Solving Linear Systems of Equations

Many practical problems could be reduced to solving a linear system of equations formulated as $A\mathbf{x} = \mathbf{b}$. This chapter studies the computational issues about directly and iteratively solving $A\mathbf{x} = \mathbf{b}$.

- A Linear System of Equations
- Vector and Matrix Norms
- Matrix Condition Number ($Cond(A) = \|A\| \cdot \|A^{-1}\|$)
- Direct Solution for $A\mathbf{x} = \mathbf{b}$

    LU-decomposition by Gaussian Elimination

    Gaussian Elimination with Partial Pivoting

    Cholesky Algorithm for $A = LL^t$ (A is positive definite)

- Iterative Solutions

    Jacobi method

    Gauss-Seidel method

    Other methods

- Applications

Overdetermined, Underdetermined, Homogeneous Systems

$$\begin{array}{ccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
. & & . & & . & & . & & . \\
. & & . & & . & & . & & . \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m
\end{array}$$

$A\mathbf{x} = \mathbf{b}$

Definition: A linear system is said to be *overdetermined* if there are more equations than unknowns ($m > n$), *underdetermined* if $m < n$, *homogeneous* if $b_i = 0$, $\forall\, 1 \leq i \leq m$.

$$(A) \begin{array}{rcrcr} x & + & y & = & 1 \\ x & - & y & = & 3 \\ -x & + & 2y & = & -2 \end{array} \quad (B) \begin{array}{rcrcr} x & + & y & = & 3 \\ x & - & y & = & 1 \\ 2x & + & y & = & 5 \end{array} \quad (C) \begin{array}{rcrcr} x & + & y & = & 2 \\ 2x & + & 2y & = & 4 \\ -x & - & y & = & -2 \end{array}$$

(A) has no solution, (B) has unique solution, (C) has infinitely many solutions

$$(D) \begin{array}{rcrcrcr} x & + & 2y & + & z & = & -1 \\ 2x & + & 4y & + & 2z & = & 3 \end{array} \qquad (E) \begin{array}{rcrcrcr} x & + & 2y & + & z & = & 5 \\ 2x & - & y & + & z & = & 3 \end{array}$$

(D) has no solution, (E) has infinitely many solutions

Some Special Matrices

$A = [a_{ij}] \in R^{n \times n}$

- *Diagonal* if $a_{ij} = 0 \; \forall \; i \neq j$
- $Lower - \Delta$ if $a_{ij} = 0$ if $j > i$
- $Unit\ lower - \Delta$ if A is lower-Δ with $a_{ii} = 1$
- $Lower\ Hessenberg$ if $a_{ij} = 0$ for $j > i + 1$
- *Band* matrix with *bandwidth* $2k + 1$ if $a_{ij} = 0$ for $|i - j| > k$

□ A band matrix with *bandwidth* 1 is *diagonal*

□ A band matrix with *bandwidth* 3 is *tridiagonal*

□ A band matrix with *bandwidth* 5 is *pentadiagonal*

□ A lower and upper Hessenberg matrix is *tridiagonal*

$$A_1 = \begin{bmatrix} 7 & 0 & 0 \\ 1 & 8 & 0 \\ 2 & 3 & 9 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 5 & 2 & 0 & 0 \\ 1 & 6 & 4 & 0 \\ 2 & 3 & 7 & 3 \\ 1 & 2 & 0 & 8 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 5 & 2 & 0 & 0 \\ 1 & 6 & 4 & 0 \\ 0 & 3 & 7 & 3 \\ 0 & 2 & 0 & 8 \end{bmatrix}$$

A Direct Solution of Linear Systems

A linear system

$$\begin{array}{rcrcrcr} 2x & + & y & + & z & = & 5 \\ 4x & - & 6y & & & = & -2 \\ -2x & + & 7y & + & 2z & = & 9 \end{array}$$

A matrix representation

$$A\mathbf{x} = \mathbf{b}, \quad or \quad \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}.$$

♣ Solution using MATLAB

$\rangle\rangle\ A = [2, 1, 2;\ 4, -6, 0;\ \ -2, 7, 2];$

$\rangle\rangle\ b = [5,\ \ -2,\ 9]';$

$\rangle\rangle\ x = A\backslash b \quad (x = [1;\ 1;\ 2])$

Elementary Row Operations

(1) Interchange two rows: $A_r \leftrightarrow A_s$

(2) Multiply a row by a nonzero real number: $A_r \leftarrow \alpha A_r$

(3) Replace a row by its sum with a multiple of another row: $A_s \leftarrow \alpha A_r + A_s$

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

♣ *Example*

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}, \; E_1 A = \begin{bmatrix} 4 & -6 & 0 \\ 2 & 1 & 1 \\ -2 & 7 & 2 \end{bmatrix}, \; E_2 A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ 4 & -14 & -4 \end{bmatrix}, \; E_3 A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ 0 & 8 & 3 \end{bmatrix}$$

Let

$$L_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \; L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \; L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$$

Then

$$L_3 L_2 L_1 A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U \quad (Upper - \Delta)$$

$$A = (L_1^{-1} L_2^{-1} L_3^{-1}) U = LU, \quad where\ L\ is\ unit\ lower - \Delta$$

Computing An Inverse Matrix By Elementary Row Operations

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 2 & 4 & 7 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

$$E_1 A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 2 & 4 & 7 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_1 I$$

$$E_2 E_1 A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = E_2 E_1 I$$

$$E_3 E_2 E_1 A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & -2 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = E_3 E_2 E_1 I$$

$$E_4 E_3 E_2 E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 9 & -2 & -3 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = E_4 E_3 E_2 E_1 I = A^{-1}$$

where the elementary matrices are

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_4 = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

LU-Decomposition

If a matrix A can be decomposed into the product of a unit lower-Δ matrix L and an upper-Δ matrix U, then the linear system $A\mathbf{x} = \mathbf{b}$ can be written as $LU\mathbf{x} = \mathbf{b}$. The problem is reduced to solving two simpler triangular systems $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$ by forward and back substitutions.

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}, \quad L_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then

$$L_3 L_2 L_1 A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U \quad \Rightarrow \quad A = L_1^{-1} L_2^{-1} L_3^{-1} U = LU$$

where

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = LU$$

◇ If $\mathbf{b} = [5, -2, 9]^t$, then $\mathbf{y} = [5, -12, 2]^t$ and $\mathbf{x} = [1, 1, 2]^t$

$$B = \begin{bmatrix} 2 & -2 & 3 \\ -2 & 3 & -4 \\ 4 & -3 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Analysis of Gaussian Elimination

♣ *Algorithm*

for $i = 1, 2, \cdots, n-1$

for $k = i+1, i+2, \cdots, n$

$m_{ki} \ \leftarrow \ a_{ki}/a_{ii}$ if $a_{ii} \neq 0$

$a_{ki} \ \leftarrow \ a_{ki}$

for $j = i+1, i+2, \cdots, n$

$a_{kj} \ \leftarrow \ a_{kj} - m_{ki} * a_{ij}$

endfor

endfor

endfor

- *The Worst Computational Complexity is* $O(\frac{2}{3}n^3)$

1. # of divisions are $(n-1) + (n-2) + \cdots + 1 = \frac{n(n-1)}{2}$
2. # of multiplications are $(n-1)^2 + (n-2)^2 + \cdots + 1^2 = \frac{n(n-1)(2n-1)}{6}$
3. # of subtractions are $(n-1)^2 + (n-2)^2 + \cdots + 1^2 = \frac{n(n-1)(2n-1)}{6}$

The Analysis of Gaussian Elimination and Back Substitution to solve Ax=b

$\frac{2}{3}n^3 + \frac{3}{2}n^2 - \frac{7}{6}n$

$$
\begin{array}{rcccccccccc}
R_1: & a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
R_2: & a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
 & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
R_i: & a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n & = & b_i \\
 & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
R_n: & a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nn}x_n & = & b_n
\end{array}
$$

By Guassian Elimination, we need $C_1 = [\sum_{k=1}^{n}(k+1)(k-1) + \sum_{k=1}^{n} k(k-1)]$ flops to reduce the above linear system of equations equivalent to the following upper triangular system.

$$
\begin{array}{rcccccccccc}
R_1: & u_{11}x_1 & + & u_{12}x_2 & + & \cdots & + & u_{1n}x_n & = & c_1 \\
 & & & \vdots & & \vdots & & \vdots & & \vdots \\
R_i: & & & u_{ii}x_i & + & \cdots & + & u_{in}x_n & = & c_i \\
 & & & & & \vdots & & \vdots & & \vdots \\
R_n: & & & & & & & u_{nn}x_n & = & c_n
\end{array}
$$

We need $C_2 = \sum_{k=1}^{n}(2k-1) = n^2$ flops to solve an upper triangular linear system of equations. Therefore, the total number of flops of solving $A\mathbf{x} = \mathbf{b}$ is summarized as

$$C_1 + C_2 = \frac{2}{3}n^3 + \frac{3}{2}n^2 - \frac{7}{6}n$$

PA=LU

Let $A \in R^{4\times 4}$ and $L_3P_3L_2P_2L_1P_1A = U$ by Gaussian Elimination with Partial Pivoting. Denote

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ \alpha_2 & 0 & 1 & 0 \\ \alpha_3 & 0 & 0 & 1 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \alpha_4 & 1 & 0 \\ 0 & \alpha_5 & 0 & 1 \end{bmatrix}, \quad L_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \alpha_6 & 1 \end{bmatrix}$$

Without loss of generality, let

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Then $P_2L_1 = (P_2L_1P_2^{-1})P_2 = L_1^{(2)}P_2$, where

$$L_1^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_3 & 1 & 0 & 0 \\ \alpha_2 & 0 & 1 & 0 \\ \alpha_1 & 0 & 0 & 1 \end{bmatrix}$$

Theorem: For any $A \in R^{n\times n}$, there exists a permutation matrix P such that $PA = LU$, where L is unit $lower - \Delta$ and U is $upper - \Delta$.

Gaussian Elimination with partial Pivoting

♣ *Algorithm*

for $i = 1, 2, \cdots, n$

p(i)=i

endfor

for $i = 1, 2, \cdots, n-1$

(a) select a pivotal element $a_{p(j),i}$ such that $|a_{p(j),i}| = max_{i \le k \le n} |a_{p(k),i}|$

(b) $p(i) \longleftrightarrow p(j)$

(c) for $k = i+1, i+2, \cdots, n$

$m_{p(k),i} = a_{p(k),i} / a_{p(i),i}$

for $j = i+1, i+2, \cdots, n$

$a_{p(k),j} = a_{p(k),j} - m_{p(k),i} * a_{p(i),j}$

endfor

endfor

endfor

• *An example*

$$A = \begin{bmatrix} 0 & 9 & 1 \\ 1 & 2 & -2 \\ 2 & -5 & 4 \end{bmatrix} \quad \Rightarrow \quad P_{23} P_{13} A = LU = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 & -5 & 4 \\ 0 & 9 & 1 \\ 0 & 0 & \frac{-9}{2} \end{bmatrix}$$

Matlab Codes for Gaussian Elimination with Partial Pivoting

```
%%-----------------------------------------------------------------------------%%
%% gausspp.m - drive of Gaussian Elimination wit Partial Pivoting              %%
%%-----------------------------------------------------------------------------%%
fin=fopen('gaussmat.dat','r');
n=fscanf(fin,'%d',1);
A=fscanf(fin,'%f',[n n]);   A=A';
b=fscanf(fin,'%f',n);
X=gausspivot(A,b,n)

%%-----------------------------------------------------------------------------%%
%% gausspivot.m - Gaussian elimination with Partial Pivoting                   %%
%%-----------------------------------------------------------------------------%%
function X=gausspivot(A,b,n)

if (abs(det(A))<eps)
  disp(sprintf('A is singular with det=%f\n',det(A)))
end

C=[A, b];
%------- Gaussian Elimination with Partial Pivoting ------%
for i=1:n-1
  [pivot, k]=max(abs(C(i:n,i)));
  if (k>1)
    temp=C(i,:);
    C(i,:)=C(i+k-1,:);
    C(i+k-1,:)=temp;
  end
  m(i+1:n,i)= -C(i+1:n,i)/C(i,i);
  C(i+1:n,:)=C(i+1:n,:)+m(i+1:n,i)*C(i,:);
end
%------- Back substitution -------%
X=zeros(n,1); %% Let X be a column vector of size n
X(n)=C(n,n+1)/C(n,n);
for i=n-1:-1:1
  X(i)=(C(i,n+1)-C(i,i+1:n)*X(i+1:n))/C(i,i);
end
```

Doolittle's LU Factorization

♣ *Algorithm:* $A \in R^{n\times n}$, $A = LU$, L is unit lower-Δ, U is upper-Δ.

for $k = 0, 1, \cdots, n-1$

$L_{kk} \leftarrow 1$

for $j = k, k+1, \cdots, n-1$

$U_{kj} \leftarrow A_{kj} - \sum_{s=0}^{k-1} L_{ks} U_{sj}$

endfor

for $i = k+1, k+2, \cdots, n-1$

$L_{ik} \leftarrow \left[A_{ik} - \sum_{s=0}^{k-1} L_{is} U_{sk}\right] / U_{kk}$

endfor

endfor

$$A = \begin{bmatrix} 9 & 3 & -3 \\ 3 & 17 & 3 \\ -3 & 3 & 27 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ -\frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 & -3 \\ 0 & 16 & 4 \\ 0 & 0 & 25 \end{bmatrix} = LU$$

Crout's LU Factorization

♣ *Algorithm:* $A \in R^{n\times n}$, $A = LU$, L is lower-Δ, U is unit upper-Δ.

for $k = 0, 1, \cdots, n-1$

$U_{kk} \leftarrow 1$

for $i = k, k+1, \cdots, n-1$

$L_{ik} \leftarrow A_{ik} - \sum_{s=0}^{k-1} L_{is} U_{sk}$

endfor

for $j = k+1, k+2, \cdots, n-1$

$U_{kj} \leftarrow \left[A_{kj} - \sum_{s=0}^{k-1} L_{ks} U_{sj}\right] / L_{kk}$

endfor

endfor

$$A = \begin{bmatrix} 9 & 3 & -3 \\ 3 & 17 & 3 \\ -3 & 3 & 27 \end{bmatrix} = \begin{bmatrix} 9 & 0 & 0 \\ 3 & 16 & 0 \\ -3 & 4 & 25 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix} = LU$$

Cholesky Algorithm

♣ *Algorithm:* $A \in R^{n\times n}$, $A = LL^t$, A is positive definite and L is lower-Δ.

for $j = 0, 1, \cdots, n-1$

$L_{jj} \;\leftarrow\; \left[A_{jj} - \sum_{k=0}^{j-1} L_{jk}^2\right]^{1/2}$

for $i = j+1, j+2, \cdots, n-1$

$L_{ij} \;\leftarrow\; \left[A_{ij} \;-\; \sum_{k=0}^{j-1} L_{ik}L_{jk}\right]/L_{jj}$

endfor

endfor

$$A = \begin{bmatrix} 9 & 3 & -3 \\ 3 & 17 & 3 \\ -3 & 3 & 27 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 4 & 0 \\ -1 & 1 & 5 \end{bmatrix} \begin{bmatrix} 3 & 1 & -1 \\ 0 & 4 & 1 \\ 0 & 0 & 5 \end{bmatrix} = LL^t$$

Vector Norms

Definition: A vector norm on R^n is a function

$$\tau \ : \ R^n \ \rightarrow \ R^+ = \{x \geq 0 | \ x \in R\}$$

that satisfies

(1) $\tau(\mathbf{x}) > 0 \ \ \forall \ \mathbf{x} \neq \mathbf{0}, \ \tau(\mathbf{0}) = 0$

(2) $\tau(c\mathbf{x}) = |c|\tau(\mathbf{x}) \ \forall \ c \in R, \ \mathbf{x} \in R^n$

(3) $\tau(\mathbf{x} + \mathbf{y}) \leq \tau(\mathbf{x}) + \tau(\mathbf{y}) \ \forall \ \mathbf{x}, \mathbf{y} \in R^n$

Hölder norm (p-norm) $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \ \ for \ p \geq 1.$

(p=1) $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ (Mahattan or City-block distance)

(p=2) $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$ (Euclidean distance)

(p=∞) $\|\mathbf{x}\|_\infty = max_{1 \leq i \leq n}\{|x_i|\}$ (∞-norm)

Matrix Norms

Definition: A matrix norm on $R^{m \times n}$ is a function

$$\tau \;:\; R^{m \times n} \;\rightarrow\; R^{+} = \{x \geq 0 |\; x \in R\}$$

that satisfies

(1) $\tau(A) > 0 \;\; \forall\; A \neq O,\, \tau(O) = 0$

(2) $\tau(cA) = |c|\tau(A) \;\forall\; c \in R,\;\; A \in R^{m \times n}$

(3) $\tau(A + B) \leq \tau(A) + \tau(B) \;\; \forall\; A, B \in R^{m \times n}$

Consistency Property: $\tau(AB) \leq \tau(A)\tau(B) \;\; \forall\; A, B$

(a) $\tau(A) = max\{|a_{ij}| \mid 1 \leq i \leq m,\; 1 \leq j \leq n\}$

(b) $\|A\|_F = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2\right]^{1/2}$ (Fröbenius norm)

Subordinate Matrix Norm: $\|A\| = max_{\|\mathbf{x}\| \neq \mathbf{0}}\{\|A\mathbf{x}\|/\|\mathbf{x}\|\}$

(1) If $A \in R^{m \times n}$, then $\|A\|_1 = max_{1 \leq j \leq n}\left(\sum_{i=1}^{m} |a_{ij}|\right)$

(2) If $A \in R^{m \times n}$, then $\|A\|_\infty = max_{1 \leq i \leq m}\left(\sum_{j=1}^{n} |a_{ij}|\right)$

(3) Let $A \in R^{n \times n}$ be real symmetric, then $\|A\|_2 = max_{1 \leq i \leq n}|\lambda_i|$, where $\lambda_i \in \lambda(A)$

Matrix Condition Number

♣ $Cond(A) = \|A\| \cdot \|A^{-1}\|$

For the matrix

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 3 & -1 & 4 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 2.5 & -0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix}$$

$$Cond_1(A) = \|A\|_1 \cdot \|A^{-1}\|_1 = 6 \times 4.5 = 27$$
$$Cond_\infty(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty = 8 \times 3.5 = 28$$
$$Cond_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = 5.7229 \times 3.0566 = 17.4930$$

Sensitivity and Error Bounds

Let $\mathbf{x}$ and $\hat{\mathbf{x}}$ be the solutions to $A\mathbf{x} = \mathbf{b}$ and $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, respectively, where $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$, $\hat{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b}$. Then

$$\|\mathbf{b}\| = \|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\| \quad or \quad \|\mathbf{x}\| \geq \frac{\|\mathbf{b}\|}{\|A\|}$$

$$\|\Delta\mathbf{x}\| = \|A^{-1}\Delta\mathbf{b}\| \leq \|A^{-1}\| \cdot \|\Delta\mathbf{b}\|$$

Thus

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\| \cdot \|\Delta\mathbf{b}\| \cdot \frac{\|A\|}{\|\mathbf{b}\|} \leq Cond(A) \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$$

Similarly, suppose that $(A + E)\tilde{\mathbf{x}} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{b}$, we have

$$\frac{\|\Delta\mathbf{x}\|}{\|\tilde{\mathbf{x}}\|} \leq Cond(A) \cdot \frac{\|E\|}{\|A\|}$$

Further derivations to get

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq Cond(A) \cdot \left(\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|E\|}{\|A\|}\right)$$

Iterative Methods for Solving Linear Systems

1. Iterative methods are msot useful in solving *large sparse* system.
2. One advantage is that the iterative methods may not require any extra storage and hence are more practical.
3. One disadvantage is that after solving $A\mathbf{x} = \mathbf{b}_1$, one must start over again from the beginning in order to solve $A\mathbf{x} = \mathbf{b}_2$.

♣ *Jacobi Method*

Given $A\mathbf{x} = \mathbf{b}$, write $A = C - M$, where C is nonsingular and easily invertible. Then

$$A\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad (C - M)\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad C\mathbf{x} = M\mathbf{x} + \mathbf{b}$$
$$\mathbf{x} = C^{-1}M\mathbf{x} + C^{-1}\mathbf{b} \quad \Rightarrow \quad \mathbf{x} = B\mathbf{x} + \mathbf{c}, \text{ where}$$
$$B = C^{-1}M,\ \mathbf{c} = C^{-1}\mathbf{b}$$

Suppose we start with an inital $\mathbf{x}^{(0)}$, then

$$\mathbf{x}^{(1)} = B\mathbf{x}^{(0)} + \mathbf{c} \quad \textit{and} \quad \mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{c}$$

Jacobi Iterative Method for Solving Linear Systems

Suppose $\mathbf{x}$ is a solution to $A\mathbf{x} = \mathbf{b}$, then

$$
\begin{aligned}
&\mathbf{x}^{(1)} - \mathbf{x} = (B\mathbf{x}^{(0)} + \mathbf{c}) - (B\mathbf{x} + \mathbf{c}) = B(\mathbf{x}^{(0)} - \mathbf{x}) \\
&\mathbf{x}^{(k)} - \mathbf{x} = B^k(\mathbf{x}^{(0)} - \mathbf{x}) \\
&\|\mathbf{x}^{(k)} - \mathbf{x}\| \le \|B^k\| \cdot \|\mathbf{x}^{(0)} - \mathbf{x}\| \le \|B\|^k \cdot \|\mathbf{x}^{(0)} - \mathbf{x}\| \\
&\|\mathbf{x}^{(k)} - \mathbf{x}\| \;\to\; 0 \text{ as } k \;\to\; \infty \text{ if } \|B = C^{-1}M\| < 1.
\end{aligned}
$$

- For simplest computations, 1-norm or ∞-norm is used.
- The simplest choice of C, M with $A = C - M$ is $C = Diag(A)$, $M = -(A - C)$.

Theorem: Let $\mathbf{x}^{(0)} \in R^n$ be arbitrary and define $\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{c}$ for $k = 0, 1, \cdots$ If $\mathbf{x}$ is a solution to $A\mathbf{x} = \mathbf{b}$, then the necessary and sufficient condition for $\mathbf{x}^{(k)} \;\to\; \mathbf{x}$ is $\|B\| < 1$.

Theorem: If A is diagonally dominant, i.e., $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$ for $1 \le i \le n$. Let $A = C - M$ and $B = C^{-1}M$, then

$$
\|B\|_\infty = max_{1\le i\le n} \left\{ \sum_{j=1}^{n} |b_{ij}| \right\} = max_{1\le i\le n} \left\{ \sum_{j\neq i} \frac{|a_{ij}|}{|a_{ii}|} \right\} \;<\; 1
$$

Example:

$$
A = \begin{bmatrix} 10 & 1 \\ 2 & 10 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 11 \\ 12 \end{bmatrix}, \quad \mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

Let

$$
C = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}, \quad M = \begin{bmatrix} 0 & -1 \\ -2 & 0 \end{bmatrix},
$$

Then

$$
B = C^{-1}M = \begin{bmatrix} 0 & -0.1 \\ -0.2 & 0 \end{bmatrix}, \quad \mathbf{c} = C^{-1}\mathbf{b} = \begin{bmatrix} 1.1 \\ 1.2 \end{bmatrix}
$$

$$
\mathbf{x}^{(1)} = B\mathbf{x}^{(0)} + \mathbf{c} = \begin{bmatrix} 1.1 \\ 1.2 \end{bmatrix}, \quad \mathbf{x}^{(2)} = \begin{bmatrix} 0.98 \\ 0.98 \end{bmatrix}, \quad \mathbf{x}^{(3)} = \begin{bmatrix} 1.002 \\ 1.004 \end{bmatrix}
$$

Gauss-Seidel Method

Given $A\mathbf{x} = \mathbf{b}$, write $A = C - M$, where C is nonsingular and easily invertible.

Jacobi: $C = diag(a_{11}, a_{22}, \cdots, a_{nn})$, $M = -(A - C)$

Gauss-Seidel: $A = (D-L)-U = C-M$, where $C = D-L$, $M = U$ for Jacobi

Let $\mathbf{x}^{(0)} \in R^n$ be nonzero, then $C\mathbf{x}^{(k+1)} = M\mathbf{x}^{(k)} + \mathbf{b}$ implies that

$$(D - L)\mathbf{x}^{(k+1)} = U\mathbf{x}^{(k)} + \mathbf{b}$$

$$D\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)} + \mathbf{b}$$

$$x_1^{(k+1)} = \frac{1}{a_{11}}\left(-\sum_{j=2}^{n} a_{1j}x_j^{(k)} + b_1\right)$$

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + b_i\right), \quad for \quad i = 2, 3, \cdots, n$$

The difference between *Jacobi* and *Gauss-Seidel* iteration is that in the latter case, one is using the coordinates of $\mathbf{x}^{(k+1)}$ as soon as they are calculated rather than in the next iteration. The program for *Gauss-Seidel* is much simpler.

Convergence of Gauss-Seidel Iterations

Theorem: If A is diagonally dominant, i.e., $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$ for $1 \leq i \leq n$. Then *Gauss-Seidel* iteration converges to a solution of $A\mathbf{x} = \mathbf{b}$.

Proof: Denote $A = (D - L) - U$ and let

$$\alpha_j = \sum_{i=1}^{j-1} |a_{ji}|, \quad \beta_j = \sum_{i=j+1}^{n} |a_{ji}|, \quad and \quad R_j = \frac{\beta_j}{|a_{jj}| - \alpha_j}$$

Since A is diagonally dominant, then $|a_{jj}| > \alpha_j + \beta_j$, thus

$$R_j = \frac{\beta_j}{|a_{jj}| - \alpha_j} < \frac{|a_{jj}| - \alpha_j}{|a_{jj}| - \alpha_j} = 1 \;\; \forall\ 1 \leq j \leq n$$

Therefore, $R = Max_{1\leq j\leq n} R_j < 1$.

The remaining problem is to show that $\|B\|_\infty = Max_{\|\mathbf{x}\|_\infty=1}\|B\mathbf{x}\|_\infty \leq R < 1$, where $B = C^{-1}M = (D - L)^{-1}U$.

Let $\|\mathbf{x}\|_\infty = 1$ and $\mathbf{y} = B\mathbf{x}$, then $\|\mathbf{y}\|_\infty = Max_{1\leq i\leq n}|y_i| = |y_k|$ for some k.
Then

$$\mathbf{y} = B\mathbf{x} = (D - L)^{-1}U\mathbf{x}$$

$$(D - L)\mathbf{y} = U\mathbf{x} \;\Rightarrow\; D\mathbf{y} = L\mathbf{y} + U\mathbf{x}$$

$$\mathbf{y} = D^{-1}(L\mathbf{y} + U\mathbf{x})$$

Then

$$y_k = \tfrac{1}{a_{kk}} \left(- \textstyle\sum_{i=1}^{k-1} a_{ki}y_i - \sum_{i=k+1}^{n} a_{ki}x_i \right)$$

$$\|\mathbf{y}\|_\infty = |y_k| \leq \tfrac{1}{|a_{kk}|} \left(\alpha_k\|\mathbf{y}\|_\infty + \beta_k\|\mathbf{x}\|_\infty\right)$$

which implies that for all $\mathbf{x}$ with $\|\mathbf{x}\|_\infty = 1$,

$$\|B\mathbf{x}\|_\infty = \|\mathbf{y}\|_\infty \leq \tfrac{\beta_k}{|a_{kk}|-\alpha_k} = R_k < 1$$

Thus, $\|B\|_\infty = Max_{\|\mathbf{x}\|_\infty=1}\|B\mathbf{x}\|_\infty \leq R < 1$

Example for Gauss-Seidel Iterations

Example:

$$A = \begin{bmatrix} 10 & 1 \\ \\ 2 & 10 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ \\ 0 & 10 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ \\ -2 & 0 \end{bmatrix} - \begin{bmatrix} 0 & -1 \\ \\ 0 & 0 \end{bmatrix} = D - L - U$$

$$\mathbf{b} = \begin{bmatrix} 11 \\ 12 \end{bmatrix}, \quad \mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Then

$$x_1^{(1)} = \tfrac{1}{10}(-a_{12}x_2^{(0)} + b_1) = 1.1$$

$$x_2^{(1)} = \tfrac{1}{10}(-a_{21}x_1^{(1)} - 0 + b_2) = 0.98$$

Moreover,

$$x_1^{(2)} = \tfrac{1}{10}(-a_{12}x_2^{(1)} + b_1) = 1.002$$

$$x_2^{(2)} = \tfrac{1}{10}(-a_{21}x_1^{(2)} - 0 + b_2) = 0.9996$$

Thus

$$\mathbf{x}^{(1)} = [1.1, 0.98]^t$$

$$\mathbf{x}^{(2)} = [1.002, 0.9996]^t$$

$$\mathbf{x}^{(3)} = [1.00004, 0.99992]^t$$