ITERATION METHODS

Direct methods are good for dense systems of small to moderate size. This generally means the entire coefficient matrix for the linear system can be stored in the computer's memory. Then Gaussian elimination can be applied to solve the linear system.

In contrast, iteration methods are used with large sparse systems and large dense systems. Discretizations of elliptic boundary value problems for partial differential equations often lead to large sparse systems. Large dense systems arise from solving boundary integral equations, the radiosity equation, and from some optimization problems.

With large sparse systems arising from solving PDEs, there is generally a pattern to the sparsity, and we can take advantage of this.

ITERATION METHODS - REMARKS

The numerical solution of partial differential equations (PDEs) leads to large sparse systems of linear and nonlinear algebraic equations. These must often be solved by iteration, although there are direct methods for many such problems.

- Traditional iteration - Jacobi, Gauss-Seidel, SOR, red/black iteration, line iteration

- Multigrid iteration - Uses many levels of discretization

- Conjugate gradient iteration and variants of it.

We look only at discretizing the Poisson equation $\Delta u = g$. The same discretization procedure can be applied to most other PDEs. *Finite element methods* also lead to large sparse banded systems which are often solved by iteration.

POISSON'S EQUATION

The Dirichlet boundary value problem for Poisson's equation is given by

$$\begin{aligned} \Delta u(x,y) &= g(x,y), \quad (x,y) \in R \\ u(x,y) &= f(x,y), \quad (x,y) \in \Gamma \end{aligned} \tag{1}$$

where R is a planar region and Γ is the boundary of R. In the book in §8.8, we examine this problem for $R = \{(x,y) \mid 0 < x, y < 1\}$.

For an integer $N > 1$, we introduce a mesh size $h = 1/N$. With it, we define a rectangular mesh on $\overline{R} = R \cup \Gamma$:

$$(x_j, y_k) = \left(\tfrac{j}{N}, \tfrac{k}{N}\right), \qquad j,k = 0,1,...,N$$

At mesh points (=grid points) inside of R, we approximate the equation $\Delta u = g$. To do so, we recall the approximation

$$G''(x) = \frac{G(x+h) - 2G(x) + G(x-h)}{h^2} - \frac{h^2}{12} G^{(4)}(\xi)$$

with some $\xi \in [x-h, x+h]$. This assumes G is four times continuously differentiable on $[x-h, x+h]$.

Since

$$\Delta u(x,y) = \frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2}$$

we obtain at each $(x_j, y_k) \in R$ that

$$\begin{aligned} &\frac{u(x_{j+1},y_k) - 2u(x_j,y_k) + u(x_{j-1},y_k)}{h^2} \\ &\quad + \frac{u(x_j,y_{k+1}) - 2u(x_j,y_k) + u(x_j,y_{k-1})}{h^2} \\ &\quad - \frac{h^2}{12}\left[\frac{\partial^4 u(\xi_j,y_k)}{\partial x^4} + \frac{\partial^4 u(x_j,\eta_k)}{\partial y^4}\right] \\ &\quad = g(x_j,y_k) \end{aligned} \tag{2}$$

with $\xi_j \in \Big[x_{j-1}, x_{j+1}\Big]$, $\eta_k \in [y_{k-1}, y_{k+1}]$. We can delete the error terms involving the fourth derivatives of u. This leads to a new set of equations for approximating unknown $u_h \approx u$:

$$\Big\{u_h(x_j,y_k) \mid 1 < j,k < N\Big\}$$

The values of u_h at boundary mesh points on Γ are given by

$$u_h(x_j,y_k) = f(x_j,y_k), \qquad (x_j,y_k) \in \Gamma \tag{3}$$

This leads to the linear system

$$\frac{u_h(x_{j+1}, y_k) - 2u_h(x_j, y_k) + u_h(x_{j-1}, y_k)}{h^2}$$
$$+\frac{u_h(x_j, y_{k+1}) - 2u_h(x_j, y_k) + u_h(x_j, y_{k-1})}{h^2}$$
$$= g(x_j, y_k), \qquad (x_j, y_k) \in R$$

$$u_h(x_j, y_k) = f(x_j, y_k), \qquad (x_j, y_k) \in \Gamma \tag{4}$$

For interior mesh points, we can simplify this to

$$\begin{aligned} u_h(x_j, y_{k-1}) + u_h(x_{j-1}, y_k) - 4u_h(x_j, y_k) \\ +u_h(x_{j+1}, y_k) + u_h(x_j, y_{k+1}) \\ = h^2 g(x_j, y_k), \quad (x_j, y_k) \in R \end{aligned} \tag{5}$$

Thus we have a system of $(N+1)^2$ linear equations in the $(N+1)^2$ unknowns

$$\Big\{u_h(x_j, y_k) \mid 1 \le j, k \le N\Big\}$$

Is it solvable?

SOLVABILITY

Consider the homogeneous linear system:

$$\begin{aligned} v_h(x_j, y_{k-1}) + v_h(x_{j-1}, y_k) - 4v_h(x_j, y_k) \\ +v_h(x_{j+1}, y_k) + v_h(x_j, y_{k+1}) \\ = 0, \quad (x_j, y_k) \in R \end{aligned} \tag{6}$$

$$v_h(x_j, y_k) = 0, \qquad (x_j, y_k) \in \Gamma \tag{7}$$

At the interior node points $(x_j, y_k) \in R$, this leads to

$$\begin{aligned} v_h(x_j, y_k) \\ = \frac{1}{4} \Big\{ v_h(x_j, y_{k-1}) + v_h(x_{j-1}, y_k) \\ +v_h(x_{j+1}, y_k) + v_h(x_j, y_{k+1}) \Big\} \end{aligned}$$

At boundary points $(x_j, y_k) \in \Gamma$, $v_h(x_j, y_k) = 0$.

Introduce

$$v_h(\overline{x}_j, \overline{y}_k) = \max_R v_h(x_j, y_k) \equiv c$$

Now suppose $c > 0$. Using (6) at the point $(\overline{x}_j, \overline{y}_k)$ and the fact that it a maximal point for v_h, we can conclude that at all four surrounding points, to the top and bottom, to the left and right, v_h will also equal c. Continue this process with those four points. Eventually, we will obtain that v_h will equal zero at a boundary point, a contradiction. Thus $c = 0$. The same argument can be used to show

$$\min_R v_h(x_j, y_k) = 0$$

Thus $v_h(x_j, y_k) \equiv 0$ on R, and the homogeneous linear system has only the zero solution. Thus the original linear system (4)-(5) has a unique solution for every given right hand side.

If $u(x, y)$ is four times continuously differentiable over $\overline{R}$, it can be shown that for some c,

$$\max_R \left| u(x_j, y_k) - u_h(x_j, y_k) \right| \leq c\, h^2$$

GAUSS-JACOBI ITERATION

To solve $Ax = b$, begin by rewriting it in the form

$$x_i = \frac{1}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{i,j} x_j \right], \qquad i = 1, ..., n \tag{8}$$

assuming all $a_{i,i} \neq 0$. Define the iteration by

$$x_i^{(m+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{i,j} x_j^{(m)} \right], \qquad i = 1, ..., n \tag{9}$$

for $m = 0, 1, ...$, with some given $x^{(0)}$.

Sometimes it is written slightly differently. For the system $(I - B)x = b$, we might instead use the form $x = b + Bx$, and then define

$$x^{(m+1)} = b + Bx^{(m)}, \qquad m = 0, 1, ...$$

We will analyze only the first form.

GAUSS-SEIDEL ITERATION

This is very similar to the Gauss-Jacobi iteration. Define

$$x_i^{(m+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{i,j} x_j^{(m+1)} - \sum_{j=i+1}^{n} a_{i,j} x_j^{(m)} \right] \tag{10}$$

for $i = 1, ..., n$ and for $m = 0, 1, ...,$ with some given $x^{(0)}$. For serial computers, this method has been preferred over Gauss-Jacobi iteration, for reasons discussed in §8.8. However, when using parallel and vector computers, the advantage is by no means so clear.

The idea for (10) is to make immediate use of each newly computed iterate. However, on a parallel or vector-pipeline computer, one may wish to computer several iterates simultaneously. We return later to this discussion.

GAUSS-JACOBI: CONVERGENCE

Subtract (9) from (8), obtaining

$$e_i^{(m+1)} = \frac{-1}{a_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{i,j} e_j^{(m)}, \qquad i = 1, ..., n \tag{11}$$

with $e^{(m)} = x - x^{(m)}$. In matrix form,

$$e^{(m+1)} = M e^{(m)}, \qquad m \geq 0 \tag{12}$$

$$M = - \begin{bmatrix} 0 & \frac{a_{1,2}}{a_{1,1}} & \dots & & \frac{a_{1,n}}{a_{1,1}} \\ \frac{a_{2,1}}{a_{2,2}} & 0 & & & \frac{a_{2,n}}{a_{2,2}} \\ \vdots & \ddots & & & \vdots \\ & & & 0 & \frac{a_{n-1,n}}{a_{n-1,n-1}} \\ \frac{a_{n,1}}{a_{n,n}} & & \dots & \frac{a_{n,n-1}}{a_{n,n}} & 0 \end{bmatrix} \tag{13}$$

From (12),

$$e^{(m)} = M^m e^{(0)}, \qquad m \geq 0 \tag{14}$$

$$e^{(m)} = M^m e^{(0)}, \qquad m \geq 0$$

For arbitrary $x^{(0)}$, we have $x^{(m)} \to x$ if and only if

$$r_\sigma(M) < 1 \tag{15}$$

This is satisfied if $\|M\| < 1$ for some operator matrix norm.

Case 1: $\|M\|_\infty < 1$. This means

$$\sum_{j=1,\, j\neq i}^{n} \left|a_{i,j}\right| < \left|a_{i,i}\right|, \qquad i = 1, ..., n$$

We say A is *diagonally dominant*. In this case,

$$\left\|x - x^{(m+1)}\right\|_\infty \leq \|M\|_\infty \left\|x - x^{(m)}\right\|_\infty, \qquad m \geq 0$$

Case 2: $\|M\|_1 < 1$. This means

$$\sum_{i=1}^{n} \left|\frac{a_{i,j}}{a_{i,i}}\right| < 1, \qquad j = 1, ..., n$$

In this case,

$$\left\|x - x^{(m+1)}\right\|_1 \leq \|M\|_1 \left\|x - x^{(m)}\right\|_1, \qquad m \geq 0$$

EXAMPLE

$$\begin{bmatrix} 9 & 1 & 1 \\ 2 & 10 & 3 \\ 3 & 4 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 19 \\ 0 \end{bmatrix}$$

m	$x_1^{(m)}$	$x_2^{(m)}$	$x_3^{(m)}$	$\left\|x - x^{(m)}\right\|_\infty$	$Ratio$
0	0	0	0	$2.00E+0$	
1	1.1111	1.9000	-0.9939	$1.00E+0$	.500
2	0.9000	1.6778	-0.9939	$3.22E-1$	.322
3	1.0351	2.0182	-0.8556	$1.44E-1$	.448
4	0.9819	1.9496	-1.0162	$5.06E-2$	.462
5	1.0074	2.0085	-0.9768	$2.32E-2$	.462
6	0.9965	1.9915	-1.0051	$8.45E-3$	.364
30	1.0000	2.0000	-1.0000	$3.01E-11$	.447
31	1.0000	2.0000	-1.0000	$1.35E-11$	.447

$$M = \begin{bmatrix} 0 & -\frac{1}{9} & -\frac{1}{9} \\ -\frac{2}{10} & 0 & -\frac{3}{10} \\ -\frac{3}{11} & -\frac{4}{11} & 0 \end{bmatrix}$$

$$\|M\|_\infty = \frac{7}{11} \doteq .636$$

Note that for the actual ratios,

$$\frac{\left\|x - x^{(m+1)}\right\|_\infty}{\left\|x - x^{(m)}\right\|_\infty} \to 0.447$$

GAUSS-SEIDEL: CONVERGENCE

To obtain an error equation, subtract (10) from (8), obtaining

$$e_i^{(m+1)} = -\sum_{j=1}^{i-1} \frac{a_{i,j}}{a_{ii}} e_j^{(m+1)} - \sum_{j=i+1}^{n} \frac{a_{i,j}}{a_{ii}} e_j^{(m)} \qquad (16)$$

for $i = 1, ..., n$. We can again write the error equation in the form

$$e^{(m+1)} = \widetilde{M} e^{(m)}, \qquad m \geq 0 \qquad (17)$$

for a suitable $\widetilde{M}$; but bounding $\left\| \widetilde{M} \right\|$ is much more difficult in this case.

We first give an alternative error analysis which is valid in the case $\|M\|_\infty < 1$ for the matrix M of (13) and the Gauss-Jacobi method.

Introduce

$$\alpha_i = \sum_{j=1}^{i-1} \left| \frac{a_{i,j}}{a_{ii}} \right|, \qquad \beta_i = \sum_{j=i+1}^{n} \left| \frac{a_{i,j}}{a_{ii}} \right|$$

with $\alpha_1 = \beta_n = 0$. Taking bounds in (16),

$$\left| e_i^{(m+1)} \right| \leq \alpha_i \left\| e^{(m+1)} \right\|_\infty + \beta_i \left\| e^{(m)} \right\|_\infty, \quad i = 1, ..., n \tag{18}$$

Let k be an index for which

$$\left| e_k^{(m+1)} \right| = \left\| e^{(m+1)} \right\|_\infty$$

Then using $i = k$ in (18),

$$\left\| e^{(m+1)} \right\|_\infty \leq \alpha_k \left\| e^{(m+1)} \right\|_\infty + \beta_k \left\| e^{(m)} \right\|_\infty$$

$$\left\| e^{(m+1)} \right\|_\infty \leq \frac{\beta_k}{1 - \alpha_k} \left\| e^{(m)} \right\|_\infty$$

Define

$$\eta = \max_i \frac{\beta_i}{1 - \alpha_i}$$

Then

$$\left\| e^{(m+1)} \right\|_\infty \leq \eta \left\| e^{(m)} \right\|_\infty$$

$$\left\|e^{(m+1)}\right\|_\infty \leq \eta \left\|e^{(m)}\right\|_\infty \tag{19}$$

We can show

$$\|M\|_\infty < 1 \quad \Rightarrow \quad \eta \leq \|M\|_\infty \tag{20}$$

Note that

$$\|M\|_\infty = \max_i (\alpha_i + \beta_i)$$

Then (20) follows from

$$(\alpha_i + \beta_i) - \frac{\beta_i}{1-\alpha_i} = \frac{\alpha_i \left[1 - \alpha_i - \beta_i\right]}{1-\alpha_i} \geq 0$$

Combining (19)-(20), we have that if Gauss-Jacobi iteration converges because $\|M\|_\infty < 1$, then Gauss-Seidel iteration also converges, but probably more rapidly.

EXAMPLE

$$\begin{bmatrix} 9 & 1 & 1 \\ 2 & 10 & 3 \\ 3 & 4 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 19 \\ 0 \end{bmatrix}$$

m	$x_1^{(m)}$	$x_2^{(m)}$	$x_3^{(m)}$	$\left\|x - x^{(m)}\right\|_\infty$	$Ratio$
0	0	0	0	$2.00E+0$	
1	1.1111	1.6778	-0.9131	$3.22E-1$	.161
2	1.0262	1.9687	-0.9958	$3.13E-2$	.097
3	1.0030	1.9981	-1.0001	$3.00E-3$	.096
4	1.0002	2.0000	-1.0001	$2.24E-4$	.074
5	1.0000	2.0000	-1.0000	$1.65E-5$	.074
6	1.0000	2.0000	-1.0000	$2.58E-6$	.155

In this case,

$$\eta = \frac{3}{8} = 0.375$$

$$\left\|\widetilde{M}\right\|_\infty = 0.3$$

A GENERAL FRAMEWORK

We want to solve $Ax = b$ by iteration. Suppose we decompose A as

$$A = N - P$$

for two matrices N and P. Then $Ax = b$ can be rewritten as

$$Nx = b + Px \tag{21}$$

Define an iteration by

$$Nx^{(m+1)} = b + Px^{(m)}, \qquad m = 0, 1, ... \tag{22}$$

In this instance, we must be able to solve linear systems $Nz = g$ with greater ease than we can solve systems $Ax = b$.

For the error, subtract (22) from (21), obtaining

$$\begin{aligned} Ne^{(m+1)} &= Pe^{(m)} \\ e^{(m+1)} &= N^{-1}Pe^{(m)} \equiv Me^{(m)} \end{aligned}$$

$$e^{(m)} = M^m e^{(0)}, \qquad m \geq 0 \tag{23}$$

Gauss-Jacobi: Let $P = N - A$ and

$$N = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}$$

Solving $Nz = g$ is equivalent to solving a linear system with a diagonal coefficient matrix.

Gauss-Seidel: Let $P = N - A$ and

$$N = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

Solving $Nz = g$ amounts to solving a lower triangular linear system.

In the language of the residual correction method of §8.5, we are using $A^{-1} \approx N^{-1}$. Equivalently, we are approximating the linear system $Ax = b$ by the linear system $Nx \approx b$.

From (23),

$$e^{(m)} = Me^{(0)}, \qquad m \geq 0$$

the method converges if and only if $r_\sigma(M) < 1$, with $M = N^{-1}P$. In the case of the Gauss-Seidel method, it is not easy to examine this matrix $N^{-1}P$ exactly. However, by other means, the following can be shown.

Theorem: Let A be Hermitian with positive diagonal elements. Then the Gauss-Seidel method for solving $Ax = b$ converges for any choice of $x^{(0)}$ if and only if the matrix A is positive definite.

Later we will discuss other iteration methods which can be put into this general framework.