L18

Gauss-Seidel Method & Convergence

GAUSS-SEIDEL

Let us reconsider the Jacobi Method:

$$a_{ii} x_i^{(k+1)} = -\sum_{\substack{j=1 \\ j\neq n}}^{n} a_{ij} x_j^{(k)} + b_i \qquad (1)$$

Let's write this in the following way:

$k+1$	k
x_1	$x_2 \; x_3 \; x_4 \; \cdots \; x_{n-1} \; x_n$
x_2	$x_1 \; x_3 \; x_4 \; \cdots \; x_{n-1} \; x_n$
x_3	$x_1 \; x_2 \; x_4 \; \cdots \; x_{n-1} \; x_n$
$\vdots$	
x_n	$x_1 \; x_2 \; x_3 \; x_4 \; \cdots \; x_{n-1}$

Note that the computation of $x_2^{(k+1)}$ makes use of $x_1^{(k)}$. But by that pt in the calculation, we have already computed $x_1^{(k+1)}$, which is expected to be closer to x_1 than $x_1^{(k)}$. We therefore expect that if we replace $x_1^{(k)}$ with $x_1^{(k+1)}$ in the calculation of $x_2^{(k+1)}$ we will get a better approx. of x_2.

A similar argument holds for

making the replacements:

$$x_1^{(k)} \longrightarrow x_1^{(k+1)}$$

$$x_2^{(k)} \longrightarrow x_2^{(k+1)}$$

in the computation of $x_3^{(k+1)}$.

We can visualize this as follows:

$k+1$	k						
x_1		x_2	x_3	x_4	$\cdots$	x_{n-1}	x_n
x_2	x_1		x_3	x_4	$\cdots$	x_{n-1}	x_n
x_3	x_1	x_2		x_4	$\cdots$	x_{n-1}	x_n
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_n	x_1	x_2	x_3	x_4	$\cdots$	x_{n-1}	

The corresponding modification to (1) is:

$$\boxed{a_{ii} x_i^{(k+1)} = -\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} + b_i} \qquad (*)$$

This algorithm is known as the Gauss-Seidel method, and is often faster than the Jacobi Method.

The Gauss-Seidel Method also fits into the general paradigm of:

$$A = S - T$$

$$S x_{k+1} = T x_k + b$$

To see this, move all $(k+1)$ vector components to the LHS:

$$\sum_{j=1}^{i} a_{ij} x_j^{(k+1)} = -\sum_{j=i+1}^{n} a_{ij} x_j^{(k)} + b_i .$$

or $$\underbrace{(L+D)}_{S} x_{k+1} = \underbrace{-U}_{T} x_k + b \qquad (**)$$

(Notice that I've switched the position of the iteration index again.)

Here: L = lower triangular part of A
D = diagonal
U = upper triangular

We recognize $(**)$ as a lower triangular system for the variable x_{k+1}, which can be solved by forward substitution (compare $(**)$ w/ Ex 1 of Lab05).

CONDITION FOR CONVERGENCE

Consider an iterative method

$$x_{k+1} = Bx_k + c \qquad (2)$$

Here $B = S^{-1}T$, $c = S^{-1}b$, using our earlier notation. Let x be the solution of:

$$x = Bx + c. \qquad (3)$$

Then $(3) - (2) \Rightarrow$

$$e_{k+1} = Be_k \qquad (4)$$

Where

$$e_k = x - x_k$$
$$= \text{"error" @ } k^{th} \text{ iteration}$$

We would like to find necessary and sufficient conditions for $e_k \to 0 \iff x_k \to x$.

Eq (4) says that @ every iteration step, the error is multiplied by B. Apply (4) repeatedly to get:

$$e_k = B^k e_0 \qquad (5)$$

If B were a scalar, then you'd immediately say that $e_k \to 0$ if $|B| < 1$. But B is a matrix. Is there a scalar property of a matrix that plays a similar role to absolute value of a scalar? If so, what is it? Norm? Spectral radius?

To find out, suppose

$$B v_i = \lambda_i v_i$$

where v_i are orthonormal and span the space.

Then

$$e_0 = \sum_{i=1}^{n} c_i v_i$$

$$\Rightarrow \quad e_k = \sum_{i=1}^{n} c_i \lambda_i^k v_i \quad \ldots \text{ by } (5)$$

Evector v_i decays if $|\lambda_i| < 1$. If this is true for all i, then $e_k \to 0$, as we want. Thus, a sufficient condition for convergence of

$$x_{k+1} = B x_k + c$$

is that

$$\underbrace{\rho(B)}_{\text{spectral radius of } B!} < 1.$$

Now consider the reverse direction: if $e_k \to 0$ then is it true that $\rho(B) < 1$?

Well: $e_k \to 0$

$\Rightarrow \quad B^k e_0 \underset{(5)}{\to} 0 \quad$ for any e_0

$\Rightarrow \quad \rho(B) < 1 \quad$ by Thm 7.17 p. 449

$$\therefore \boxed{\begin{array}{l}\rho(B) < 1 \iff \\ (x_{k+1} = Bx_k + c) \to (x = Bx + c).\end{array}}$$

RATE OF CONVERGENCE

What can we say about $\|e_k\|$ for large k, eg $k \gg 1$?

$$(5) \Rightarrow \quad e_k = \sum_i c_i \lambda_i^k v_i$$

Rearrange indices s.t. λ_1 is largest in absolute value:

$$e_k = \lambda_1^k \sum_i c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k v_i$$

$$\xrightarrow[k \gg 1]{} \lambda_1^k c_1 v_1$$

$$\Rightarrow \quad \|e_k\| \xrightarrow[k \gg 1]{} \underbrace{|\lambda_1|^k}_{\rho(B)^k} |c_1| \underbrace{\|v_1\|}_{1}$$

$$\propto \rho(B)^k$$

Not only must $\rho(B)$ be less than 1 for convergence, the magnitude of $\rho(B)$ controls the rate of convergence: iterative methods w/ smaller spectral radii converge faster!