Numerical Analysis : [MA214]
Lecture 11

Instructor: Prof. Tony J. Puthenpurakal

Department of Mathematics
Indian Institute of Technology Bombay

tputhen@math.iitb.ac.in

September 5, 2017

Recall

Last time we have discussed order of convergence

$$x_n \longrightarrow \xi$$

$$e_n = \xi - x_n$$

If there exists $p \geq 0$ and a constant $C \neq 0$ such that

$$\lim_{n \to \infty} \frac{|e_{n+1}|}{|e_n|^p} = C.$$

This p is called "*order of convergence*" and C is called the *asymptotic error constant.*

Examples *1. Fixed point iteration*

ξ fixed point of $g : I \longrightarrow I$ and $g'(\xi) \neq 0$.
Then $p = 1$ and $C = |g'(\xi)|$.

2. For Newtons Method

$$\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|^2} = \frac{1}{2}\left|\frac{f''(\xi)}{f'(\xi)}\right|, \text{ provided } f'(\xi) \neq 0$$

So order of convergence is 2 and error constant is $\frac{1}{2}\left|\frac{f''(\xi)}{f'(\xi)}\right|$
(if ξ is a double root then $p = 1$)

3. For Secant Method

$$|e_{n+1}| = C_n |e_n||e_{n-1}|$$

$$p = \frac{1+\sqrt{5}}{2} = 1.618\cdots$$

$$\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|^p} = \left|\frac{1}{2}\frac{f''(\xi)}{f'(\xi)}\right|^{1/p}, \text{ provided } f'(\xi) \neq 0$$

4. $\{\frac{1}{n^r}\} \longrightarrow 0$, here $r \geq 1$ $p = 1$.

In theory if order of convergence $p > 1$, then it converges "*fast*" to ξ.

Numerical method to solve Linear system of equations

Suppose we have a system of equations

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}
$$

We want to find a solution

In applications, n is large (at least 1000). So doing it by hand is out of equation. We have to use computers.

It is convenient to use matrices

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Thus we have to solve

$$Ax = b$$

This system either has

1. a unique solution
2. no solution
3. infinitely many solutions

Example

1. $x_1 + x_2 = 2, x_1 - x_2 = 0$ has a unique solution $x_1 = x_2 = 1$.
2. $x_1 + x_2 = 1, 2x_1 + 2x_2 = 3$ has no solution.
3. $2x_1 - x_2 = 3, 4x_1 - 2x_2 = 6$ has infinitely many solutions $\{(x_1, x_2) | 2x_1 - x_2 = 3\}$.

For many applications the system $Ax = b$ has a unique solution.

Theorem

$Ax = b$ has a unique solution if and only if A is an invertible matrix, i.e., there exists a matrix B such that $BA = AB = I_n$, $I_n = n \times n$ identity

$$matrix = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

B is usually denoted by A^{-1}

$$Ax = b$$

$$A^{-1}(Ax) = A^{-1}b$$

So $x = A^{-1}b$ is the unique solution to $Ax = b$

In practice, computation of A^{-1} takes too many computations. Usually we don't need it.

In practice, there is a two step procedure to find solution of $Ax = b$.

Step 1 (Gaussian Elimination)

The system $Ax = b$ is transformed to an equivalence system $U\overline{x} = \overline{b}$, where $U = (u_{ij})$ is an upper triangular matrix $i.e.$, $u_{ij} = 0$ for $i > j$.

$Ax = b$ is equivalent to $U\overline{x} = \overline{b}$ means that x_0 is a solution of $Ax = b$ if and only if x_0 is a solution of $U\overline{x} = \overline{b}$.

Step 2 Solving $U\overline{x} = \overline{b}$.

We do step 2 first

Note that we are assuming $Ax = b$ has a unique solution.

So $U\overline{x} = \overline{b}$ has a unique solution.

$\implies$ U is an invertible matrix.

Exercise:
Show that an upper triangular matrix $U = (u_{ij})$ is invertible iff all diagonal entries $(i.e., u_{ii})$ are non-zero.

Step 2 Solution of $U\overline{x} = \overline{b}$

$$
\begin{aligned}
u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n &= \overline{b_1} \\
u_{22}x_2 + \cdots + u_{2n}x_n &= \overline{b_2} \\
&\vdots \\
u_{n-1,n-1}x_{n-1} + u_{n-1,n}x_n &= \overline{b_{n-1}} \\
u_{nn}x_n &= \overline{b_n}
\end{aligned}
$$

$$
\begin{aligned}
x_n &= \frac{\overline{b_n}}{u_{nn}} \\
x_{n-1} &= \frac{\overline{b_{n-1}} - u_{n-1,n}x_n}{u_{n-1,n-1}} \\
x_i &= \frac{\overline{b_i} - \sum_{j>i} u_{ij}x_j}{u_{ii}} \text{ for } i = n-1, n-2, \cdots, 2, 1
\end{aligned}
$$

This process is called back-substitution.

Example:

$$\begin{aligned} 3x_1 + x_2 + 2x_3 &= 6 \\ 4x_2 + 2x_3 &= 7 \\ 3x_3 &= 9 \end{aligned}$$

$$\begin{aligned} x_3 &= \frac{9}{3} = 3 \\ 4x_2 + 6 &= 7 \implies x_2 = \frac{1}{4} \end{aligned}$$

$$3x_1 + \frac{1}{4} + 6 = 6 \implies x_1 = -\frac{1}{12}$$

Gaussian Elimination

Recall two linear system $Ax = b$ and $\overline{A}x = \overline{b}$ are equivalent if any solution of one is a solution of the other.

Theorem

Let $Ax = b$ be a linear system and suppose we subject this system to a sequence of operation of the following kind

1. *Multiplication of one equation by a non-zero constant*
2. *Addition of a multiple of one equation to another equation*
3. *Interchange of two equations*

If this system produces a new system $\overline{A}x = \overline{b}$ then the system $Ax = b$ and $\overline{A}x = \overline{b}$ are equivalent. In particular, A is invertible iff $\overline{A}$ is invertible.

Gaussian Elimination It is possible to convert $Ax = b$ to equivalent system $Ux = \overline{b}$ (U upper triangular matrix) by using the above 3 operations.

Example

$$\begin{aligned} x_1 - x_2 + 2x_3 &= -6 \\ 2x_1 - 2x_2 + 3x_3 &= -14 \\ x_1 + x_2 + x_3 &= -2 \end{aligned}$$

$$\begin{bmatrix} 1 & -1 & 2 & : & -6 \\ 2 & -2 & 3 & : & -14 \\ 1 & 1 & 1 & : & -2 \end{bmatrix} \xrightarrow{R_2 - 2R_1} \begin{bmatrix} 1 & -1 & 2 & : & -6 \\ 0 & 0 & -1 & : & -2 \\ 1 & 1 & 1 & : & -2 \end{bmatrix}$$

$$\xrightarrow{R_3 - R_1} \begin{bmatrix} 1 & -1 & 2 & : & -6 \\ 0 & \mathbf{0} & -1 & : & -2 \\ 0 & 2 & -1 & : & 4 \end{bmatrix} \xrightarrow{R_2 \longleftrightarrow R_3} \begin{bmatrix} 1 & -1 & 2 & : & -6 \\ 0 & 2 & -1 & : & 4 \\ 0 & 0 & -1 & : & -2 \end{bmatrix}$$

$$\begin{aligned} -x_3 = -2 &\implies x_3 = 2 \\ 2x_2 - x_3 = 4 &\implies x_2 = 3 \\ x_1 - x_2 + 2x_3 = -6 &\implies x_1 = -7 \end{aligned}$$

Algorithm for Gaussian Elimination

To solve $Ax = b$

$W = [A : b]$ "augmented matrix"

Step 1 For $i = 1, 2, \cdots, n-1$ do steps 2,3,4.

Step 2 Let p be the smallest integer with $i \le p \le n$ and $a_{pi} \neq 0$. If no integer p can be found then output "no unique solution exists and stop".

Step 3 If $p \neq i$ then interchange Row $R_i \leftrightarrow$ Row R_j.

Step 4 For $j = i+1, \cdots, n$ do steps 5,6.

Step 5 Set $m_{ji} = \frac{a_{ji}}{a_{ii}}$.

Step 6 perform $R_j - m_{ji} R_i$.

Step 7 If $a_{nn} = 0$ "no unique solution exists and stop".

Step 8 $U =$ first n columns of W. $\overline{b} =$ last column of W.

Then $Ax = b$ is equivalent to $Ux = \overline{b}$ where U is an upper triangular matrix.

Operation Count

We count the number of multiplication/division and addition/subtraction to do GE.

In general the amount of time required to perform a multiplication or division on a computer is approximately the same and is considerably greater than that required to perform addition or subtraction.

No arithmetic operation is performed until step 5 in the algorithm.

Step 5 requires about $n - i$ division be performed.

In step 6 we replace row R_j by $R_j - m_{ji}R_i$. This require m_{ji} to be multiplied to each terms in R_i.

This requires $(n-i)(n-i+1)$ multiplication.

Afterwards each term of the resulting equation is subtracted from the corresponding term in R_j. This requires $(n-i)(n-i+1)$ subtraction.

Thus for each $i = 1, 2, \cdots, n-1$ the operations required are

Multiplication/division: $n-1+(n-i)(n-i+1) = (n-i)(n-i+2)$

Addition/subtraction: $(n-i)(n-i+1)$

Total multiplication/division: $\sum_{i=1}^{n-1}(n-i)(n-i+2) = \frac{2n^3+3n^2-5n}{6}$

Total addition/subtraction: $\sum_{i=1}^{n-1}(n-i)(n-i+1) = \frac{n^3-n}{3}$

For back substitution (*i.e.*, step 2)

One can show one requires

$\frac{n^2+n}{2}$ multiplication/division

$\frac{n^2-n}{2}$ addition/subtraction

Note for large n, n^3 is considerably larger than n^2, for example when $n = 100$, 100^2 is 1% of 100^3.

Thus GE is $\theta(n^3/3)$ operation.

Tridiagonal matrix

$$\begin{bmatrix} a_1 & b_1 & 0 & 0 & \cdots & 0 & 0 \\ c_1 & a_2 & b_2 & 0 & \cdots & 0 & 0 \\ 0 & c_3 & a_3 & b_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & b_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & c_{n-1} & a_n \end{bmatrix}$$

$A = [a_{ij}]$ is said to be tridiagonal if $a_{ij} = 0$ for $|i - j| > 1$

Exercise

$Ax = b$ for tridiagonal systems can be solved in $\theta(n)$ steps.

LU-factorization

Steps used to solve a system $Ax = b$ can be used to factor the matrix A.

The factorization is particularly useful if it has the form $A = LU$ where L is a lower triangular matrix and U is upper triangular matrix.

Not all matrices have this type of representation. However many matrices that occur in practice have this property.

Application:- We would want to solve $Ax = b$ for many different values of b.

If we do GE each time then we would need $\theta(n^3/3)$ operation each time we solve $Ax = b$.

On the other hand once $A = LU$. Then we can solve $Ax = b$ as follows

Set $y = Ux$. We solve $LUx = b, \implies Ly = b$

L is lower triangular. So determining y requires $\theta(n^2)$ operation.

Then solve $Ux = y$ (U is upper triangular), requires only $\theta(n^2)$ operation.

Thus number of operation to solve the system $Ax = b$ is reduced from $\theta(n^3/3)$ to $\theta(2n^2)$.

When $n \geq 1000$ this reduces number of computation by more than 99%.

Construction of LU factorization

Assumption:- $Ax = b$ can be solved without row interchange.

First step in GE process consists of performing for each $j = 2, 3, \cdots, n$ the operations $R_j - m_{j1}R_1$ where $m_{j1} = \frac{a_{j1}^{(1)}}{a_{11}^{(1)}}$.

Equivalently, we can multiply the original matrix A on the left by the matrix

$$M^{(1)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -m_{21} & 1 & 0 & \cdots & 0 \\ -m_{31} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -m_{n-1,1} & 0 & 0 & \cdots & 0 \\ -m_{n1} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Set $A = A^{(1)}$, $b = b^{(1)}$, $A^{(1)}x = b^{(1)}$.

$$M^{(1)}A^{(1)}x = M^{(1)}b^{(1)}.$$

Set $A^{(2)} = M^{(1)}A^{(1)}$, $b^{(2)} = M^{(1)}b^{(1)}$

So we have system $A^{(2)}x = b^{(2)}$

$A^{(2)}$ has $a_{i1}^{(2)} = 0$ for $i \geq 2$.

In a similar manner we construct $M^{(2)}$, the identity matrix with entries below the diagonal in the second column replaced by the negative of the multiple

$$m_{j2} = \frac{a_{j2}^{(2)}}{a_{22}^{(2)}}$$

$$A^{(3)} = M^{(2)}A^{(2)}$$

$$b^{(3)} = M^{(2)}b^{(2)}$$

$A^{(3)}$ has zeros below the diagonal in the first 2 columns.

So we have $A^{(3)}x = b^{(3)}$

In general with $A^{(k)}x = b^{(k)}$ already formed, multiply both sides by

$$M^{(k)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & -m_{k+1,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -m_{n,k} & \cdots & 1 \end{bmatrix}$$

$$A^{(k+1)} = M^{(k)}A^{(k)} = M^{(k)} \cdots M^{(1)}A$$

$$b^{(k+1)} = M^{(k)}b^{(k)} = M^{(k)} \cdots M^{(1)}b$$

So we have $A^{(k+1)}x = b^{(k+1)}$

The process ends with $A^{(n)}x = b^{(n)}$ where

$$A^{(n)} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^{(n)} \end{bmatrix}$$

is upper-triangular matrix.

Set $U = A^{(n)}$

Therefore $= M^{(n-1)}M^{(n-2)}\cdots M^{(1)}A$ and

$$\begin{aligned} L &= \; = [M^{(n-1)}M^{(n-2)}\cdots M^{(1)}]^{-1} \\ &= \; = [M^{(1)}]^{-1}[M^{(2)}]^{-1}\cdots[M^{(n-1)}]^{-1} \end{aligned}$$

Note each $M^{(k)}$ is lower triangular. So $[M^{(k)}]^{-1}$ is lower triangular.
$\implies$ L is lower triangular.

Also $A = LU$.

$$M^{(1)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -m_{21} & 1 & 0 & \cdots & 0 \\ -m_{31} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -m_{n-1,1} & 0 & 0 & \cdots & 0 \\ -m_{n1} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$[M^{(1)}]^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ m_{21} & 1 & 0 & \cdots & 0 \\ m_{31} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1,1} & 0 & 0 & \cdots & 0 \\ m_{n1} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and so on

One can prove that

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ m_{21} & 1 & 0 & \cdots & 0 \\ m_{31} & m_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1,1} & m_{n-1,2} & m_{n-1,3} & \cdots & 0 \\ m_{n1} & m_{n2} & m_{n3} & \cdots & 1 \end{bmatrix}$$

Example:-

$$\begin{aligned} x_1 + x_2 + 0x_3 + 3x_4 &= 4 \\ 2x_1 + x_2 - x_3 + x_4 &= 1 \\ 3x_1 - x_2 + -x_3 + 2x_4 &= -3 \\ -x_1 + 2x_2 + 3x_3 - x_4 &= 4 \end{aligned}$$

Step 1 :- $R_2 - 2R_1,\ R_3 - 3R_1,\ R_4 + R_1$ gives

$$A^{(2)} = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & -4 & -1 & -7 \\ 0 & 3 & 3 & 2 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

Step 2 $R_3 - 4R_2,\ R_4 + 3R_2$ gives

$$A^{(3)} = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} = U$$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix}$$

Solve $Ax = b = \begin{bmatrix} 8 \\ 7 \\ 14 \\ -7 \end{bmatrix}$

$$\text{Set} \qquad y = Ux$$

$$Ly = b$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 14 \\ -7 \end{bmatrix}$$

$$y_1 = 8$$

$$2y_1 + y_2 = 7 \implies y_2 = -9$$

$$3y_1 + 4y_2 + y_3 = 14 \implies y_3 = 26$$

$$-y_1 - 3y_2 + y_4 = -7 \implies y_4 = -26$$

We then solve $Ux = y$

$$\begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ -9 \\ 26 \\ -26 \end{bmatrix}$$

We use "back substitution"

$$\begin{aligned} -13x_4 = -26 &\quad\Longrightarrow\quad x_4 = 2 \\ 3x_3 + 13x_4 = 26 &\quad\Longrightarrow\quad x_3 = 0 \\ -x_2 - x_3 - 5x_4 = -9 &\quad\Longrightarrow\quad x_2 = -1 \\ x_1 + x_2 + 3x_4 = 8 &\quad\Longrightarrow\quad x_1 = 3 \end{aligned}$$