AMSC 600/CMSC 760
Fall 2007
Solution of Sparse Linear Systems
Multigrid, Part 2
Dianne P. O'Leary

Multigrid, part 2

There are three important topics left to consider:

- Programs for multigrid (SCSCwebpage/cs_multigrid)
- Convergence of multigrid. (Saad Section 13.5)
- Algebraic multigrid. (Saad Section 13.6)

Programs for multigrid (SCSCwebpage/cs_multigrid)

Convergence of multigrid

We'll follow Saad in talking about the analysis of a 2-level V-cycle.

Solving the problem exactly on the $2h$ grid is usually too expensive, so that is not a practical algorithm.

But the general V-cycle is a perturbed version of this, since the coarse mesh problem is then solved inexactly rather than exactly.

Notation

- We let $\mathbf{R}_h$ be the operator takes values on grid $h/2$ and produces values on grid h. (This moves from fine to coarse grid.) We let $\mathbf{P}_h$ be the operator takes values on grid h and produces values on grid $h/2$. (This moves from coarse to fine grid.)
- We choose $\mathbf{R}_h = \mathbf{P}_h^T$.
- $\mathbf{A}_{2h} = \mathbf{R}_{2h}\mathbf{A}_h\mathbf{P}_{2h}$

Since we only work with 2 meshes, we'll drop the subscripts on $\mathbf{R}$ and $\mathbf{P}$.

V-Cycle: 2 levels

$\mathbf{v}_h$ = V-Cycle($\mathbf{v}_h$, $\mathbf{A}_h$, $\mathbf{f}_h$, η_1,η_2)

Perform η_1 G-S iterations on $\mathbf{A}_h\mathbf{u}_h = \mathbf{f}_h$ using $\mathbf{v}_h$ as the initial guess, obtaining an approximate solution that we still call $\mathbf{v}_h$.

Compute $\mathbf{v}_{2h}$ to solve $\mathbf{A}_{2h}\mathbf{v}_{2h} = \mathbf{R}(\mathbf{f}_h - \mathbf{A}_h\mathbf{v}_h)$.

Set $\mathbf{v}_h = \mathbf{v}_h + \mathbf{P}\mathbf{v}_{2h}$.

Perform η_2 G-S iterations on $\mathbf{A}_h\mathbf{u}_h = \mathbf{f}_h$ using $\mathbf{v}_h$ as the initial guess, obtaining an approximate solution that we still call $\mathbf{v}_h$.

Analysis

- We want to express this as a stationary iterative method:

$$\mathbf{v}_h^{(new)} = \mathbf{M}\mathbf{v}_h + \mathbf{c}.$$

- The algorithm:
  - Apply η_1 G-S iterations. This iteration matrix is $\mathbf{M}_{GS}^{\eta_1}$. This operator is called the smoother.
  - Solve the coarse grid problem and add the correction.
    * We get the residual by taking $-\mathbf{A}_h$ times the current guess.
    * Then we restrict to the coarse grid with $\mathbf{R}$ and apply $\mathbf{A}_{2h}^{-1}$.
    * Then we prolong using $\mathbf{P}$.
    * Then we add this onto the iterate.
  - Apply η_2 G-S iterations. This iteration matrix is $\mathbf{M}_{GS}^{\eta_2}$.
  - So the matrix $\mathbf{M}$ is

$$\mathbf{M} = \mathbf{M}_{GS}^{\eta_2}(\mathbf{I} - \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h)\mathbf{M}_{GS}^{\eta_1} \equiv \mathbf{M}_{GS}^{\eta_2}\mathbf{T}\mathbf{M}_{GS}^{\eta_1}.$$

- The following analysis (without loss of generality) takes $\eta_1 = 0$.

Assumptions

(Saad, Section 13.5) We make two (standard) assumptions, one concerning the smoother and one concerning the grids.

- Smoothing property:

$$\|\mathbf{M}_{GS}^{\eta_2}\mathbf{z}\|_A^2 \le \|\mathbf{z}\|_A^2 - \alpha\|\mathbf{A}_h\mathbf{z}\|_{D^{-1}}^2,$$

  for all $\mathbf{z}$ and for some constant α independent of h.
  ($A = \mathbf{A}_h$, $D =$diag$(\mathbf{A}_h)$.)

- Approximability:

$$\min_{\mathbf{w}} \|\mathbf{z} - \mathbf{P}\mathbf{w}\|_D^2 \leq \beta \|\mathbf{z}\|_A^2,$$

where the minimum is taken over all vectors $\mathbf{w}$ on the coarse grid, and where the constant β is independent of h.

Properties

- P1: $\mathbf{I} - \mathbf{T} = \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h$ is a projection operator, meaning that $(\mathbf{I} - \mathbf{T})^2 = \mathbf{I} - \mathbf{T}$ and $\mathbf{T}^2 = \mathbf{T}$.

  Proof:

$$\begin{aligned}
(\mathbf{I} - \mathbf{T})^2 &= (\mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h)(\mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h) \\
&= \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{A}_{2h}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h \\
&= \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h \\
&= \mathbf{I} - \mathbf{T}.
\end{aligned}$$

  The proof that $\mathbf{T}^2 = \mathbf{T}$ follows easily.

- P2:

$$\begin{aligned}
\mathbf{T}^T\mathbf{A}_h &= \mathbf{A}_h - \mathbf{A}_h\mathbf{R}^T\mathbf{A}_{2h}^{-1}\mathbf{P}^T\mathbf{A}_h \\
&= \mathbf{A}_h(\mathbf{I} - \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h) \\
&= \mathbf{A}_h\mathbf{T}.
\end{aligned}$$

- P3:

$$\begin{aligned}
\mathbf{T}\mathbf{P} &= (\mathbf{I} - \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h)\mathbf{P} \\
&= \mathbf{P} - \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{R}\mathbf{A}_h\mathbf{P} \\
&= \mathbf{P} - \mathbf{P}\mathbf{A}_{2h}^{-1}\mathbf{A}_{2h} \\
&= \mathbf{P} - \mathbf{P} = \mathbf{0}.
\end{aligned}$$

- P4: $\|\mathbf{T}\mathbf{z}\|_A^2 \leq \|\mathbf{z}\|_A^2$.

  Proof:

$$\begin{aligned}
\|\mathbf{z}\|_A^2 &= \|(\mathbf{T} + (\mathbf{I} - \mathbf{T}))\mathbf{z}\|_A^2 \\
&= \mathbf{z}^T\mathbf{T}^T\mathbf{A}_h\mathbf{T}\mathbf{z} - 2\mathbf{z}^T\mathbf{T}^T\mathbf{A}_h(\mathbf{I} - \mathbf{T})\mathbf{z} + \mathbf{z}^T(\mathbf{I} - \mathbf{T})^T\mathbf{A}_h(\mathbf{I} - \mathbf{T})\mathbf{z} \\
&= \|\mathbf{T}\mathbf{z}\|_A^2 + \|(\mathbf{I} - \mathbf{T})\mathbf{z}\|_A^2,
\end{aligned}$$

  and the result follows. (We have used P1 and P2 to get rid of the plum colored term in the middle line.)

- CS: The Cauchy-Schwartz inequality says

$$|\mathbf{z}^T\mathbf{y}| \leq \|\mathbf{z}\|\|\mathbf{y}\|.$$

- P5: For all $\mathbf{y}$, if $\mathbf{z} = \mathbf{T}\mathbf{w}$ (i.e., $\mathbf{z}$ is in the range of $\mathbf{T}$), then by P3,

$$\mathbf{z}^T \mathbf{A}_h \mathbf{P}\mathbf{y} = \mathbf{w}^T \mathbf{T}\mathbf{A}_h \mathbf{P}\mathbf{y} = \mathbf{w}^T \mathbf{A}_h \mathbf{T}\mathbf{P}\mathbf{y} = \mathbf{0}.$$

- P6: Therefore, if $\mathbf{z}$ is in the range of $\mathbf{T}$, then $\|\mathbf{z}\|_A \le \sqrt{\beta}\|\mathbf{A}_h\mathbf{z}\|_{D^{-1}}$.

  Proof: We let $\mathbf{y}$ be the minimizer in the definition of approximability. Then

$$\begin{aligned}
\|\mathbf{z}\|_A^2 &= \mathbf{z}^T \mathbf{A}_h \mathbf{z} \\
&= \mathbf{z}^T \mathbf{A}_h \mathbf{D}^{-1/2}\mathbf{D}^{1/2}(\mathbf{z} - \mathbf{P}\mathbf{y}) && \textit{(by P5)} \\
&\le \|\mathbf{D}^{-1/2}\mathbf{A}_h\mathbf{z}\|\;\|\mathbf{D}^{1/2}(\mathbf{z} - \mathbf{P}\mathbf{y})\| && \textit{(by CS)} \\
&= \|\mathbf{A}_h\mathbf{z}\|_{D^{-1}}\|(\mathbf{z} - \mathbf{P}\mathbf{y})\|_D \\
&\le \sqrt{\beta}\|\mathbf{A}_h\mathbf{z}\|_{D^{-1}}\|\mathbf{z}\|_A\,. && \textit{(by Approximability)}
\end{aligned}$$

- Finally,

$$\begin{aligned}
\|\mathbf{M}_{GS}^{\eta_2}\mathbf{T}\mathbf{z}\|_A^2 &\le \|\mathbf{T}\mathbf{z}\|_A^2 - \alpha\|\mathbf{A}_h\mathbf{T}\mathbf{z}\|_{D^{-1}}^2 && \textit{(by Smoothing)} \\
&\le \|\mathbf{T}\mathbf{z}\|_A^2 - \frac{\alpha}{\beta}\|\mathbf{T}\mathbf{z}\|_A^2 && \textit{(by P6)} \\
&= \left(1 - \frac{\alpha}{\beta}\right)\|\mathbf{T}\mathbf{z}\|_A^2 \\
&\le \left(1 - \frac{\alpha}{\beta}\right)\|\mathbf{z}\|_A^2. && \textit{(by P4)}
\end{aligned}$$

- If we let $\mathbf{z}$ denote the error in the solution before applying a multigrid iteration, the inequality above guarantees reduction in the error (when measured in the $\mathbf{A}_h$-norm) at a rate independent of h, so the number of iterations necessary to reduce the error by a given amount is constant, independent of h, so it can be done using an amount of work proportional to a work unit.

Algebraic multigrid

We introduced multigrid by presenting grids.

There have been many attempts to extend the ideas to gridless problems.

Let's think of h as a parameter related to the size of the matrix problem, so that $\mathbf{A}_h$ is (at most) twice as big as $\mathbf{A}_{2h}$.

The definition of a multigrid method depends on only two things:

- We choose $\mathbf{R}_h = \mathbf{P}_h^T$.
- $\mathbf{A}_{2h} = \mathbf{R}_{2h}\mathbf{A}_h\mathbf{P}_{2h}$ and this matrix is (at most) half the size of $\mathbf{A}_h$.

This is easy to accomplish in general.

The convergence analysis depends on two additional assumptions:

- Smoothing property:

$$\|\mathbf{M}_{GS}^{\eta_2}\mathbf{z}\|_A^2 \le \|\mathbf{z}\|_A^2 - \alpha\|\mathbf{A}_h\mathbf{z}\|_{D^{-1}}^2,$$

  for all $\mathbf{z}$ and for some constant α independent of h.

- Approximability:

$$\min_{\mathbf{w}} \|\mathbf{z} - \mathbf{P}\mathbf{w}\|_D^2 \le \beta\|\mathbf{z}\|_A^2,$$

This is not easy, and we will abandon hope of it.

One approach to Algebraic Multigrid: Graph coloring

- Use the graph of the matrix $\mathbf{A}$ to define the geometry of the problem.
- Color the nodes of the graph to define the coarse grid.
- Define the restriction operator $\mathbf{R}$ to take a weighted average of values on the fine grid to create a value on the coarse grid.

See Saad's "guiding principles" on p. 442.

A second approach to Algebraic Multigrid: ILU

Suppose, by coloring the graph or by other means, we can obtain a permutation matrix $\mathbf{Q}$ so that

$$\mathbf{QAQ}^T = \begin{bmatrix} \mathbf{B} & \mathbf{F} \\ \mathbf{E} & \mathbf{C} \end{bmatrix}$$

where $\mathbf{B}$ is block diagonal.

Then our matrix problem can be written

$$\begin{bmatrix} \mathbf{B} & \mathbf{F} \\ \mathbf{E} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}.$$

Note that

$$\begin{bmatrix} \mathbf{B} & \mathbf{F} \\ \mathbf{E} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{EB}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{F} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}$$

where the Schur complement is

$$\mathbf{S} = \mathbf{C} - \mathbf{EB}^{-1}\mathbf{F}.$$

So we have reduced our problem to:

- Let $\mathbf{w}_1 = \mathbf{f}_1$ and solve $\mathbf{w}_2 = \mathbf{f}_2 - \mathbf{E}\mathbf{B}^{-1}\mathbf{w}_1$.
- Solve $\mathbf{S}\mathbf{u}_2 = \mathbf{w}_2$.
- Solve $\mathbf{B}\mathbf{u}_1 = \mathbf{w}_1 - \mathbf{F}\mathbf{u}_2$.

and we only need to be able to solve linear systems involving $\mathbf{B}$ and $\mathbf{S}$ easily.

The problem involving $\mathbf{S}$ can be thought of as the coarse grid problem.

Repeating this recursively (next, looking for a similar partition of $\mathbf{S}$) gives a multilevel algorithm.

When the Schur complement gets too dense, we can substitute an approximation to it and then apply a V-cycle.

Final words

- We used GS as a smoother. Other SIMs (e.g., Jacobi) can also be used.
- (Geometric) multigrid is a terrific method for pde's and certain other problems.
- Algebraic multigrid needs a lot more work.
- Saad's Section 13.7 is titled, "Multigrid vs. Krylov Methods." A better approach is to think of multigrid as a preconditioner.
  - If multigrid works well, then only 1 iteration of the Krylov method is used, and the algorithm is just multigrid.
  - If multigrid is slow, the Krylov method will accelerate it.