Marrakech Jan 2003

MULTIGRID METHODS

Gérard MEURANT

CEA

• This is a part of the quest for an optimal method for elliptic PDE's: a number of operations proportional to n for a precision of the order of the truncation error

$$\texttt{Jacobi}: \lambda n^3$$

$$\texttt{Gauss-Seidel}: \beta n^3$$

$$\texttt{SOR} \text{ with optimal } \omega: \gamma n^2$$

$$\texttt{ICCG}: \gamma n^{1.5}$$

Some variants of the multigrid: $O(n)$

• MG was first studied in the Soviet Union in the 60's (Fedorenko (1962))

• the method received very little attention until A. Brandt popularized it in the 70's by solving many different and difficult problems. Since that time the method has received considerable attention and has emerged almost as a new branch of numerical mathematics

• Books: W. Hackbush and U. Trottenberg (1982), Hackbush (1985) and (1993, Briggs (1987), new edition (2001), Trottenberg (2001)

The two–grid method

2^{nd} order linear elliptic PDE:

$$Lu = f$$

in $\Omega =]0,1[^2$ with Dirichlet bc

• regular mesh with $h = 1/(m+1)$, m odd, 5 point finite difference scheme

$$A_h u_h = b_h$$

• coarse grid with $H = 2h$

1) Study why classical iterative methods like `Jacobi` or `Gauss-Seidel` have a poor convergence rate and try to improve this by using the coarse grid
2) Define a two–grid method, study why it does not work so well and correct this using for instance `Jacobi` or `Gauss-Seidel` "smoothing"

Answer to 1): they do not damp efficiently smooth components of the error

If we have an approximation u^k of u s.t. $Au = b$

let $\varepsilon^k = u - u^k$, then

$$A\varepsilon^k = Au - Au^k = b - Au^k = r^k$$

If we know an approximation v^k to ε^k, then

$$v^k + u^k$$

should be a better approximation

• The idea behind multigrid is to compute an approximation w^k of ε^k on a coarser grid Ω_H consisting of every other point in each direction

• So we need a way to go from Ω_h to Ω_H: a linear restriction operator R

$$R : \Omega_h \to \Omega_H, \quad (\Re^{m^2} \to \Re^{p^2})$$

$$P : \Omega_H \to \Omega_h, \quad (\Re^{p^2} \to \Re^{m^2})$$

• How to define A_H?

1) Use the same approximation as for A_h but on Ω_H (5 point)

2) $A_H = RA_hP$

With usual choices for R and P, this gives a nine point approximation scheme

• One step of the two–grid algorithm is defined by:

1) $r^k = b - Au^k$,

2) $r_H^k = Rr^k$,

3) Solve exactly $A_H\varepsilon_H^k = r_H^k$,

4) $v^k = P\varepsilon_H^k$,

5) $u^{k+1} = v^k + u^k$.

• The iteration matrix of this iterative method is given by

$$v^k = P\varepsilon_H^k = PA_H^{-1}r_H^k = PA_H^{-1}R(b - Au^k)$$

The iteration matrix is $K = I - PA_H^{-1}RA$

Can we have $\rho(K) < 1$?

This is the answer to question 2)!

Theorem

Let A be symmetric positive definite and suppose $A_H = RAP$ and $P = R^T$, then the eigenvalues of $I - PA_H^{-1}RA$ are 0 and 1

This is because some frequencies (eigenvectors) cannot be represented on the coarse grid

Before going to the coarse grid we need to "smooth" the residual

This can be done with relaxation iterations (which damp high frequencies quite fast)

1) do ν_1 iterations with an iteration matrix $S \rightarrow \bar{u}^k$

2) $\bar{r}^k = b - A\bar{u}^k$,

3) $\bar{r}_H^k = R\bar{r}^k$,

4) solve exactly $A_H \varepsilon_H^k = \bar{r}_H^k$,

5) $v^k = P\varepsilon_H^k$,

6) starting from $\bar{u}^k + v^k$, do ν_2 iterations with $S \rightarrow u^{k+1}$

The iteration matrix of this method is

$$M = S^{\nu_2}(I - PA_H^{-1}RA)S^{\nu_1} = S^{\nu_2}KS^{\nu_1}$$

M is similar to $S^{-\nu_2}MS^{\nu_2} = KS^{\nu_1+\nu_2}$ or $S^{\nu_1+\nu_2}K$

We have to choose

- the relaxation method S,
- the integers ν_1 and ν_2,
- how to construct the coarse grid,
- the restriction operator R,
- the prolongation operator P,
- how to define A_H

Let's study a 1D example:

$$-\frac{d^2u}{dx^2} = f \quad \text{in} \quad \Omega =]0,1[, \quad u(0) = u(1) = 0$$

Ω is divided into $n+1$ equal intervals, so $h = 1/(n+1)$, n odd and $n > 3$

The coarse grid Ω_H is defined with $H = 2h$

$$\frac{-u_{i-1} + 2u_i - u_{i+1}}{h^2} = f_i, \quad u_0 = u_{n+1} = 0$$

$$Au = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} u = b.$$

The eigenvalues of A are $\lambda_k = \frac{1}{h^2}(2 - 2\cos(k\pi h))$

The related eigenvector ϕ_k is such that $(\phi_k)_i = \sin(ik\pi h)$

We choose $\nu_1 + \nu_2 = 1$ and the relaxed `Jacobi` method as a smoother

$$A = \frac{2}{h^2}I + L + L^T = D + L + L^T$$

The iteration matrix is $S = I - \alpha D^{-1}A = I - \frac{\alpha h^2}{2}A$

The eigenvalues of S are $\mu_k = 1 - \alpha(1 - \cos(k\pi h))$, same eigenvectors as those of A

We must have $\alpha < \frac{2}{1+\cos(\pi h)}$ for convergence

$$\rho(S) = 1 - \alpha(1 - \cos(\pi h)) = 1 - O(h^2)$$

But, now we are only interested in smoothing properties!

◦ low frequencies: vectors φ_k with $k < \frac{n}{2}$

◦ high frequencies: vectors φ_k with $\frac{n}{2} \leq k \leq n$

we define something analogous to the spectral radius but restricted to high frequencies

$$\mu(h, \alpha) = \max_k \{|\lambda_k|, \frac{n}{2} \leq k \leq n\}$$

and

$$\mu^*(\alpha) = \sup_h \{\mu(h, \alpha), h \leq \frac{1}{4}\}$$

For our model example,

$$\mu(h, \alpha) = \max\{|1 - \alpha + \alpha \cos(\frac{n}{2}\pi h)|, |1 - \alpha - \alpha \cos(\pi h)|\}$$

$$\mu^*(\alpha) = \max\{|1 - \alpha|, |1 - 2\alpha|\}$$

The value of α which minimizes $\mu^*(\alpha)$ is $\alpha = \frac{2}{3}$ and $\mu^*(\frac{2}{3}) = \frac{1}{3}$

◦ Restriction: weighted average

$$w_H = Rw$$

$$(w_H)_{2i} = \frac{1}{4}(w_{2i-1} + 2w_{2i} + w_{2i+1})$$

◦ Prolongation: linear interpolation

$$w = Pw_H$$

$$(w)_{2i} = (w_H)_{2i},\ (w)_{2i+1} = \frac{1}{2}((w_H)_{2i} + (w_H)_{2i+2})$$

$$P = 2R^T$$

◦ Coarse grid matrix

$$(A_H v)_{2i} = \frac{-v_{2i-2} + 2v_{2i} - v_{2i+2}}{H^2}$$

Let Φ_k be the eigenvectors of A_H

$$(\Phi_k)_{2i} = \sin(ik\pi H) = \sin(2ik\pi h) = (\varphi_k)_{2i}$$

Then, we look for $K\varphi_k$

span$(\varphi_k, \varphi_{n+1-k})$ is invariant by K. If we order the basis vectors φ_1, φ_n, φ_2, φ_{n-1}, φ_3, $\ldots$, $\varphi_{\frac{n+1}{2}}$, K is block diagonal
The kth block is

$$\frac{1}{2}\begin{pmatrix} 1-c_k & 1+c_k \\ 1-c_k & 1+c_k \end{pmatrix}$$

with $c_k = \cos(k\pi h)$. The last block is equal to 1

• The eigenvalues of K are 0 and 1

$M = SK$ is also block diagonal

$$M_k = \frac{1}{2}\begin{pmatrix} (1-c_k)[1-\alpha(1-c_k)] & (1+c_k)[1-\alpha(1-c_k)] \\ (1-c_k)[1-\alpha(1+c_k)] & (1+c_k)[1-\alpha(1+c_k)] \end{pmatrix},$$

$$M_{\frac{n+1}{2}} = 1-\alpha$$

• The eigenvalues of M_k are 0 and $1-\alpha(1+c_k^2)$

Theorem

For our $1D$ example, the two–grid method converges if $\alpha < \frac{2}{1+c_1^2}$
The optimal value of α is $\alpha_{\text{opt}} = \frac{2}{2+c_1^2}$ and $\rho(M) = \frac{c_1^2}{2+c_1^2} \leq \frac{1}{3}$

We remark that as $h \to 0$, $\alpha_{\text{opt}} \to \frac{2}{3}$ and $\rho(M) \to \frac{1}{3}$ (optimal values for the smoother)

If one takes $\nu = 2$, then $\alpha_{\text{opt}} \approx \frac{2}{3}$ and $\rho(M) \approx \frac{1}{9}$

In $2D$

◦ Restriction:

$$(Ru)_{i,j} = \frac{1}{16}[4u_{i,j} + 2(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) + u_{i-1,j-1} + u_{i-1,j+1} + u_{i+1,j+1} + u_{i+1,j-1}]$$

◦ Prolongation: bilinear interpolation

For $2D$

α	1	$\frac{4}{5}$	$\frac{1}{2}$
$\mu(h,\alpha)$	c_1	$\frac{1+2c_1}{5}$	$\frac{2+c_1}{4}$
$\mu^*(\alpha)$	1	$\frac{3}{5}$	$\frac{3}{4}$

$\alpha = \frac{4}{5}$ is optimal

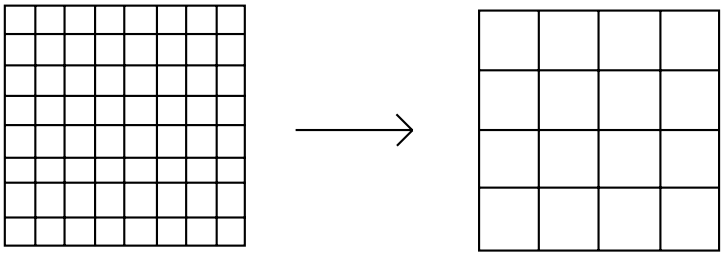
Setting $\rho^* = \sup_h \rho(M)$ we have

ν	μ^*	ρ^*
1	0.6	0.6
2	0.36	0.36
3	0.216	0.216
4	0.130	0.137
5	0.078	0.113

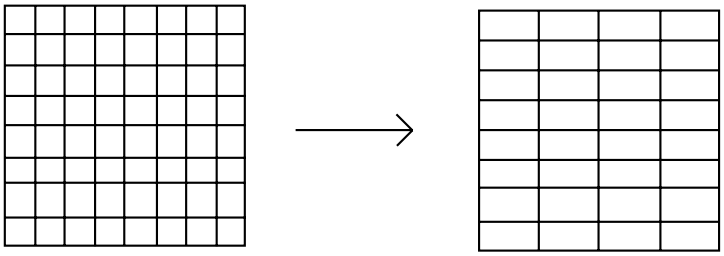
Smoothers

1) `Gauss-Seidel`
   - lexicographic ordering,
   - red–black ordering,
   - many color orderings, etc. . .
2) block relaxations
   - by lines,
   - by columns,
   - line zebra (odd–even ordering on the lines),
   - column zebra,
   - alternating directions (lines–columns or zebra lines–zebra columns),
3) `ILU` or block `ILU`,
4) Conjugate gradient

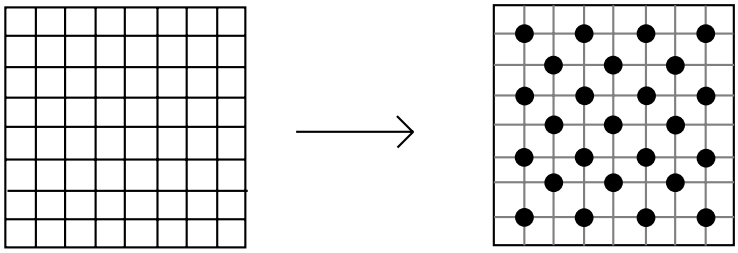
Coarsening



Standard coarsening



Semi coarsening



RB coarsening

Grid transfers

Prolongation:

1) bilinear interpolation. This is symbolically written as the stencil

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

2) a seven point interpolation

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

3) quadratic or cubic interpolation.

4) interpolation using the equations

Restriction:

1) $R = cP^T$, ($c = 1/4$ for finite differences)

   If $P=$ bilinear interpolation, $R=$ full weighting (FW)

2) half weighting (HW)

$$\frac{1}{8}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

3) Injection

   Generally the trivial restriction is not considered to be robust and is not used

   The prolongation and restriction must satisfy a "compatibility" condition

   If p and r are the orders of P and R and $2l$ is the order of the PDE, we must have $p + r \geq 2l - 1$

Coarse grid operator

1. using the same discretization as for the fine grid
2. using a "Galerkin" approximation: $A_H = RA_hP$

The multigrid method

• Solve the coarse problem $A_H \varepsilon_H^k = \bar{r}_H^k$ recursively

We have a sequence Ω_l of grids whose stepsizes are h_l, $l = 0$ corresponding to the coarsest grid and $l = L$ to the finest one

◦ A_l the approximation of A on Ω_l,

◦ R_l the restriction operator: $\Omega_l \rightarrow \Omega_{l-1}$,

◦ P_{l-1} the interpolation operator: $\Omega_{l-1} \rightarrow \Omega_l$,

◦ S_l the iteration matrix of the smoothing operator on Ω_l.

◦ let $\bar{w}_l = smooth^{\nu}(w_l, A_l, b_l)$ be the result of ν smoothing iterations for the problem $A_l u_l = b_l$ starting from w_l

Multigrid for $l+1$ levels

If $l = 1$ apply the two–grid algorithm

If $l > 1$

1) $\bar{u}_l^k = smooth^{\nu_1}(u_l^k, A_l, b_l)$,

2) $\bar{r}_l^k = b_l - A_l \bar{u}_l^k$,

3) $\bar{r}_{l-1}^k = R_l \bar{r}_l^k$,

4) compute $\bar{v}_{l-1}^k$ as the approximate solution of

$$A_{l-1} v_{l-1}^k = \bar{r}_{l-1}^k$$

on Ω_{l-1} by doing γ iterations of the l-grid algorithm $(\Omega_{l-1}, \ldots, \Omega_0)$ starting from 0

5) $\bar{v}_l^k = P_{l-1} \bar{v}_{l-1}^k$,

6) $u_l^{k+1} = smooth^{\nu_2}(\bar{u}_l^k + \bar{v}_l^k, A_l, b_l)$

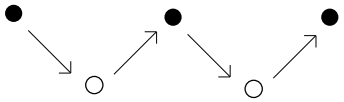
$\searrow$ the restriction,

$\swarrow$ the interpolation,

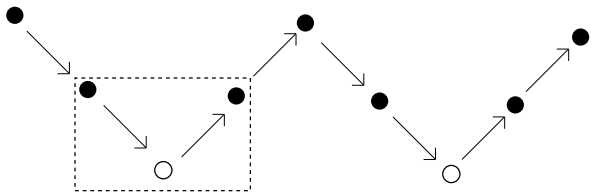
$\bullet$ the smoothing iterations,

$\circ$ the exact solution of a problem.

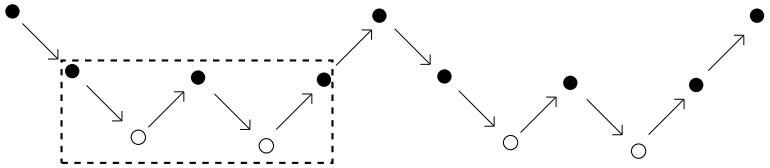
When $L = 1$ this is the two–grid algorithm which graphically looks like



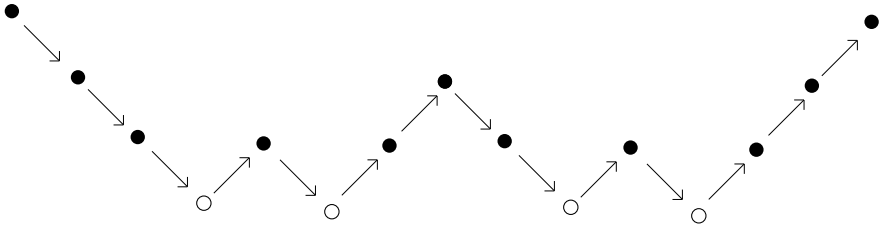
3 grids, $\gamma = 1$, V–cycle



3 grids, $\gamma = 2$, W–cycle



4 grids, $\gamma = 2$, W–cycle



• The iteration matrix is

$$M_l = S_l^{\nu_2}(I - P_{l-1}(I - M_{l-1}^{\gamma})A_{l-1}^{-1}R_l A_l)S_l^{\nu_1}$$

for $l > 2$ and

$$M_1 = S_1^{\nu_2}(I - P_0 A_0^{-1} R_1 A_1)S_1^{\nu_1}$$

This is also

$$M_l = M_l^{l-1} + (S_l^{\nu_2} P_{l-1})M_{l-1}^{\gamma}(A_{l-1}^{-1}R_l A_l S_l^{\nu_1})$$

There are many theoretical results for MG, see Hackbush

The Full multigrid method

• Starts from the coarsest level

let $M^r(w, A_l, b_l)$ be the operator corresponding to r iterations of the multigrid method for $A_l u = b_l$ on $\Omega_l, \ldots, \Omega_0$ starting from w

1) let $\widetilde{u}_0$ st $A_0 \widetilde{u}_0 = b_0$

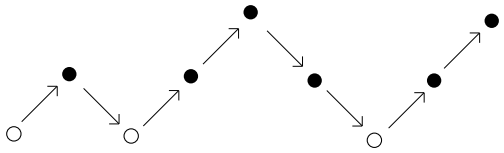
2) for $k = 1, \ldots, l$

$$u_k^0 = \Pi_{k-1} \widetilde{u}_{k-1},$$

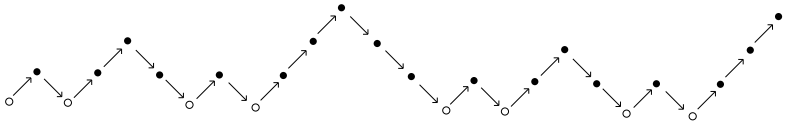
$$\widetilde{u}_k = M^r(u_k^0, A_k, b_k),$$

Π_{k-1} is an interpolation operator, $\Omega_{k-1} \to \Omega_k$

FMG, 3 grids, $\gamma = 1, r = 1$



FMG, 4 grids, $\gamma = 2, r = 1$



Algebraic multigrid

$$Ax = b$$

A symmetric positive definite of order n

The iterative method must be **scalable**

We will use PCG. The preconditioner must such that:

• the number of iterations is (almost) constant, when the problem size is increased (the condition number must be independent of h for PDEs or n)

• the complexity of applying the preconditioner is proportional to n

• easy to use on a parallel computer

Three known possibilities:

◦ multilevel (multigrid–like) methods

◦ domain decomposition

◦ combination of both

Multilevel preconditioners

◦ Algebraic methods (grid $\equiv$ (sub) set of unknowns)

◦ Algebraic MultiGrid (AMG)–like (V–cycle):

Starting from the zero vector:

0– if we are on the coarsest level, solve exactly by Gaussian elimination, otherwise

1– do ν iterations of smoothing

2– restrict the residual r to $r_c = Rr$

3– recursively solve $A_c e_c = r_c$, $A_c = RAP$, $R = P^T$

4– interpolate e_c to $e = Pe_c$

5– add the correction e to the current iterate

6– do ν iterations of smoothing

We have to define

◦ the smoother

$\rightarrow$ ◦ how to construct the coarse grids (coarsening alg.)

$\rightarrow$ ◦ the interpolation

Smoothers

◦ Symmetric Gauss–Seidel (not naturally //)

◦ Incomplete Cholesky (not // either) LDL^T

- IC(0)
- IC with some fill–in on values
- IC with some fill–in on levels of fill

$$LD^{-1}L^T(x^{k+1} - x^k) = b - Ax^k$$

◦ Least squares polynomial

$$\int_a^b (1 - \lambda q(\lambda))^2 w(\lambda)\, d\lambda, \quad q \in \mathcal{Q}_k,$$

a and b Gerschgorin bounds for the eigenvalues

It is usually more efficient to use a low order (1 or 2) polynomial

◦ Approximate inverse AINV from M. Benzi (Emory Univ.)

$$M = ZD^{-1}Z^T$$

where the matrix Z is upper triangular with $1's$ on the diagonal and D is diagonal

Approximate A–orthogonalization:

$Z = I$ $d_1 = a_{1,1}$

for $i = 2, \ldots, n$

drop the entries $z_{k,i}$ such that $|z_{k,i}| \leq \tau \|a_i\|_\infty$

for $j = i, \ldots, n$

$d_j = a_i^T z_j$

end

for $j = i + 1, \ldots, n$

$z_j = z_j - \frac{d_j}{d_i} z_i$

end

end

The parameter τ defines which entries of Z are kept

Works fine for H–matrices, for general SPD matrices use SAINV (Stabilized AINV)

Smoother: Richardson iteration defined as (matrix $\times$)

$$x^{k+1} = x^k + M(b - Ax^k)$$

For a two–grid algorithm using one step of pre–smoothing and one step of post–smoothing starting from $x^0 = 0$, the preconditioner $\tilde{M}_1$ is defined as

$$\tilde{M}_1 = M + M(I - AM) + (I - MA)(P(RAP)^{-1}R)(I - AM)$$

Theorem

$\tilde{M}_1$ is symmetric

We can show that $\tilde{M}_1$ is positive definite if we suppose that M is such that $\rho(I - AM) < 1$ (sufficient condition)

Moreover 1 is a multiple eigenvalue of $\tilde{M}_1 A$ and all the eigenvalues of $\tilde{M}_1 A$ are smaller or equal to 1

This occurs whatever the choice of M, R and P

The same results apply if the coarse matrix is obtained by using the same algorithm recursively, that is in the multilevel case

Other smoothers

◦ A crude approximate inverse (Tang and Wan)

The approximate inverse M for a general matrix A is computed to minimize

$$\|I - MA\|_F$$

This problem is equivalent to solving n l_2 minimization problems,

$$\|A^T m_i - e_i\|$$

where m_i^T is the ith row of M and e_i is the ith column of the identity matrix

There are clever algorithms to adaptively chose the sparsity pattern (Huckle and Grote)

Tang and Wan suggested to use the $(0, 1)$ level pattern

M might not be symmetric, use $1/2(M+M^T)$. Positive definiteness?

The influence matrix

$$\mathcal{N} = \{1, \ldots, n\}$$

$$\mathcal{N} = F \cup C$$

Influence matrix: standard AMG choice for M–matrices

$$S_i = \{j | -a_{i,j} > \theta \max_{k \neq i}(-a_{i,k}), \quad \theta < 1\}$$

This gives S (by padding with zeros)

Generalization:

$$S_i^A = \{j \mid |a_{i,j}| > \tau \max_{k \neq i} |a_{i,k}|, \quad \tau < 1\}$$

This choice is denoted by 'a'

Keep at least a 1 for the largest element in magnitude ('b')

The coarsening algorithm

The "standard" algorithm is:

Weights w_i = the number of points that depend on i (using S)

1- choose the first point i with maximal weight as a C point

2- assign the points that i influences (using S) as F points

3- increment by 1 the weights of the points influencing these new F points (to give more chances to be selected as C points)

4- decrease by 1 the weights of points that depends on i

repeat from step 1- until all points are labeled

◦ There are many other choices for the coarsening algorithm

◦ Remark that if we change the interpolation algorithm the number and location of coarse nodes change (on coarse levels)

The interpolation algorithm

◦ AMG uses $Ae = 0$, $i \in F, j \in C$

$$\omega_{i,j} = -\frac{a_{i,j} + \sum_{k \in D_i^S} \frac{a_{i,k} a_{k,j}}{\sum_{m \epsilon C_i} a_{k,m}}}{a_{i,i} + \sum_{k \in D_i^W} a_{i,k}}$$

This uses $e_j \approx e_i$ for weak connections and a weighted average for F connections

Note that the given F point needs to have at least one coarse point in its neighborhood in the graph of A

The coarse matrices

The interpolation algorithm defines P and $R = P^T$

$$A_C = RAP$$

Numerical experiments

5 point finite differences, unit square, $m \times m$ mesh

b random

$x^0 = 0$

stopping criterion $\|r^k\| \leq 10^{-10}\|r^0\|$

Done on my PC (Sony XG9, Intel 500 Mhz) using Matlab

◦ Poisson equation

◦ Discontinuous problem

diffusion coeff=1 except 100 in the strip $[0, 1] \times [1/4, 3/4]$

scaled matrix with 1s on the diagonal

Matrices from the Matrix Market

◦ 1138-bus admittance matrix

◦ nos 7

with diagonal scaling