Numerical Analysis

Solving Linear Systems of Equations

OL Say

ol.say@itc.edu.kh

Institute of Technology of Cambodia

May 1, 2023

2. Outline

1. A system of algebraic equations has the form

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

2. In matrix notation the equations are written as

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

or simply $AX = B$.

3. A particularly useful representation of the equations for computational purposes is the augmented coefficient matrix obtained by adjoining the constant vector B to the coefficient matrix A in the following fashion:

$$(A|B) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right)$$

4. In the first few sections, we discuss about direct methods for solving the system. The three popular methods are listed below

Method	Initial form	Final form
Gauss elimination	$AX = B$	$UX = C$
LU decomposition	$AX = B$	$LUX = B$
Gauss-Jordan elimination	$AX = B$	$IX = C$

1. U represents an upper triangular matrix

$$U = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & u_{nn} \end{pmatrix}$$

2. L is a lower triangular matrix

$$L = \begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{pmatrix}$$

3. And $A = LU$.

2. Gauss Elimination and Backward Substitution

Example 1

Represent the linear system

$$\begin{cases} x_1 - x_2 + 2x_3 - x_4 = -8, \\ x_1 - x_2 + 2x_3 - x_4 = -8, \\ x_1 + x_2 + x_3 = -2, \\ x_1 - x_2 + 4x_3 + 3x_4 = 4, \end{cases}$$

as an augmented matrix and use Gaussian Elimination to find its solution.

$$A_0 = (A|B) = \left(\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -1 & 4 & 3 & 4 \end{array}\right)$$

$$A_1^{(0)} = \left(\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right)$$

$$A_1^{(1)} = \left(\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right)$$

$$A_2 = \left(\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right)$$

Use the backward substitution from the result of Gaussian elimination

$$
\begin{aligned}
x_4 &= \frac{4}{2} = 2,\\
x_3 &= \frac{-4-(-1)x_4}{-1} = 2,\\
x_2 &= \frac{6-x_4-(-1)x_3}{2} = 3,\\
x_1 &= \frac{-8-(-1)x_4-2x_3-(-1)x_2}{1} = -7.
\end{aligned}
$$

The system has a unique solution $X = (-7, 3, 2, 2)^T$.

$$
\begin{array}{c|cccccc}
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & & a_{k,k} & a_{k,k+1} & a_{k,k+2} & \cdots & a_{k,n} \\
i = k+1 & & a_{k+1,k} & a_{k+1,k+1} & a_{k+1,k+2} & \cdots & a_{k+1,n} \\
i = k+2 & (0) & a_{k+2,k} & a_{k+2,k+1} & a_{k+2,k+2} & \cdots & a_{k+2,n} \\
\vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
i = n & & a_{n,k} & a_{n,k+1} & a_{n,k+2} & \cdots & a_{n,n} \\
\hline
 & & j = k & j = k+1 & j = k+2 & \cdots & j = n
\end{array}
$$

Gaussian Elimination and Backward Substitution

To solve the $n \times n$ linear system $AX = B$.

INPUT ??

OUTPUT ??

Body ??

1. It is possible to show that any square matrix A can be expressed as a product of a lower triangular matrix L and an upper triangular matrix U : $A = LU$.
2. The process of computing L and U for a given A is known as *LU decomposition* or *LU factorization*.
3. LU decomposition is not unique (the combinations of L and U for a prescribed A are endless), unless certain constraints are placed on L or U.
4. These constraints distinguish one type of decomposition from another.
5. Three commonly used decompositions are listed below

Name	Constraints
Doolittle LU decomposition	$l_{ii} = 1, \ i = 1, 2, \ldots, n$
Doolittle LDLt decomposition	$U = DL^T, l_{ii} = 1, \ i = 1, 2, \ldots, n$
Cholesky LLt decomposition	$U = L^T$
Crout LU decomposition	$u_{ii} = 1, \ i = 1, 2, \ldots, n$

1. Consider a 3×3 matrix A and assume that there exist triangular matrices

$$L = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}, \quad \text{such that } A = LU.$$

2. After completing the multiplication on the right-hand side, we get

$$A = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{11}l_{21} & u_{12}l_{21} + u_{22} & u_{13}l_{21} + u_{23} \\ u_{11}l_{31} & u_{12}l_{31} + u_{22}l_{32} & u_{13}l_{31} + u_{23}l_{32} + u_{33} \end{pmatrix}$$

3. $R_2 \leftarrow R_2 - l_{21}R_1$ (eliminates A_{21})
   $R_3 \leftarrow R_3 - l_{31}R_1$ (eliminates A_{31})

$$A_1 = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & u_{22}l_{32} & u_{23}l_{32} + u_{33} \end{pmatrix}$$

4. $R_3 \leftarrow R_3 - l_{32}R_2$

$$A_2 = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

Example 2

Use Doolittle's decomposition method to solve the equations $AX = B$, where

$$A = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 6 & -1 \\ 2 & -1 & 2 \end{pmatrix}, B = \begin{pmatrix} 7 \\ 13 \\ 5 \end{pmatrix}$$

1. Decompose A into LU by Gaussian Elimination method.
2. Solve $LY = B$ for Y by Forward Substitution method where $Y = UX$.
3. Solve $UX = Y$ for X by Backward Substitution method.

Doolittle's Decomposition Algorithm

Decompose an $n \times n$ square matrix A into LU with $L_{ii} = 1$ for $i = 1, 2, \dots, n$.

INPUT Square matrix A of size $n \times n$

OUTPUT Lower triangular matrix L and upper triangular matrix U

1. Set $U \leftarrow A$
2. For $i = 1$ to n set $l_{ii} \leftarrow 1$
3. For $k = 1$ to $n - 1$ do
    a. For $i = k + 1$ to n do
        i. Set $r \leftarrow u_{ik}/u_{kk}$; $l_{ik} \leftarrow r$; $u_{ik} \leftarrow 0$
        ii. For $j = k + 1$ to n set $u_{ij} \leftarrow u_{ij} - r \cdot u_{kj}$
4. OUTPUT L and U

Definition 3 (Positive Definite)

A matrix A is positive definite if it is *symmetric* and if $X^T AX > 0$ for every n-dimensional vector $X \neq 0$.

Example 4

A symmetric matrix $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ is positive definite because

$$\begin{aligned} X^T AX &= (x_1, x_2, x_3) \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= x_1^2 + (x_1 + x_2)^2 + (x_2 + x_3)^2 + x_3^2 > 0, \text{ for all } X \neq 0. \end{aligned}$$

Definition 5 (Leading Pricipal Submatrix)

A leading principal submatrix of a matrix A is a matrix of the form

$$A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

for some $1 \leq k \leq n$.

Theorem 6

A symmetric matrix A is positive definite if and only if each of its leading principal submatrices has a positive determinant.

Example 7

A symmetric matrix $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ is positive definite because

$$A_1 = \det\begin{pmatrix} 2 \end{pmatrix} = 2 > 0$$

$$A_2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3 > 0$$

$$A_3 = \det\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} = 4 > 0.$$

3.2 Doolittle LDLt Decomposition

Theorem 8

The symmetric matrix A is positive definite if and only if Gaussian elimination without row interchanges can be performed on the linear system $AX = B$ with all pivot elements positive. Moreover, in this case, the computations are stable with respect to the growth of round-off errors.

Corollary 9 (Doolittle LDLt Decomposition)

The matrix A is positive definite if and only if A can be factored in the form LDL^T, where L is lower triangular with 1's on its diagonal and D is a diagonal matrix with positive diagonal entries.

Corollary 10 (Cholesky LLt Decomposition Existence)

The matrix A is positive definite if and only if A can be factored in the form LL^T, where L is lower triangular with nonzero diagonal entries.

1. Consider a 4×4 symmetric positive definite matrix A, a lower triangular matrix L and a diagonal matrix D such that $A = LDL^T$ where L, D and A are denoted respectively as the following

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{pmatrix}, \begin{pmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{pmatrix}, \begin{pmatrix} a_{11} & & & (\text{sym}) \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

2. The equality $L(DL^T) = A$ can be displayed as

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{pmatrix} \begin{pmatrix} d_1 & d_1 l_{21} & d_1 l_{31} & d_1 l_{41} \\ 0 & d_2 & d_2 l_{32} & d_2 l_{42} \\ 0 & 0 & d_3 & d_3 l_{43} \\ 0 & 0 & 0 & d_4 \end{pmatrix} = \begin{pmatrix} a_{11} & & & (\text{s}) \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

3. After completing the matrix multiplication on the left-hand side, equate the elements in each row
4. First row $d_1 = a_{11}$

1. Second row

$$d_1 l_{21} = a_{21} \Rightarrow l_{21} = a_{21}/d_1$$

$$d_1 l_{21}^2 + d_2 = a_{22} \Rightarrow d_2 = a_{22} - d_1 l_{21}^2$$

2. Third row

$$d_1 l_{31} = a_{31} \Rightarrow l_{31} = a_{31}/d_1$$

$$d_1 l_{31} l_{21} + d_2 l_{32} = a_{32} \Rightarrow l_{32} = (a_{32} - d_1 l_{31} l_{21})/d_2$$

$$d_1 l_{31}^2 + d_2 l_{32}^2 + d_3 = a_{33} \Rightarrow d_3 = a_{33} - d_1 l_{31}^2 - d_2 l_{32}^2$$

3. Fourth row

$$d_1 l_{41} = a_{41} \Rightarrow l_{41} = a_{41}/d_1$$

$$d_1 l_{41} l_{21} + d_2 l_{42} = a_{42} \Rightarrow l_{42} = (a_{42} - d_1 l_{41} l_{21})/d_2$$

$$d_1 l_{41} l_{31} + d_2 l_{42} l_{32} + d_3 l_{43} = a_{43} \Rightarrow l_{43} = (a_{42} - d_1 l_{41} l_{31} - d_2 l_{42} l_{32})/d_3$$

$$d_1 l_{41}^2 + d_2 l_{42}^2 + d_3 l_{43}^2 + d_4 = a_{44} \Rightarrow d_4 = a_{44} - d_1 l_{41}^2 - d_2 l_{42}^2 - d_3 l_{43}^2$$

3.2 Doolittle LDLt Decomposition

Doolittle LDLt Decomposition Algorithm

Decompose an $n \times n$ symmetric positive definite matrix A into LDL^T, where L is a lower triangular matrix with 1's along the diagonal and D is a diagonal matrix with positive entries on thediagonal:

INPUT the dimension n; entries a_{ij}, for $1 \leq i, j \leq n$ of A.

OUTPUT the entries l_{ij}, for $1 \leq j \leq i$ and $1 \leq i \leq n$ of L and d_i for $1 \leq i \leq n$.

1. Set $d_1 = a_{11};\ l_{21} = a_{21}/d_1;\ d_2 = a_{22} - d_1 l_{21}^2$
2. For $i = 3$ to n do
   - a. Set $l_{i1} = a_{i1}/d_1$
   - b. For $j = 2$ to $i - 1$ set $l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} d_k l_{ik} l_{jk})/d_j$
   - c. Set $d_i = a_{ii} - \sum_{k=1}^{i-1} d_k l_{ik}^2$
3. OUTPUT l_{ij} for $1 \leq j \leq i$ and $1 \leq i \leq n$ and d_i for $1 \leq i \leq n$

1. Consider a 4×4 symmetric positive definite matrix A and a lower triangular matrix L such that

$$\begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{43} \\ 0 & 0 & 0 & l_{44} \end{pmatrix} = \begin{pmatrix} a_{11} & & & (\text{sym}) \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

2. After completing the matrix multiplication on the left-hand side, equate the elements in each row
3. First row $l_{11}^2 = a_{11} \Rightarrow l_{11} = \sqrt{a_{11}}$
4. Second row

$$l_{21}l_{11} = a_{21} \Rightarrow l_{21} = a_{21}/l_{11}$$

$$l_{21}^2 + l_{22}^2 = a_{22} \Rightarrow l_{22} = \sqrt{a_{22} - l_{21}^2}$$

5. Third row

$$l_{31}l_{11} = a_{31} \Rightarrow l_{31} = a_{31}/l_{11}$$
$$l_{31}l_{21} + l_{32}l_{22} = a_{32} \Rightarrow l_{32} = (a_{31} - l_{31}l_{21})/l_{22}$$
$$l_{31}^2 + l_{32}^2 + l_{33}^2 = a_{33} \Rightarrow l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2}$$

6. Fourth row

$$l_{41}l_{11} = a_{41} \Rightarrow l_{41} = a_{41}/l_{11}$$
$$l_{41}l_{21} + l_{42}l_{22} = a_{42} \Rightarrow l_{42} = (a_{42} - l_{41}l_{21})/l_{22}$$
$$l_{41}l_{31} + l_{42}l_{32} + l_{43}l_{33} = a_{43} \Rightarrow l_{43} = (a_{43} - l_{41}l_{31} - l_{42}l_{32})/l_{33}$$
$$l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 = a_{44} \Rightarrow l_{44} = \sqrt{a_{44} - l_{41}^2 - l_{42}^2 - l_{43}^2}$$

Cholesky LLt Decomposition Algorithm

Decompose an $n \times n$ positive definite, symmetric matrix A into LL^T
INPUT the dimension n; entries a_{ij}, for $1 \le i, j \le n$ of A.
OUTPUT the entries l_{ij}, for $1 \le j \le i$ and $1 \le i \le n$ of L.
(The entries of $U = L^T$ is $u_{ij} = l_{ji}$, for $i \le j \le n$ and $1 \le i \le n$.)

1. Set $l_{11} = \sqrt{a_{11}}$; $l_{21} = a_{21}/l_{11}$; $l_{22} = \sqrt{a_{22} - l_{21}^2}$
2. For $i = 3$ to n do
   a. Set $l_{i1} = a_{i1}/l_{11}$
   b. For $j = 2$ to $i - 1$ set $l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk})/l_{jj}$
   c. Set $l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2}$
3. OUTPUT l_{ij} for $1 \le j \le i$ and $1 \le i \le n$.

Example 11

Use Cholesky's decomposition method to solve the equations $AX = B$, where

$$A = \begin{pmatrix} 4 & -2 & 2 \\ -2 & 2 & -4 \\ 2 & -4 & 11 \end{pmatrix}, B = \begin{pmatrix} 8 \\ 0 \\ -9 \end{pmatrix}$$

1. By Cholesky decomposition,

$$L = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -3 & 1 \end{pmatrix}, L^T = \begin{pmatrix} 2 & -1 & 1 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}$$

2. $LY = B \Rightarrow Y = (4, 4, -1)^T$
3. $UX = Y \Rightarrow X = (3, 1, -1)^T$.

3.4 Crout's Decomposition

Follow the Doolittle's Decomposition method to derive the Crout's Decomposition method.

Crout's Decomposition

Decompose an $n \times n$ square matrix A into LU with $U_{ii} = 1$ for $i = 1, 2, \ldots, n$.
INPUT ??
OUTPUT ??
Body ??

4. Gauss-Jordan Elimination

1. The Gauss-Jordan method is essentially Gauss elimination taken to its limit.
2. In the Gauss elimination method only the equations that lie below the pivot equation are transformed.
3. In the Gauss-Jordan method the elimination is also carried out on equations above the pivot equation, resulting in a diagonal coefficient matrix.
4. The main disadvantage of Gauss-Jordan elimination is that it involves about $n^3/2$ long operations, which is 1.5 times the number required in Gauss elimination.
5. For this reason, we will not discuss the method into more detail.

1. Engineering problems often lead to coefficient matrices that are *sparsely populated,* meaning that most elements of the matrix are zero.
2. If all the nonzero terms are clustered about the leading diagonal, then the matrix is said to be *banded.*
3. Tridiagonal and pentadiagonal matrices are examples of banded matrices.

$$\begin{pmatrix} d_1 & e_1 & 0 & 0 & 0 & \cdots & 0 \\ c_1 & d_2 & e_2 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & d_3 & e_3 & 0 & \cdots & 0 \\ 0 & 0 & c_3 & d_4 & e_4 & \cdots & 0 \\ 0 & 0 & 0 & c_4 & d_5 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & e_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & c_{n-1} & d_n \end{pmatrix}, \begin{pmatrix} d_1 & e_1 & f_1 & 0 & 0 & \cdots & 0 \\ c_1 & d_2 & e_2 & f_2 & 0 & \cdots & 0 \\ b_1 & c_2 & d_3 & e_3 & f_3 & \cdots & 0 \\ 0 & b_2 & c_3 & d_4 & e_4 & \ddots & \vdots \\ 0 & 0 & b_3 & c_4 & d_5 & \ddots & f_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & e_{n-1} \\ 0 & 0 & 0 & \cdots & b_{n-2} & c_{n-1} & d_n \end{pmatrix}$$

Consider the Tridiagonal matrix mentioned above. Let us now apply LU decomposition to the coefficient matrix.

1. We reduce row k by getting rid of c_{k-1} with the elementary operation $R_k \leftarrow R_k - \lambda R_{k-1},\ \lambda = c_{k-1}/d_{k-1},\ k = 2, 3, \ldots, n.$
2. The corresponding change in d_k is $d_k = d_k - \lambda e_{k-1}$ whereas e_k is not affected.
3. To finish up with Doolittle's decomposition of the form $(L\backslash U)$, we store the multiplier $\lambda = c_{k-1}/d_{k-1}$ in the location previously occupied by c_{k-1} : $c_{k-1} \leftarrow \lambda$.
4. The resulting factors L and U are

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \cdots & 0 \\
c_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\
0 & c_2 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & c_3 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & c_4 & 1 & \ddots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\
0 & 0 & 0 & 0 & \cdots & c_{n-1} & 1
\end{pmatrix},
\begin{pmatrix}
d_1 & e_1 & 0 & 0 & 0 & \cdots & 0 \\
0 & d_2 & e_2 & 0 & 0 & \cdots & 0 \\
0 & 0 & d_3 & e_3 & 0 & \cdots & 0 \\
0 & 0 & 0 & d_4 & e_4 & \cdots & 0 \\
0 & 0 & 0 & 0 & d_5 & \ddots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & e_{n-1} \\
0 & 0 & 0 & 0 & \cdots & 0 & d_n
\end{pmatrix}
$$

where the values of c_i and d_i are modified by the Doolittle's decomposition above.

4. Solve $LY = B$ for Y by Forward Substitution method

$$y_1 = b_1, \; y_i = b_i - c_{i-1} * y_{i-1}, \; i = 2, 3, \dots, n.$$

5. Solve $UX = Y$ for X by Forward Substitution method

$$x_n = y_n/d_n, \; x_i = (y_i - e_i x_{i+1})/d_i, \; i = n-1, n-2, \dots, 1.$$

Doolittle's Decomposition for Tridiagonal Matrix

Decompose an $n \times n$ square tridiagonal matrix A into LU with $L_{ii} = 1$ for $i = 1, 2, \dots, n$.

INPUT ??

OUTPUT ??

Body ??

Example 12 (LU Decomposition)

Determine L and U that result from Doolittle's decomposition of a tridiagonal matrix A and solve for X of $AX = B$ where

$$A = \begin{pmatrix} 1 & -2 & 0 \\ 2 & -2 & 3 \\ 0 & 6 & 7 \end{pmatrix}, B = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}$$

1. Apply Gauss elimination $R_2 \leftarrow R_2 - (2)R_1$,

$$A \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & 3 \\ 0 & 6 & 7 \end{pmatrix}, L \to \begin{pmatrix} 1 & 0 & 0 \\ \mathbf{2} & 1 & 0 \\ 0 & ? & 1 \end{pmatrix}, U \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & ? \end{pmatrix}$$

2. $R_3 \leftarrow R_3 - (3)R_2$,

$$A \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & -2 \end{pmatrix}, L \to \begin{pmatrix} 1 & 0 & 0 \\ \mathbf{2} & 1 & 0 \\ 0 & \mathbf{3} & 1 \end{pmatrix}, U \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & -2 \end{pmatrix}$$

3. Therefore,

$$A = LU,\ L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}, U = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & -2 \end{pmatrix}$$

4. Solve $LY = B \Rightarrow Y = (0, -1, 2)^T$
5. Solve $UX = Y \Rightarrow X = (2, 1, -1)^T$.

1. We encounter pentadiagonal (bandwidth = 5) coefficient matrices in the solution of fourth-order, ordinary differential equations by finite differences.
2. Often these matrices are symmetric, in which case an $n \times n$ coefficient matrix has the form

$$\begin{pmatrix} d_1 & e_1 & f_1 & 0 & 0 & \cdots & 0 \\ e_1 & d_2 & e_2 & f_2 & 0 & \cdots & 0 \\ f_1 & e_2 & d_3 & e_3 & f_3 & \cdots & 0 \\ 0 & f_2 & e_3 & d_4 & e_4 & \ddots & \vdots \\ 0 & 0 & f_3 & e_4 & d_5 & \ddots & f_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & e_{n-1} \\ 0 & 0 & 0 & \cdots & f_{n-2} & e_{n-1} & d_n \end{pmatrix}$$

3. Let us now look at the solution of the equations $AX = B$ by Doolittle's decomposition.
4. The first step is to transform A to upper triangular form by Gauss elimination.
5. If elimination has progressed to the stage where the k-th row has become the pivot row, we have the following situation

$$
A \longrightarrow \left(\begin{array}{cc|ccc|cccc}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\hline
\cdots & 0 & d_k & e_k & f_k & 0 & 0 & 0 & \cdots \\
\cdots & 0 & e_k & d_{k+1} & e_{k+1} & f_{k+1} & 0 & 0 & \cdots \\
\cdots & 0 & f_k & e_{k+1} & d_{k+2} & e_{k+2} & f_{k+2} & 0 & \cdots \\
\hline
\cdots & 0 & 0 & f_{k+1} & e_{k+2} & d_{k+3} & e_{k+3} & f_{k+3} & \cdots \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right)
$$

6. The elements e_k and f_k below the pivot row (the k-th row) are eliminated by the operations

$$R_{k+1} \leftarrow R_{k+1} - \lambda_1 R_k, \ \lambda_1 = e_k/d_k$$
$$R_{k+2} \leftarrow R_{k+2} - \lambda_2 R_k, \ \lambda_2 = f_k/d_k$$

7. The only terms (other than those being eliminated) that are changed by the operations are

$$d_{k+1} \leftarrow d_{k+1} - \lambda_1 e_k, \ \lambda_1 = e_k/d_k$$
$$e_{k+1} \leftarrow d_{k+1} - \lambda_1 f_k, \ \lambda_1 = e_k/d_k$$
$$d_{k+2} \leftarrow d_{k+2} - \lambda_2 f_k, \ \lambda_2 = f_k/d_k$$

8. Storage of the multipliers in the upper triangular portion of the matrix results in

$$e_k \leftarrow \lambda_1 = e_k/d_k$$
$$f_k \leftarrow \lambda_2 = f_k/d_k$$

9. Apply above iterations for $k = 1, 2, n-2$, the matrix has the form (do not confuse d, e, and f with the original contents of A)

$$U^* = \begin{pmatrix} d_1 & e_1 & f_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & e_2 & f_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & e_3 & f_3 & \cdots & 0 \\ 0 & 0 & 0 & d_4 & e_4 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & \ddots & f_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & d_{n-1} & e_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & e_{n-1} & d_n \end{pmatrix}$$

10. One last step

$$\lambda_1 \leftarrow e_{n-1}/d_{n-1}$$
$$d_n \leftarrow d_n - \lambda_1 e_{n-1}$$
$$e_{n-1} \leftarrow \lambda_1.$$

11. Now comes the solution phase. The equations $LY = B$ have the augmented coefficient matrix

$$(L|B) = \left(\begin{array}{ccccccc|c} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & b_1 \\ e_1 & 1 & 0 & 0 & 0 & \cdots & 0 & b_2 \\ f_1 & e_2 & 1 & 0 & 0 & \cdots & 0 & b_3 \\ 0 & f_2 & e_3 & 1 & 0 & \ddots & \vdots & b_4 \\ 0 & 0 & f_3 & e_4 & 1 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 & b_{n-1} \\ 0 & 0 & 0 & \cdots & f_{n-2} & e_{n-1} & 1 & b_n \end{array}\right)$$

12. Solution by forward substitution yields

$$\begin{aligned}
y_1 &= b_1 \\
y_2 &= b_2 - e_1 y_1 \\
&\vdots \\
y_k &= b_k - e_{k-1} y_{k-1} - f_{k-2} y_{k-2}, \quad k = 3, 4, \ldots, n
\end{aligned}$$

13. The equations to be solved by back substitution, namely $UX = Y$, have the augmentedcoefficient matrix

$$
(U|Y) = \left(\begin{array}{ccccccc|c}
d_1 & d_1e_1 & d_1f_1 & 0 & 0 & \cdots & 0 & y_1 \\
0 & d_2 & d_2e_2 & d_2f_2 & 0 & \cdots & 0 & y_2 \\
0 & 0 & d_3 & d_3e_3 & d_3f_3 & \cdots & 0 & y_3 \\
0 & 0 & 0 & d_4 & d_4f_4 & \ddots & \vdots & y_4 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & d_{n-2}f_{n-2} & \vdots \\
0 & 0 & 0 & \cdots & 0 & d_{n-1} & d_{n-1}e_{n-1} & y_{n-1} \\
0 & 0 & 0 & \cdots & 0 & 0 & d_n & y_n
\end{array}\right)
$$

14. the solution of which is obtained by back substitution:

$$
\begin{aligned}
x_n &= y_n/d_n \\
x_{n-1} &= y_{n-1}/d_{n-1} - e_{n-1}x_n \\
&\vdots \\
x_k &= y_k/d_k - e_k x_{k+1} - f_k x_{k+2}, \quad k = n-2, n-3, \ldots, 1.
\end{aligned}
$$

Doolittle's Decomposition for Pentadiagonal Matrix

Decompose an $n \times n$ square symmetric pentadiagonal matrix A into LU with $L_{ii} = 1$ for $i = 1, 2, \ldots, n$.

INPUT ??

OUTPUT ??

Body ??

Example 13 (LDLt Decomposition)

Determine L and D such that $A = LDL^T$ and solve for $AX = B$ for X provided that A is a symmetric pentadiagonal matrix

$$A = \begin{pmatrix} 3 & 6 & 3 \\ 6 & 14 & 4 \\ 3 & 4 & 9 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ -2 \\ 6 \end{pmatrix}$$

1. Apply Gauss elimination $R_2 \leftarrow R_2 - (2) \cdot R_1, R_3 \leftarrow R_3 - (1) \cdot R_1$,

$$A \to \begin{pmatrix} 3 & 6 & 3 \\ 0 & 2 & -2 \\ 0 & -2 & 6 \end{pmatrix}, \qquad U^* \to \begin{pmatrix} 3 & \mathbf{2} & \mathbf{1} \\ 0 & 2 & -2 \\ 0 & -2 & 6 \end{pmatrix},$$

$$L \to \begin{pmatrix} 1 & 0 & 0 \\ \mathbf{2} & 1 & 0 \\ \mathbf{1} & ? & 1 \end{pmatrix}, \qquad D \to \begin{pmatrix} \mathbf{3} & 0 & 0 \\ 0 & \mathbf{2} & 0 \\ 0 & 0 & ? \end{pmatrix}.$$

2. $R_3 \leftarrow R_3 - (-1) \cdot R_2$,

$$A \longrightarrow \begin{pmatrix} 3 & 6 & 3 \\ 0 & 2 & -2 \\ 0 & 0 & 4 \end{pmatrix}, \qquad U^* \longrightarrow \begin{pmatrix} 3 & \mathbf{2} & \mathbf{1} \\ 0 & 2 & -\mathbf{1} \\ 0 & 0 & 4 \end{pmatrix},$$

$$L \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ \mathbf{2} & 1 & 0 \\ \mathbf{1} & -\mathbf{1} & 1 \end{pmatrix}, \qquad D \longrightarrow \begin{pmatrix} \mathbf{3} & 0 & 0 \\ 0 & \mathbf{2} & 0 \\ 0 & 0 & \mathbf{4} \end{pmatrix}.$$

3. From the result of previous step, $U = A$ and U^* contain all the information of D and L^T.

$$A = LU = LDL^T = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 6 & 3 \\ 0 & 2 & -2 \\ 0 & 0 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Definition 14 (Diagonal Dominance)

An $n \times n$ matrix A is said to be diagonally dominant if each diagonal element is larger than the sum of the other elements in the same row (we are talking here about absolute values). Thus diagonal dominance requires that $|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$ for $i = 1, 2, \ldots, n$.

Example 15

Matrix A is not diagonally dominant. But matrix B obtaining from A by rearrange rows in the following manner is diagonally dominant.

$$A = \begin{pmatrix} 1 & 4 & -2 \\ 2 & 0 & -3 \\ 3 & -1 & 1 \end{pmatrix}, B = \begin{pmatrix} 3 & -1 & 1 \\ 1 & 4 & -2 \\ 2 & 0 & -3 \end{pmatrix}$$

1. Consider the solution of $AX = B$ by Gauss elimination with row pivoting.
2. Pivoting aims at improving diagonal dominance of the coefficient matrix.
3. That is making the pivot element as large as possible in comparison to other elements in the pivot row.
4. The comparison is made easier if we establish an array s with the elements $s_i = \max_j |a_{ij}|, \ i = 1, 2, \ldots, n.$
5. Thus s_i, called the *scale factor* of row i, contains the absolute value of the largest element in the i-th row of A.
6. The *relative size* of an element a_{ij} (that is, relative to the largest element in the i-th row) is defined as the ratio $r_{ij} = |a_{ij}|/s_i$.
7. Suppose that the elimination phase has reached the stage where the k-th row has become the pivot row.
8. The augmented coefficient matrix at this point is shown in the following matrix:

$$
\left(\begin{array}{cccccc|c}
a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} & b_1 \\
0 & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} & b_2 \\
0 & 0 & a_{33} & a_{34} & \cdots & a_{3n} & b_3 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\hline
0 & \cdots & 0 & a_{kk} & \cdots & a_{kn} & b_k \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & a_{nk} & \cdots & a_{nn} & b_n
\end{array}\right)
$$

9. We do not automatically accept a_{kk} as the next pivot element, but look in the k-th column below a_{kk} for a “better” pivot.
10. The best choice is the element a_{pk} that has the largest relative size; that is, we choose p such that $r_{pk} = \max_i (r_{ik}),\ i \geq k$.

Example 16

Employ Gauss elimination with scaled row pivoting to solve the equations $AX = B$, where

$$A = \begin{pmatrix} 2 & -2 & 6 \\ -2 & 4 & 3 \\ -1 & 8 & 4 \end{pmatrix}, \ B = \begin{pmatrix} 16 \\ 0 \\ -1 \end{pmatrix}.$$

Definition 17 (Jacobi Method)

The Jacobi iterative method is obtained by solving the i-th equation in $AX = B$ for x_i to obtain (provided $a_{ii} \neq 0$)

$$x_i = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j, \; i = 1, 2, \ldots, n.$$

For each $k \geq 1$, generate the components $x_i^{(k)}$ of $x^{(k)}$ from the components of $x^{(k-1)}$ by

$$x_i^{(k)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k-1)}, \; i = 1, 2, \ldots, n.$$

As an illustration, consider a 3×3 system of linear equations

$$\begin{cases} x_1 = \dfrac{b_1}{a_{11}} - \dfrac{a_{12}}{a_{11}}x_2 - \dfrac{a_{13}}{a_{11}}x_3 \\ x_2 = \dfrac{b_2}{a_{22}} - \dfrac{a_{21}}{a_{22}}x_1 - \dfrac{a_{23}}{a_{22}}x_3 \\ x_3 = \dfrac{b_3}{a_{33}} - \dfrac{a_{31}}{a_{33}}x_2 - \dfrac{a_{32}}{a_{33}}x_3 \end{cases}$$

Jacobi Iteration is defined to be

$$\begin{cases} x_1^{(k)} = \dfrac{b_1}{a_{11}} - \dfrac{a_{12}}{a_{11}}x_2^{(k-1)} - \dfrac{a_{13}}{a_{11}}x_3^{(k-1)} \\ x_2^{(k)} = \dfrac{b_2}{a_{22}} - \dfrac{a_{21}}{a_{22}}x_1^{(k-1)} - \dfrac{a_{23}}{a_{22}}x_3^{(k-1)} \\ x_3^{(k)} = \dfrac{b_3}{a_{33}} - \dfrac{a_{31}}{a_{33}}x_2^{(k-1)} - \dfrac{a_{32}}{a_{33}}x_3^{(k-1)} \end{cases}$$

1. The components of $x^{(k-1)}$ are used to compute all the components $x_i^{(k)}$ of $X^{(k)}$.
2. But, for $i > 1$, the components $x_1^{(k)}, \dots, x_{i-1}^{(k)}$ of $X^{(k)}$ have already been computed and are expected to be better approximations to the actual solutions $x_1, \dots, x_{i-1}$ than are $x_1^{(k-1)}, \dots, x_{i-1}^{(k-1)}$.
3. It seems reasonable, then, to compute $x_i^{(k)}$ using these most recently calculated values. That is, to use

$$\begin{cases} x_1^{(k)} = \dfrac{b_1}{a_{11}} & & - \dfrac{a_{12}}{a_{11}} x_2^{(k-1)} & - \dfrac{a_{13}}{a_{11}} x_3^{(k-1)} \\ x_2^{(k)} = \dfrac{b_2}{a_{22}} & - \dfrac{a_{21}}{a_{22}} x_1^{(k)} & & - \dfrac{a_{23}}{a_{22}} x_3^{(k-1)} \\ x_3^{(k)} = \dfrac{b_3}{a_{33}} & - \dfrac{a_{31}}{a_{33}} x_2^{(k)} & - \dfrac{a_{32}}{a_{33}} x_3^{(k)} & \end{cases}$$

Definition 18 (Gauss-Siedel Method)

Gauss-Siedel iterative method is a modification of Jacobi by replacing $x_j^{(k-1)}$ by $x_j^{(k)}$ for $j = 1, 2, \dots, i-1$ in the i-the equation.

$$x_i^{(k)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k-1)}, \; i = 1, 2, \dots, n.$$

Example 19

Solving $AX = B$ for $X = (x_1, x_2, x_3)^T$ where

$$A = \begin{pmatrix} 2 & -1 & 2 \\ 4 & -2 & 2 \\ 3 & -1 & 1 \end{pmatrix}, \; B = \begin{pmatrix} 3 \\ 4 \\ 4 \end{pmatrix}$$

1. using Jacobi method with two iterations and
2. using Gauss-Siedel method with two iterations.

1. We introduce some notations about diagonal and off-diagonal parts of matrix coefficients A of the equation $AX = B$ as follow

$$
\begin{aligned}
A &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ -a_{21} & 0 & 0 \\ -a_{31} & -a_{32} & 0 \end{pmatrix} - \begin{pmatrix} 0 & -a_{12} & -a_{13} \\ 0 & 0 & -a_{23} \\ 0 & 0 & 0 \end{pmatrix} \\
&= D - L - U
\end{aligned}
$$

2. The equation $AX = B$ can be re-written as $(D - L - U)X = B$ and

$$
\begin{aligned}
DX &= (L + U)X + B &&\Rightarrow X = D^{-1}(L + U)X + D^{-1}B \\
(D - L)X &= UX + B &&\Rightarrow X = (D - L)^{-1}UX + (D - L)^{-1}B
\end{aligned}
$$

3. The Jacobi and Gauss-Siedel methods can be written in the form $X^{(k)} = D^{-1}(L+U)X^{(k-1)} + D^{-1}B = T_J X^{(k-1)} + C_J$ and $X^{(k)} = (D-L)^{-1}UX^{(k-1)} + (D-L)^{-1}B = T_G X^{(k-1)} + C_G$ respectively.
4. To study the convergence of general iteration techniques, we need to analyze the formula
$$X^{(k)} = TX^{(k-1)} + C, \text{ for } k = 1, 2, \ldots,$$
where $X^{(0)}$ is arbitrary.
5. Relaxation method represents a slight modification of the Gauss-Seidel method that is designed to enhance convergence.
6. After each new value of x is computed using Gauss-Seidel formula, that value is modified by a weighted average of the results of the previous and the present iterations:
$$x_i^{(k)} = \omega\left(\frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{(k)} - \sum_{j=i+1}^{n}\frac{a_{ij}}{a_{ii}}x_j^{(k-1)}\right) + (1-\omega)x_i^{(k-1)}$$

7. The relaxation factor ω is chosen to be positive and if $0 < \omega < 1$, it is called Under-Relaxation, if $1 < \omega$, it is called Over-Relaxation, and if $\omega = 1$, there is no Relaxation (unmodified Gauss-Seidel method.)
8. To determine the matrix form of the Relaxation method, we re-write formula as

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

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

$$X^{(k)} = T_\omega X^{(k-1)} + C_\omega$$

where

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

$$C_\omega = \omega(D - \omega L)^{-1}B$$

Definition 20 (Spectral Radius)

The spectral radius $\rho(A)$ of a square matrix A is defined by $\rho(A) = \max|\lambda|$, where λ is an eigenvalue of A and $|\lambda|$ is the absolute value or mudulus of λ.

Theorem 21

For any $X^{(0)} \in \mathbb{R}^n$, the sequence $\left\{X^{(k)}\right\}_{k=0}^{\infty}$ defined by $X^{(k)} = TX^{(k-1)} + C,\ \text{for each } k \geq 1$, converges to the unique solution of $X = TX + C$ if and only if $\rho(T) < 1$.

Theorem 22 (Kahan)

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

Theorem 23 (Ostrowski-Reich)

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

Theorem 24

If A is positive definite and tridiagonal, then $\rho(T_G) = [\rho(T_J)]^2 < 1$, and the optimal choice of ω for the Relaxation method is

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

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

Example 25 (Relaxation)

Consider an equation $AX = B$ with

$$a = \begin{pmatrix} 4 & 3 & 0 \\ 3 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}, b = \begin{pmatrix} 24 \\ 30 \\ -24 \end{pmatrix}.$$

1. Show that A is positive definite ($|A_k| > 0$, all its leading principle submatrices has positive determinant.)
2. Split $A = D - L - U$ and compute $T_J = D^{-1}(L + U)$.
3. Find all eigenvalues of T_J, then determine the spectral radius $\rho(T_J)$ of Jacobi matrix T_J and deduce the optimal value of Relaxation factor ω.
4. Use computer program to solve the equation using Relaxation method with initial guess $X^{(0)} = (1, 1, 1)^T$.

k	x_1	x_2	x_3	$\lvert x - x_0 \rvert$
0	1.0000000	1.0000000	1.0000000	None
1	6.3125000	3.5195312	-6.6501465	9.6485976
2	2.6223145	3.9585266	-4.6004238	4.2440016
3	3.1333027	4.0102646	-5.0966863	0.7141865
4	2.9570512	4.0074838	-4.9734897	0.2150575
5	3.0037211	4.0029250	-5.0057135	0.0568967
6	2.9963276	4.0009262	-4.9982822	0.0106717
7	3.0000498	4.0002586	-5.0003486	0.0043094
8	2.9997451	4.0000653	-4.9998924	0.0005816
9	3.0000025	4.0000150	-5.0000222	0.0002926
10	2.9999853	4.0000031	-4.9999935	0.0000355
11	3.0000008	4.0000005	-5.0000015	0.0000176
12	2.9999993	4.0000001	-4.9999996	0.0000024
13	3.0000001	4.0000000	-5.0000001	0.0000009
14	3.0000000	4.0000000	-5.0000000	0.0000002
15	3.0000000	4.0000000	-5.0000000	0.0000000

1. Consider the problem of finding the vector X that minimizes the scalar function $f(X) = \frac{1}{2}X^T AX - B^T X$ where the matrix A is symmetric and positive definite.
2. Because $f(X)$ is minimized when its gradient $\nabla f = AX - B$ is zero, we see that minimization is equivalent to solving $AX = B$.
3. Gradient methods accomplish the minimization by iteration, starting with an initial vector X_0.
4. Each iterative cycle k computes a refined solution $X_{k+1} = X_k + \alpha_k S_k$
5. The *step length* α_k is chosen so that X_{k+1} minimizes $f(X_{k+1})$ in the *search direction* S_k. That is X_{k+1} must satify $A(X_k + \alpha_k S_k) = B$.

6. Introducing the *residual* $R_k = B - AX_k$, the last equation becomes
$$\alpha_k = \frac{S_k^T R_k}{S_k^T A S_k}.$$
7. We choose $S_k = -\nabla f = R_k$, because this is the direction of the largest negative change in $f(X)$.
8. The resulting procedure is known as the *method of steepest descent.* It is not a popular algorithm because its convergence can be slow.
9. The more efficient conjugate gradient method uses the search direction $S_{k+1} = R_{k+1} + \beta_k S_k$.
10. The constant β_k is chosen so that the two successive search directions are conjugate to each other, meaning $S_{k+1}^T A S_k = 0$.
11. From the last two equations $\beta_k = -\dfrac{R_{k+1}^T A S_k}{S_k^T A S_k}$.

INPUT: Any initial vector X_0; symmetric and positive definite matrix coefficient A; column vector B; maximum number of iterations N; termimating tolerance TOL.
OUTPUT: Approximated solution X by Conjugate Gradient Method.

1. $R_0 \leftarrow B - AX_0$
2. $S_0 \leftarrow R_0$ (lacking a previous search direction, choose the direction of steepest descent)
3. For $k = 0$ to N do
   a. $\alpha_k \leftarrow \frac{S_k^T R_k}{S_k^T A S_k}$
   b. $X_{k+1} \leftarrow X_k + \alpha_k S_k$
   c. $R_{k+1} \leftarrow B - AX_{k+1}$
   d. If $|R_{k+1}| \leq TOL$, OUTPUT X_{k+1} and STOP.
   e. $\beta_k \leftarrow -\frac{R_{k+1}^T A S_k}{S_k^T A S_k}$
   f. $S_{k+1} \leftarrow R_{k+1} + \beta_k S_k$.

Example 26

Solve $AX = B$ for X by Conjugate Gradient method where

$$A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 4 & -2 \\ 1 & -2 & 4 \end{pmatrix}, \; B = \begin{pmatrix} 12 \\ -1 \\ 5 \end{pmatrix}, \; X_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

1. Initial guess $x = [0., 0., 0.]$.

2. First iterations

$$r = [12., -1., 5.]$$
$$|r| = 13.038404810405298$$
$$s = [12., -1., 5.]$$
$$\alpha = 0.2014218009$$
$$\beta = 0.1331169560$$
$$x = [2.4170616114, -0.2014218009, 1.0071090047]$$

3. Second iterations

$$r = [1.1232227488, 4.2369668246, -1.8483412322]$$
$$|r| = 4.757087609803571$$
$$s = [2.7206262213, 4.1038498686, -1.182756452]$$
$$\alpha = 0.2427607354$$
$$\beta = 0.0251606659$$
$$x = [3.0775228336, 0.7948318111, 0.7199821787]$$

4. Third iterations

$$\begin{aligned}
r &= [-0.2352417019, 0.3381599465, 0.632212074] \\
|r| &= 0.7545746578244767 \\
s &= [-0.1667889345, 0.4414155421, 0.602453134] \\
\alpha &= 0.4647960240 \\
x &= [3., 1., 1.]
\end{aligned}$$

5. Last check

$$\begin{aligned}
r = [&-1.7763568394e-15, \\
&8.8817841970e-16, \\
&-1.7763568394e-15] \\
|r| &= 2.6645352591003757e-15 < 1.0e-10
\end{aligned}$$