

# Numerical Analysis

(ENME 602)

**Spring 2021** 

# Lecture 7

#### Dr. Hesham H. Ibrahim

Associate Professor, Mechatronics Department. hesham.hamed-ahmed@guc.edu.eg Office C7.04



#### Lecture 7

# **Linear Systems: Iterative Methods**

- 7.1 Successive Over-Relaxation (SOR) Method
- 7.2 Compact Matrix Form
- 7.3 Optimum Relaxation Parameter
- 7.4 Convergence and Error Bounds





#### Example 1

$$2x_1 - x_2 = 0$$

$$-x_1 + 2x_2 - x_3 = 1$$

$$-x_2 + 2x_3 = 2$$

A matrix **A** is <u>diagonally dominated</u> if, in each row, the absolute value of the entry on the diagonal is greater than the sum of the absolute values of the other entries. More compactly, **A** is diagonally dominated if

$$\left|A_{ii}\right| > \sum_{i,i\neq i} \left|A_{ij}\right|$$
 for all  $i$ 

Jacobi-iteration 
$$x_1^k = \frac{1}{2}x_2^{k-1}$$

$$x_2^k = \frac{1}{2}(1 + x_1^{k-1} + x_3^{k-1})$$

$$x_3^k = \frac{1}{2}(2 + x_2^{k-1})$$

Exact Solution  $x = (1, 2, 2)^t$ 



#### Example 1

$$2x_1 - x_2 = 0$$

$$-x_1 + 2x_2 - x_3 = 1$$

$$-x_2 + 2x_3 = 2$$

$$x_{1}^{k} = \frac{1}{2}x_{2}^{k-1}$$
Gauss-Seidel iteration  $x_{2}^{k} = \frac{1}{2}(1 + x_{1}^{k} + x_{3}^{k-1})$ 

$$x_{3}^{k} = \frac{1}{2}(2 + x_{2}^{k})$$

$$x^{o} = (0, 0.5, 1)^{t}$$

$$x^{1} = (0.25, 1.125, 1.563)^{t}$$

$$x^{2} = (0.5625, 1.5625, 1.78125)^{t}$$

$$x = (1, 2, 2)^{t}$$
Exact Solution
$$x = (1, 2, 2)^{t}$$



Jacobi's iteration 
$$\longrightarrow$$
  $x_i^{(k)} = \frac{1}{a_{ii}} \left[ \sum_{\substack{j=1\\j\neq i}}^n \left( -a_{ij} x_j^{(k-1)} \right) + b_i \right], \quad \text{for } i = 1, 2, \dots, n$ 

- The components of  $\mathbf{x}^{(k-1)}$  are used to compute all the components  $x_i^{(k)}$  of  $\mathbf{x}^{(k)}$ .
- But, for i > 1, the components  $x_1^{(k)}, \ldots, x_{i-1}^{(k)}$  of  $\mathbf{x}^{(k)}$  have already been computed and are expected to be better approximations to the actual solutions  $x_1, \ldots, x_{i-1}$  than are  $x_1^{(k-1)}, \ldots, x_{i-1}^{(k-1)}$ .

Gauss-Seidel iteration 
$$\longrightarrow$$
  $x_i^{(k)} = \frac{1}{a_{ii}} \left[ -\sum_{j=1}^{i-1} (a_{ij} x_j^{(k)}) - \sum_{j=i+1}^{n} (a_{ij} x_j^{(k-1)}) + b_i \right]$  for each  $i = 1, 2, \dots, n$ .



Jacobi's iteration 
$$\longrightarrow$$
  $x_i^{(k)} = \frac{1}{a_{ii}} \left[ \sum_{\substack{j=1\\j\neq i}}^n \left( -a_{ij} x_j^{(k-1)} \right) + b_i \right], \quad \text{for } i = 1, 2, \dots, n$ 

Gauss-Seidel iteration

for each i = 1, 2, ..., n.

**SOR** - iteration 
$$\longrightarrow x_i^{(k)} = (1 - \omega)x_i^{(k-1)} + \frac{\omega}{a_{ij}} \left[ b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k-1)} \right]$$

- $\omega$  is a relaxation parameter
- $0 < \omega < 2$  in order to guarantee convergence
- $\omega$  < 1 is called under-relaxed
- $\omega = 1$  reduces to Gauss-Seidel
- $\omega > 1$  is called Over-relaxed



#### Example 1

$$2x_1 - x_2 = 0$$

$$-x_1 + 2x_2 - x_3 = 1$$

$$-x_2 + 2x_3 = 2$$

Gauss-Seidel iteration 
$$x_1^k = \frac{1}{2}x_2^{k-1}$$

$$x_2^k = \frac{1}{2}(1 + x_1^k + x_3^{k-1})$$

$$x_3^k = \frac{1}{2}(2 + x_2^k)$$

$$x_1^k = (1 - \omega)x_1^{k-1} + \omega\left(\frac{1}{2}x_2^{k-1}\right)$$
SOR iteration  $x_2^k = (1 - \omega)x_2^{k-1} + \omega\left(\frac{1}{2}(1 + x_1^k + x_3^{k-1})\right)$ 

$$x_3^k = (1 - \omega)x_3^{k-1} + \omega\left(\frac{1}{2}(2 + x_2^k)\right)$$



#### Example 1

$$2x_1 - x_2 = 0$$

$$-x_1 + 2x_2 - x_3 = 1$$

$$-x_2 + 2x_3 = 2$$

$$x_1^k = (1 - \omega) x_1^{k-1} + \omega \left(\frac{1}{2} x_2^{k-1}\right)$$
SOR iteration
$$x_2^k = (1 - \omega) x_2^{k-1} + \omega \left(\frac{1}{2} \left(1 + x_1^k + x_3^{k-1}\right)\right)$$

$$x_3^k = (1 - \omega) x_3^{k-1} + \omega \left(\frac{1}{2} \left(2 + x_2^k\right)\right)$$

$$x^{o} = (0, 0.5, 1)^{t}$$

$$\omega = 1.2$$

$$x^{1} = (0.3, 1.28, 1.708)^{t}$$

$$x^{2} = (0.708, 1.829, 1.944)^{t}$$



#### Example 2

• The linear system  $A\mathbf{x} = \mathbf{b}$  given by

$$4x_1 + 3x_2 = 24$$
  
 $3x_1 + 4x_2 - x_3 = 30$   
 $-x_2 + 4x_3 = -24$ 

has the solution  $(3, 4, -5)^t$ .

• Compare the iterations from the Gauss-Seidel method and the SOR method with  $\omega = 1.25$  using  $\mathbf{x}^{(0)} = (1, 1, 1)^t$  for both methods.

#### Example 2

For each k = 1, 2, ..., the equations for the Gauss-Seidel method are

$$x_1^{(k)} = -0.75x_2^{(k-1)} + 6$$
  
 $x_2^{(k)} = -0.75x_1^{(k)} + 0.25x_3^{(k-1)} + 7.5$   
 $x_3^{(k)} = 0.25x_2^{(k)} - 6$ 

and the equations for the SOR method with  $\omega =$  1.25 are

$$x_1^{(k)} = -0.25x_1^{(k-1)} - 0.9375x_2^{(k-1)} + 7.5$$

$$x_2^{(k)} = -0.9375x_1^{(k)} - 0.25x_2^{(k-1)} + 0.3125x_3^{(k-1)} + 9.375$$

$$x_3^{(k)} = 0.3125x_2^{(k)} - 0.25x_3^{(k-1)} - 7.5$$

#### Example 2

Gauss-Seidel Iterations										
k	0	1	2	3		7				
$X_1^{(k)}$	1	5.250000	3.1406250	3.0878906		3.0134110				
$X_2^{(k)}$	1	3.812500	3.8828125	3.9267578		3.9888241				
$X_3^{(k)}$	1	-5.046875	-5.0292969	-5.0183105		-5.0027940				

SOR Iterations ( $\omega=$ 1.25)											
k	0	1	2	3		7					
$X_1^{(k)}$	1	6.312500	2.6223145	3.1333027		3.0000498					
$X_2^{(k)}$ $X_3^{(k)}$	1	3.5195313	3.9585266	4.0102646		4.0002586					
$X_3^{(k)}$	1	-6.6501465	-4.6004238	-5.0966863		-5.0003486					

For the iterates to be accurate to 7 decimal places,

- the Gauss-Seidel method requires 34 iterations,
- as opposed to 14 iterations for the SOR method with  $\omega = 1.25$ .



#### Jacobi's Method

 In general, iterative techniques for solving linear systems involve a process that converts the system Ax = b into an equivalent system of the form

$$\mathbf{x} = T\mathbf{x} + \mathbf{c}$$

for some fixed matrix T and vector c.

ullet After the initial vector  $\mathbf{x}^{(0)}$  is selected, the sequence of approximate solution vectors is generated by computing

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$$

for each k = 1, 2, 3, ... (reminiscent of the fixed-point iteration for solving nonlinear equations).



#### Jacobi's Method

The Jacobi method can be written in the form

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$$

by splitting A into its diagonal and off-diagonal parts.

To see this, let D be the diagonal matrix whose diagonal entries are those of A, -L be the strictly lower-triangular part of A, and -U be the strictly upper-triangular part of A where

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



#### Jacobi's Method

We then write A = D - L - U where

$$D = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} \qquad L = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 \end{bmatrix}$$

and

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



#### Jacobi's Method

The equation  $A\mathbf{x} = \mathbf{b}$ , or  $(D - L - U)\mathbf{x} = \mathbf{b}$ , is then transformed into

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

and, if  $D^{-1}$  exists, that is, if  $a_{ii} \neq 0$  for each i, then

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

This results in the matrix form of the Jacobi iterative technique:

$$\mathbf{x}^{(k)} = D^{-1}(L+U)\mathbf{x}^{(k-1)} + D^{-1}\mathbf{b}, \quad k = 1, 2, \dots$$



#### Jacobi's Method

Introducing the notation  $T_j = D^{-1}(L + U)$  and  $\mathbf{c}_j = D^{-1}\mathbf{b}$  gives the Jacobi technique the form

$$\mathbf{x}^{(k)} = T_j \mathbf{x}^{(k-1)} + \mathbf{c}_j$$

In practice, this form is only used for theoretical purposes while

$$x_i^{(k)} = \frac{1}{a_{ii}} \left[ \sum_{\substack{j=1 \ j \neq i}}^n \left( -a_{ij} x_j^{(k-1)} \right) + b_i \right], \quad \text{for } i = 1, 2, \dots, n$$

is used in computation.



#### Example 3

Express the Jacobi iteration method for the linear system  $A\mathbf{x} = \mathbf{b}$  given by

$$E_1: 10x_1 - x_2 + 2x_3 = 6$$
  
 $E_2: -x_1 + 11x_2 - x_3 + 3x_4 = 25$   
 $E_3: 2x_1 - x_2 + 10x_3 - x_4 = -11$ 

$$E_4$$
:  $3x_2 - x_3 + 8x_4 = 15$ 

in the form  $\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$ .

#### Example 3

We saw earlier that the Jacobi method for this system has the form

$$x_{1} = \frac{1}{10}x_{2} - \frac{1}{5}x_{3} + \frac{3}{5}$$

$$x_{2} = \frac{1}{11}x_{1} + \frac{1}{11}x_{3} - \frac{3}{11}x_{4} + \frac{25}{11}$$

$$x_{3} = -\frac{1}{5}x_{1} + \frac{1}{10}x_{2} + \frac{1}{10}x_{4} - \frac{11}{10}$$

$$x_{4} = -\frac{3}{8}x_{2} + \frac{1}{8}x_{3} + \frac{15}{8}$$

Hence, we have

$$T = \begin{bmatrix} 0 & \frac{1}{10} & -\frac{1}{5} & 0\\ \frac{1}{11} & 0 & \frac{1}{11} & -\frac{3}{11}\\ -\frac{1}{5} & \frac{1}{10} & 0 & \frac{1}{10}\\ 0 & -\frac{3}{8} & \frac{1}{8} & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \frac{3}{5}\\ \frac{25}{11}\\ -\frac{11}{10}\\ \frac{15}{8} \end{bmatrix}$$

$$\mathbf{x}^{(k)} = D^{-1}(L+U)\mathbf{x}^{(k-1)} + D^{-1}\mathbf{b}, \quad k = 1, 2, \dots$$



#### Gauss-Seidel Method

#### Writing all *n* equations gives

$$a_{11}x_1^{(k)} = -a_{12}x_2^{(k-1)} - a_{13}x_3^{(k-1)} - \cdots - a_{1n}x_n^{(k-1)} + b_1$$

$$a_{21}x_1^{(k)} + a_{22}x_2^{(k)} = -a_{23}x_3^{(k-1)} - \cdots - a_{2n}x_n^{(k-1)} + b_2$$

$$\vdots$$

$$a_{n1}x_1^{(k)} + a_{n2}x_2^{(k)} + \cdots + a_{nn}x_n^{(k)} = b_n$$

With the definitions of D, L, and U given previously, we have the Gauss-Seidel method represented by

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



#### Gauss-Seidel Method

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

# Re-Writing the Equations (Cont'd)

Solving for  $\mathbf{x}^{(k)}$  finally gives

$$\mathbf{x}^{(k)} = (D-L)^{-1}U\mathbf{x}^{(k-1)} + (D-L)^{-1}\mathbf{b}$$
, for each  $k = 1, 2, ...$ 

Letting  $T_g = (D - L)^{-1}U$  and  $\mathbf{c}_g = (D - L)^{-1}\mathbf{b}$ , gives the Gauss-Seidel technique the form

$$\mathbf{x}^{(k)} = T_g \mathbf{x}^{(k-1)} + \mathbf{c}_g$$

For the lower-triangular matrix D-L to be nonsingular, it is necessary and sufficient that  $a_{ii} \neq 0$ , for each i = 1, 2, ..., n.



#### **SOR Method**

#### The SOR Method

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

#### Letting

$$T_{\omega} = (D - \omega L)^{-1}[(1 - \omega)D + \omega U]$$

and 
$$\mathbf{c}_{\omega} = \omega (D - \omega L)^{-1} \mathbf{b}$$

gives the SOR technique the form

$$\mathbf{x}^{(k)} = T_{\omega}\mathbf{x}^{(k-1)} + \mathbf{c}_{\omega}$$



- An obvious question to ask is how the appropriate value of  $\omega$  is chosen when the SOR method is used?
- Although no complete answer to this question is known for the general n × n linear system, the following results can be used in certain important situations.

#### Theorem (Kahan)

If  $a_{ii} \neq 0$ , for each i = 1, 2, ..., n, then  $\rho(T_{\omega}) \geq |\omega - 1|$ . This implies that the SOR method can converge only if  $0 < \omega < 2$ .

#### Theorem (Ostrowski-Reich)

If A is a positive definite matrix and  $0 < \omega < 2$ , then the SOR method converges for any choice of initial approximate vector  $\mathbf{x}^{(0)}$ .



#### Theorem (Kahan)

If  $a_{ii} \neq 0$ , for each i = 1, 2, ..., n, then  $\rho(T_{\omega}) \geq |\omega - 1|$ . This implies that the SOR method can converge only if  $0 < \omega < 2$ .

#### Theorem (Ostrowski-Reich)

If A is a positive definite matrix and  $0 < \omega < 2$ , then the SOR method converges for any choice of initial approximate vector  $\mathbf{x}^{(0)}$ .

#### Theorem

If A is positive definite and tridiagonal, then  $\rho(T_g) = [\rho(T_j)]^2 < 1$ , and the optimal choice of  $\omega$  for the SOR method is

$$\omega = \frac{2}{1 + \sqrt{1 - [\rho(T_j)]^2}}$$

With this choice of  $\omega$ , we have  $\rho(T_{\omega}) = \omega - 1$ .



#### Example 4

Find the optimal choice of  $\omega$  for the SOR method for the matrix

$$A = \left[ \begin{array}{ccc} 4 & 3 & 0 \\ 3 & 4 & -1 \\ 0 & -1 & 4 \end{array} \right]$$

- This matrix is clearly tridiagonal, so we can apply the result in the SOR theorem if we can also show that it is positive definite.
- Because the matrix is symmetric, the theory tells us that it is positive definite if and only if all its leading principle submatrices has a positive determinant.
- This is easily seen to be the case because

$$det(A) = 24$$
,  $det\left(\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}\right) = 7$  and  $det([4]) = 4$ 



#### Example 4

Find the optimal choice of  $\omega$  for the SOR method for the matrix

$$A = \left[ \begin{array}{ccc} 4 & 3 & 0 \\ 3 & 4 & -1 \\ 0 & -1 & 4 \end{array} \right]$$

We compute

$$T_{j} = D^{-1}(L+U)$$

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

$$= \begin{bmatrix} 0 & -0.75 & 0 \\ -0.75 & 0 & 0.25 \\ 0 & 0.25 & 0 \end{bmatrix}$$

so that

$$T_{j} - \lambda I = \begin{bmatrix} -\lambda & -0.75 & 0 \\ -0.75 & -\lambda & 0.25 \\ 0 & 0.25 & -\lambda \end{bmatrix}$$



#### Example 4

#### Therefore

$$\det(T_j - \lambda I) = \begin{vmatrix} -\lambda & -0.75 & 0 \\ -0.75 & -\lambda & 0.25 \\ 0 & 0.25 & -\lambda \end{vmatrix} = -\lambda(\lambda^2 - 0.625)$$

Thus

$$\rho(T_j) = \sqrt{0.625}$$

and

$$\omega = \frac{2}{1 + \sqrt{1 - [\rho(T_j)]^2}} = \frac{2}{1 + \sqrt{1 - 0.625}} \approx 1.24.$$



#### Theorem

For any  $\mathbf{x}^{(0)} \in \mathbb{R}^n$ , the sequence  $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$  defined by

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$$
, for each  $k \ge 1$ 

converges to the unique solution of

$$\mathbf{x} = T\mathbf{x} + \mathbf{c}$$

if and only if  $\rho(T) < 1$ .

# Corollary

||T|| < 1 for any natural matrix norm and  $\mathbf{c}$  is a given vector, then the sequence  $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$  defined by

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$$

converges, for any  $\mathbf{x}^{(0)} \in \mathbb{R}^n$ , to a vector  $\mathbf{x} \in \mathbb{R}^n$ , with  $\mathbf{x} = T\mathbf{x} + \mathbf{c}$ , and the following error bounds hold:

(i) 
$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \le \|T\|^k \|\mathbf{x}^{(0)} - \mathbf{x}\|$$

(ii) 
$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \le \frac{\|T\|^k}{1 - \|T\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|$$

# Using the Matrix Formulations

We have seen that the Jacobi and Gauss-Seidel iterative techniques can be written

$$\mathbf{x}^{(k)} = T_j \mathbf{x}^{(k-1)} + \mathbf{c}_j$$
 and  $\mathbf{x}^{(k)} = T_g \mathbf{x}^{(k-1)} + \mathbf{c}_g$ 

using the matrices

$$T_j = D^{-1}(L + U)$$
 and  $T_g = (D - L)^{-1}U$ 

respectively. If  $\rho(T_j)$  or  $\rho(T_g)$  is less than 1, then the corresponding sequence  $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$  will converge to the solution  $\mathbf{x}$  of  $A\mathbf{x} = \mathbf{b}$ .



The following are easily verified sufficiency conditions for convergence of the Jacobi and Gauss-Seidel methods.

#### **Theorem**

If A is strictly diagonally dominant, then for any choice of  $\mathbf{x}^{(0)}$ , both the Jacobi and Gauss-Seidel methods give sequences  $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$  that converge to the unique solution of  $A\mathbf{x} = \mathbf{b}$ .



# Thank You



"Numerical Differentiation and Integration"

