine-point stencil for a Laplacian on a uniform grid. The operator stencil is

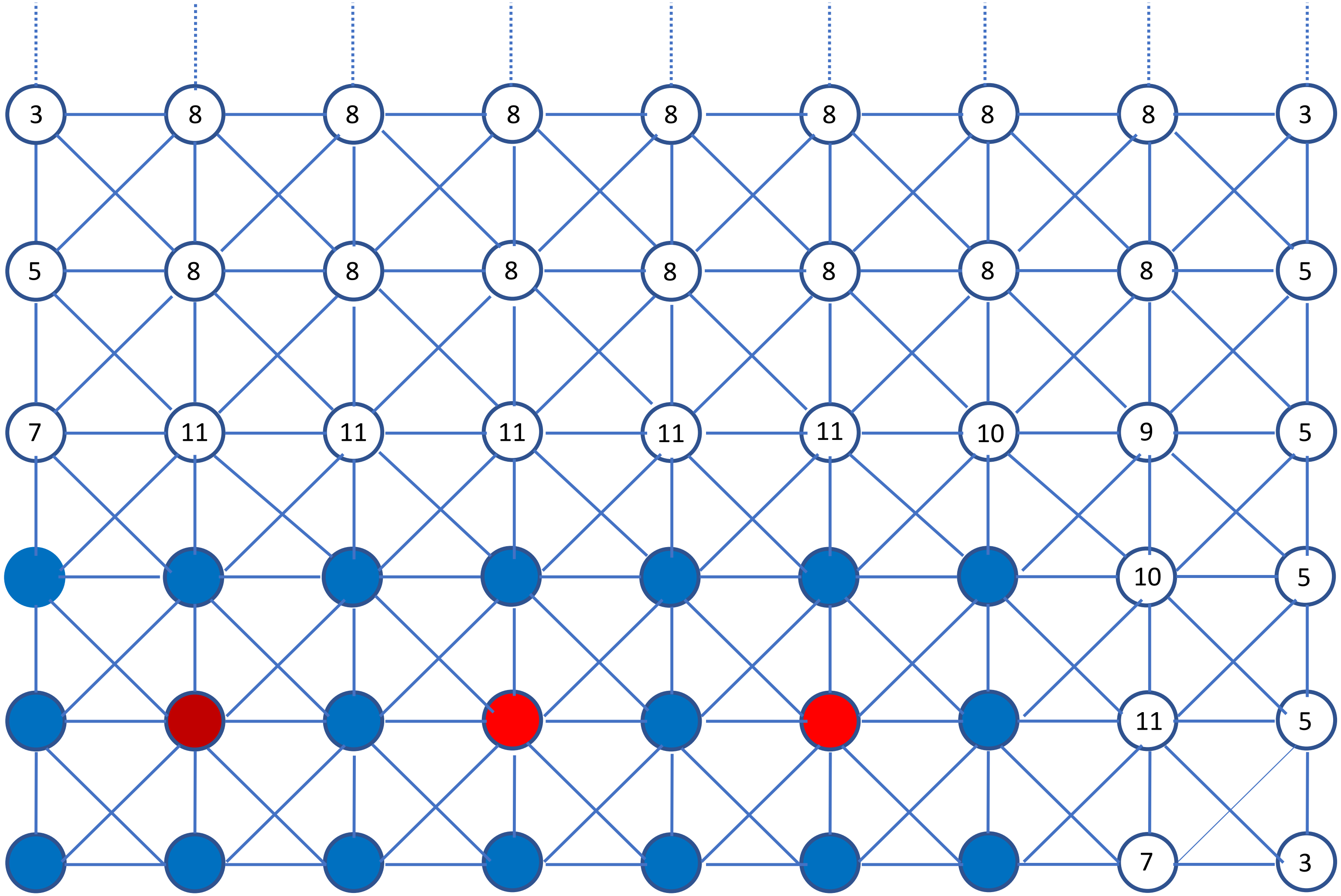
$$\frac{1}{h^2}\begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

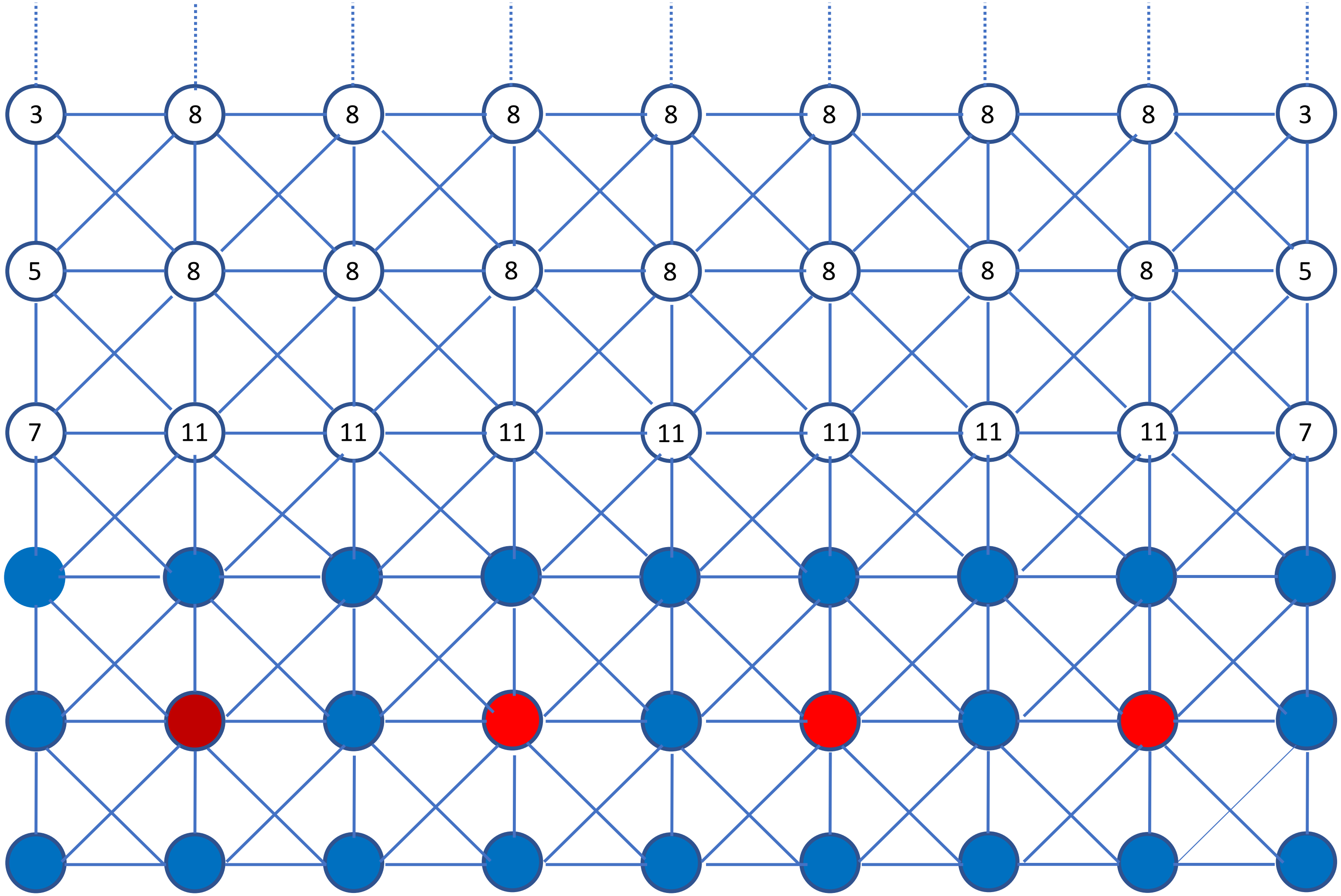
No matter which θ, every connection is of strong dependence $\Rightarrow$ each point strongly influences and depends strongly upond each of its neighbors.

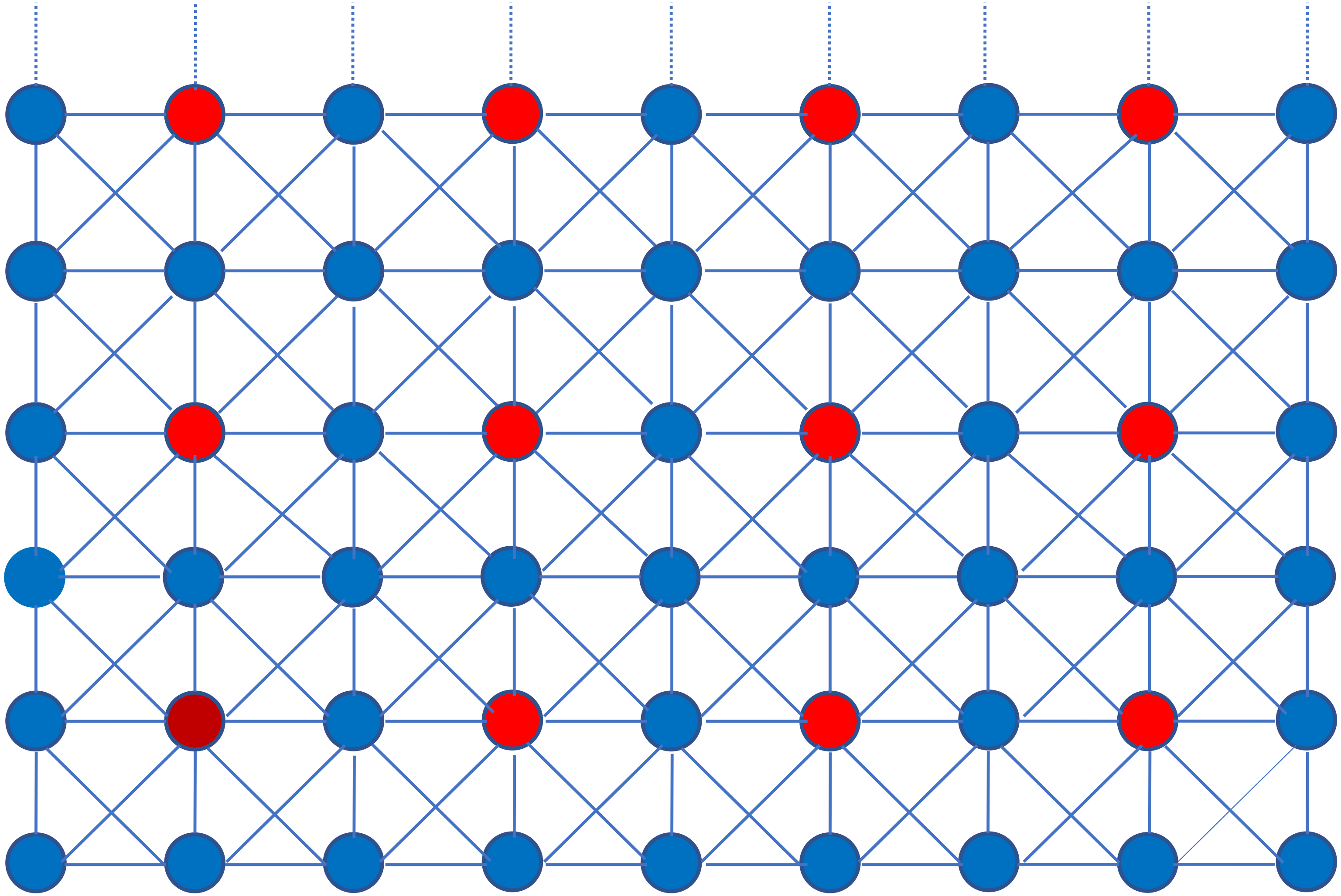

3 8 8 8 8 8 8 8 3
5 8 8 8 8 8 8 8 5
5 8 8 8 8 8 8 8 5
5 8 8 8 8 8 8 8 5
5 8 8 8 8 8 8 8 5
3 5 5 5 5 5 5 5 3


3 8 8 8 8 8 8 8 3
5 8 8 8 8 8 8 8 5
7 11 10 9 8 8 8 8 5
10 8 8 8 8 5
11 8 8 8 8 5
7 5 5 5 5 3


3 8 8 8 8 8 8 8 3
5 8 8 8 8 8 8 8 5
7 11 11 11 10 9 8 8 5
10 8 8 5
11 8 8 5
7 5 5 3


3
8
8
8
8
8
8
8
3
5
8
8
8
8
8
8
8
5
7
11
11
11
11
11
10
9
5
10
5
11
5
7
3


3
8
8
8
8
8
8
8
3
5
8
8
8
8
8
8
8
5
7
11
11
11
11
11
11
11
7



Next steps…

As in geometric multigrid,

- Select a coarse grid so that the smooth components can be represented accurately,
- Select an interpolation operator, so that the smooth components can be accurately transferred from the coarse grid to the fine grid,
- Define a restriction operator and a coarse grid version of A using the variational (Galerkin) properties.

Coarse grid operators

The restriction operator is given by the transpose of the prolongation, i.e.

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

The coarse grid operator is constructed using the Galerkin condition

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

The AMG algorithm

Let us now define the AMG algorithm. In case of AMG, we have

1. A setup step
2. The solution step using the components defined in the setup step

Coarse-fine AMG Setup Algorithm

Input:
A_0, the fine grid operator
Max_size: threshold for maximal size of coarsest problem

Output:
$A_1, \dots, A_L$ and $P_0, \dots, P_{L-1}$

$l = 0$
While size(A_l) > max_size

$S_l = \text{strength}(\mathrm{A}_l)$	(strength of connection)
$C_l, F_l = \text{splitting }(\mathrm{S}_l)$	(C/F-splitting)
$W = \text{weights}(S_l, \mathrm{A}_l, C_l, F_l)$	(Interpolation weigths)
$P_l = [W; I]$	(Form Interpolation)
$A_{l+1} = P_l^T A_l P_l$	(Coarse grid operator)
$l = l + 1$	

AMG Two-Grid Correction Cycle

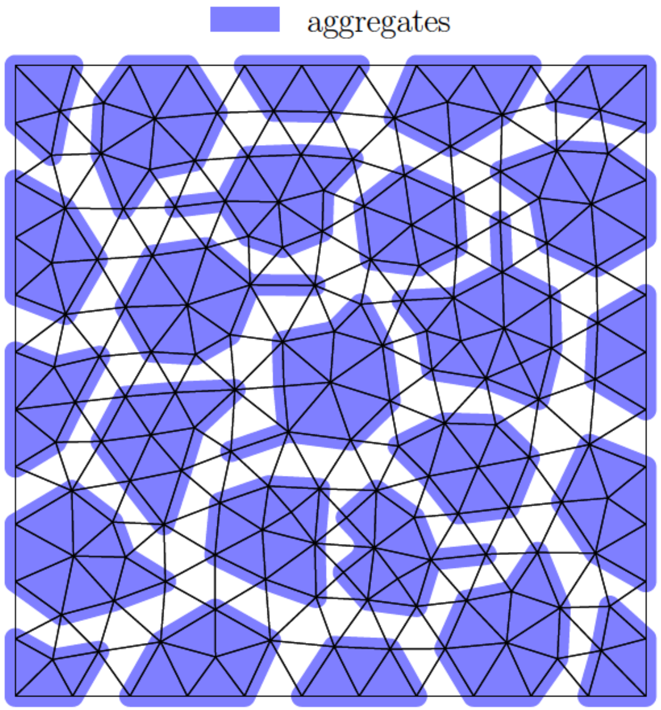
$\boldsymbol{v^h} \leftarrow \boldsymbol{AMG}(\boldsymbol{v^h}, \boldsymbol{f^h})$

- Relax ν_1 times on $A^h\boldsymbol{u}^h = \boldsymbol{f}^h$ with initial guess $\boldsymbol{v}^h$.
- Compute the fine-grid residual $\mathbf{r}^h = f^h - A^h\boldsymbol{v}^h$ and restrict it to the coarse grid by $\mathbf{r}^h = I_h^{2h}\boldsymbol{r}^h$
- Solve $A^{2h}\boldsymbol{e}^{\boldsymbol{2h}} = \boldsymbol{r}^{2h}$ on Ω^{2h}.
- Interpolate the coarse-grid error to the fine grid by $\boldsymbol{e}^h = I_{2h}^h\boldsymbol{e}^{2h}$ and correct the fine-grid approximation by $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h + \boldsymbol{e}^h$.
- Relax ν_2 times on $A^h\boldsymbol{u}^h = \boldsymbol{f}^h$ with initial guess $\boldsymbol{v}^h$.

Any other kind of cycles can then be build in strict analogy to geometric multigrid.

Different kind of coarsening strategies

Aggregation based coarsening



Coarse-fine AMG or Runge-Stueben

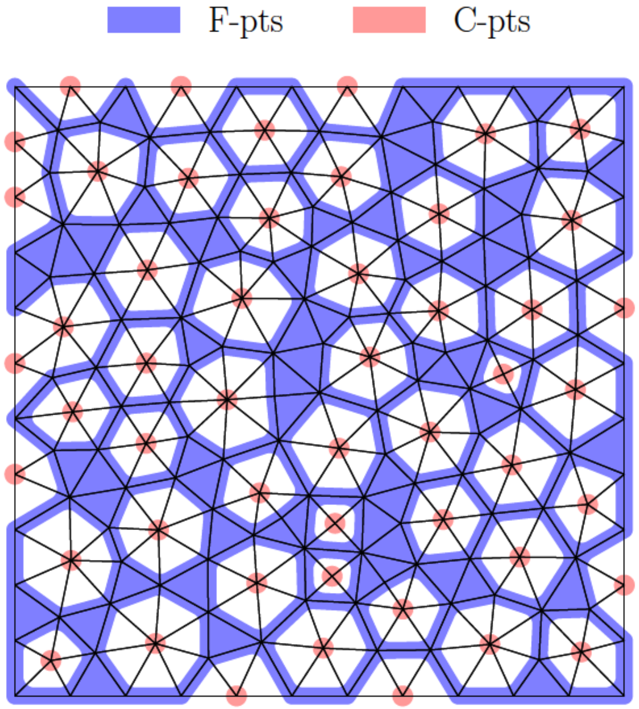


Figure taken of Copper Mountain AMG tutorial 2021

Some final comments

Blackbox solvers exist. Just try them out for a linear system that you want to solve!

- Library for python: **PyAMG** (https://github.com/pyamg/pyamg)

- **AGMG** by Yvan Notay (http://agmg.eu/)

- In PETSc:
  - **Hypre – BoomerAMG**
  - **GAMG**

And certainly other…

Additional references:

- Falgout, *An introduction to algebraic multigrid*, (2006)
- Stuben, *Algebraic Multigrid AMG An Introduction with Applications*, (1999)