M.Sc. Course in
Applied Mathematics
with Oceanology and Computer Programming
Vidyasagar University

Semester-II

Paper-MTM 202 **Paper Name: Numerical Analysis**

Solution of System of Linear Equations

Module No. 4

Solution of Inconsistent and Ill Conditioned Systems

Objective

(a) Ill-conditioned system of equations

(b) Least squares method for inconsistent system

(c) Relaxation method

(d) Successive over relaxation method

Keywords

Ill-conditioned equations, least squares method, inconsistent system, relaxation method

Developed by

Professor Madhumangal Pal
Department of Applied Mathematics
Vidyasagar University, Midnapore-721102
email: mmpalvu@gmail.com

....

In previous modules, we have discussed several methods to solve a system of linear equations. In these modules, it is assumed that the given system is well-posed, i.e. if one (or more) coefficient of the system is slightly changed, then there is no major change in the solution. Otherwise the system of equations is called ill-posed or ill-conditioned. In this module, we will discussed about the solution methods of the ill-conditioned system of equations.

Before going to discuss the ill-conditioned system, we define some basic terms from linear algebra which are used to described the methods.

4.1 Vector and matrix norms

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a vector of dimension n. The **norm** of the vector $\mathbf{x}$ is the size or length of $\mathbf{x}$, and it is denoted by $\|\mathbf{x}\|$. The norm is a mapping from the set of vectors to a real number. That is, it is a real number which satisfies the following conditions:

$$\text{(i)}\ \ \|\mathbf{x}\| \geq 0 \quad \text{and} \quad \|\mathbf{x}\| = \mathbf{0} \quad \text{iff} \quad \mathbf{x} = \mathbf{0} \tag{4.1}$$

$$\text{(ii)}\ \ \|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\| \quad \text{for any real scalar } \alpha \tag{4.2}$$

$$\text{(iii)}\ \ \|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \ \ \text{(triangle inequality).} \tag{4.3}$$

Several type of norms are defined by many authors. The most use full vector norms are defined below.

$$\text{(i)}\ \ \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \tag{4.4}$$

$$\text{(ii)}\ \ \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \ \ \text{(Euclidean norm)} \tag{4.5}$$

$$\text{(iii)}\ \ \|\mathbf{x}\|_\infty = \max_i |x_i| \ \text{(maximum norm or uniform norm).} \tag{4.6}$$

Now, we define different type of matrix norms. Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices such that $\mathbf{A}+\mathbf{B}$ and $\mathbf{AB}$ are defined. The **norm of a matrix A** is denoted by $\|\mathbf{A}\|$ and it satisfies the following conditions

$$\text{(i)}\ \ \|\mathbf{A}\| \geq 0 \text{ and } \|\mathbf{A}\| = 0 \text{ iff } \mathbf{A} = \mathbf{0} \tag{4.7}$$

$$\text{(ii)}\ \ \|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|,\ \ \alpha \text{ is a real scalar} \tag{4.8}$$

$$\text{(iii)}\ \|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\| \tag{4.9}$$

$$\text{(iv)}\ \|\mathbf{AB}| \leq \|\mathbf{A}\|\|\mathbf{B}\|. \tag{4.10}$$

From (4.10), it follows that

$$\|\mathbf{A}^{\mathbf{k}}\| \leq \|\mathbf{A}\|^{k}, \tag{4.11}$$

for any positive integer k.

Like the vector norms, some common matrix norms are

$$\text{(i)}\ \ \|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}| \text{ (the column norm)} \tag{4.12}$$

$$\text{(ii)}\ \ \|\mathbf{A}\|_2 = \sqrt{\sum_i \sum_j |a_{ij}|^2} \quad \text{(the Euclidean norm)} \tag{4.13}$$

$$\text{(iii)}\ \|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}| \quad \text{(the row norm).} \tag{4.14}$$

The Euclidean norm is also known as **Erhard-Schmidt norm** or **Schur norm** or the **Frobenius norm**.

The concept of matrix norm is used to study the convergence of iterative methods to solve the system of linear equations. It is also used to study the **stability** of a system of equations.

Example 4.1 *Let*

$\mathbf{A} = \begin{bmatrix} 1 & 0 & -4 & 1 \\ 4 & 5 & 7 & 0 \\ 1 & -2 & 0 & 3 \end{bmatrix}$ *be a matrix. Find the matrix norms* $\|\mathbf{A}\|_1, \|A\|_2$ *and* $\|\mathbf{A}\|_\infty$.

Solution.

$\|\mathbf{A}\|_1 = \max\{1+4+1, 0+5-2, -4+7+0, 1+0+3\} = 6$

$\|\mathbf{A}\|_2 = \sqrt{1^2+0^2+(-4)^2+1^2+4^2+5^2+7^2+0^2+1^2+(-2)^2+0^2+3^2} = \sqrt{122}$ and

$\|\mathbf{A}\|_\infty = \max\{1+0-4+1, 4+5+7+0, 1-2+0+3\} = 16.$

4.2 Ill-conditioned system of linear equations

Let us consider the following system of linear equations.

$$\begin{aligned} x + \frac{1}{3}y &= 1.33 \\ 3x + y &= 4. \end{aligned} \tag{4.15}$$

It is easy to verify that this system of equations has no solution. But, for different approximate values of $\frac{1}{3}$ the system has different interesting results.

First we take $\frac{1}{3} \simeq 0.3$. Then the system becomes

$$\begin{aligned} x + 0.3y &= 1.33 \\ 3x + y &= 4. \end{aligned} \tag{4.16}$$

The solution of these equations is $x = 1.3,\ y = 0.1$.

If we approximate $\frac{1}{3}$ as 0.33, then the reduced system of equations is

$$\begin{aligned} x + 0.33y &= 1.33 \\ 3x + y &= 4 \end{aligned} \tag{4.17}$$

and its solution is $x = 1,\ y = 1$.

If the approximation is 0.333 then the system is

$$\begin{aligned} x + 0.333y &= 1.33 \\ 3x + y &= 4 \end{aligned} \tag{4.18}$$

and its solution is $x = -2, y = 10$.

When $\frac{1}{3} \simeq 0.3333$, then the system is

$$\begin{aligned} x + 0.3333y &= 1.33 \\ 3x + y &= 4 \end{aligned} \tag{4.19}$$

and its solution is $x = 100,\ y = -32$.

Note the systems of equations (4.15)-(4.19) and their solutions. These are very confusing situations. What is the best approximation of $\frac{1}{3}$? 0.3 or 0.3333. Observe that the solutions are significantly increased when the coefficient of y in first equation is sightly increased. That is, a small change in the coefficient of y in first equation of the system produces large change in the solution. These systems are called **ill-conditioned** or **ill-posed** system. On the other hand, if the change in the solution is small for small changes in the coefficients, then the system is called **well-conditioned** or **well-posed** system.

Let us consider the following system of equations

$$\mathbf{Ax} = \mathbf{b}. \tag{4.20}$$

Suppose one or more elements of the matrices $\mathbf{A}$ and/or $\mathbf{b}$ be changed and let them be $\mathbf{A}'$ and $\mathbf{b}'$. Also, let $\mathbf{y}$ be the solution of the new system, i.e.

$$\mathbf{A}'\mathbf{y} = \mathbf{b}'. \tag{4.21}$$

Assumed that the changes in the coefficients are very small.

The system of equations (4.20) is called ill-conditioned when the change in $\mathbf{y}$ is too large compared to the solution vector $\mathbf{x}$ of (4.20). Otherwise, the system of equations is called well-conditioned. If a system is ill-conditioned then the corresponding coefficient matrix is called an **ill-conditioned matrix**.

For the above problem, i.e. for the system of equations (4.17) coefficient matrix is $\begin{bmatrix} 1 & 0.33 \\ 3 & 1 \end{bmatrix}$ and it is an ill-conditioned matrix.

When $|\mathbf{A}|$ is small then, in general, the matrix $\mathbf{A}$ is ill-conditioned. But, the term small has no definite meaning. So many methods are suggested to measure the ill-conditioned of a matrix. One of the simple methods is defined below.

Let $\mathbf{A}$ be a matrix and the condition number (denoted by $Cond(\mathbf{A})$) of it is define by

$$Cond(\mathbf{A}) = \|\mathbf{A}\| \, \|\mathbf{A^{-1}}\| \tag{4.22}$$

where $\|\mathbf{A}\|$ is any type of matrix norm. If $Cond(\mathbf{A})$ is large then the matrix is called ill-conditioned and corresponding system of equations is called ill-conditioned system of equations. If $Cond(\mathbf{A})$ is small then the matrix $\mathbf{A}$ and the corresponding system of equations are called well-conditioned.

. . . .

Let us consider the following two matrices to illustrated the ill-conditioned and well-conditioned cases. Let $\mathbf{A} = \begin{bmatrix} 0.33 & 1 \\ 1 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 4 & 4 \\ 3 & 5 \end{bmatrix}$ be two matrices.

Then $\mathbf{A^{-1}} = \begin{bmatrix} -300 & 100 \\ 100 & -33 \end{bmatrix}$ and $\mathbf{B^{-1}} = \frac{1}{11}\begin{bmatrix} 0.625 & -0.500 \\ -0.375 & 0.500 \end{bmatrix}$.

The Euclidean norms of $\mathbf{A}$ and $\mathbf{B}$ are $\|\mathbf{A}\|_2 = \sqrt{0.10890 + 1 + 1 + 9} = 3.3330$ and $\|\mathbf{A^{-1}}\|_2 = 333.300$.

Thus, $Cond(\mathbf{A}) = \|\mathbf{A}\|_2 \ \|\mathbf{A^{-1}}\|_2 = 3.3330 \times 333.300 = 1110.8889$, a very large number. Hence $\mathbf{A}$ is ill-conditioned.

For the matrix $\mathbf{B}$, $\|\mathbf{B}\|_2 = \sqrt{16 + 16 + 9 + 25} = 8.1240$ and $\|\mathbf{B^{-1}}\|_2 = 1.01550$
Then $Cond(\mathbf{B}) = 8.24992$, a relatively small quantity.

Thus, the matrix $\mathbf{B}$ is well-conditioned.

The value of $Cond(A)$ lies between 0 and ∞. If it is large then we say that the matrix is ill-conditioned. But, there is no definite meaning of large number. So, this measure is not good.

Now, we define another parameter whose value lies between 0 and 1.

Let $\mathbf{A} = [a_{ij}]$ be a matrix and $r_i = \left(\sum\limits_{j=1}^{n} a_{ij}^2\right)^{1/2}$, $i = 1, 2, \ldots, n$. The quantity

$$\nu(\mathbf{A}) = \frac{|\mathbf{A}|}{r_1 r_2 \cdots r_n} \tag{4.23}$$

measures the smallness of the determinant $|\mathbf{A}|$. It can be shown that $-1 \leq \nu \leq 1$. If $|\nu(\mathbf{A})|$ is closed to zero, then the matrix $\mathbf{A}$ is ill-conditioned and if it is closed to 1, then $\mathbf{A}$ is well-conditioned.

For the matrix $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 0.22 & 1 \end{bmatrix}$, $r_1 = \sqrt{17},\ r_2 = 1.0239,\ |\mathbf{A}| = 0.12$,
$\nu(\mathbf{A}) = \dfrac{0.12}{\sqrt{17} \times 1.0239} = 0.0284$ and for the matrix $\mathbf{B} = \begin{bmatrix} 3 & 5 \\ -2 & 2 \end{bmatrix}$, $r_1 = \sqrt{34},\ r_2 = \sqrt{8}$,
$|\mathbf{B}| = 16,\ \nu(\mathbf{B}) = \dfrac{16}{\sqrt{34} \times \sqrt{8}} = 0.9702$.

Thus the matrix $\mathbf{A}$ is ill-conditioned while the matrix $\mathbf{B}$ is well-conditioned as its value is very closed to 1.

4.3 Least squares method for inconsistent system

Let us consider a system of equations whose number of equations is not equal to number of variables. Let such system be

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4.24}$$

where $\mathbf{A}, \mathbf{x}$ and $\mathbf{b}$ are of order $m \times n$, $n \times 1$ and $m \times 1$ respectively. Note that the coefficient matrix is rectangular. Thus, either the system has no solution or it has infinite number of solutions. Assumed that the system is inconsistent. So, it does not have any solution. But, the system may have a least squares solution. A solution $\mathbf{x}'$ is said to be least squares if $\mathbf{A}\mathbf{x}' - \mathbf{b} \neq \mathbf{0}$, but $\|\mathbf{A}\mathbf{x}' - \mathbf{b}\|$ is minimum. The solution $\mathbf{x_m}$ is called the minimum norm least squares solution if

$$\|\mathbf{x_m}\| \leq \|\mathbf{x_l}\| \tag{4.25}$$

for any $\mathbf{x_l}$ such that

$$\|\mathbf{A}\mathbf{x_l} - \mathbf{b}\| \leq \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \qquad \text{for all } \mathbf{x}. \tag{4.26}$$

Since $\mathbf{A}$ is rectangular matrix, so its solution can be determined by the following equation

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b}, \tag{4.27}$$

where $\mathbf{A}^{+}$ is the g-inverse of $\mathbf{A}$.

Since the Moore-Penrose inverse $\mathbf{A}^{+}$ is unique, therefore the minimum norm least squares solution is unique.

The solution can also be determined by without finding the g-inverse of $\mathbf{A}$. This method is described below.

If $\mathbf{x}$ is the exact solution of the system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, then $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0}$, otherwise $\mathbf{A}\mathbf{x} - \mathbf{b}$ is a non-null matrix of order $m \times 1$. In explicit form this vector is

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n - b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n - b_2 \\ \cdots\cdots \quad \cdots\cdots \quad \cdots \quad \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n - b_m \end{bmatrix}.$$

. . . .

Let square of the norm $\|\mathbf{Ax} - \mathbf{b}\|$ be denoted by S. Therefore,

$$\begin{aligned} S &= (a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n - b_1)^2 \\ &\quad +(a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n - b_2)^2 + \cdots \\ &\quad +(a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n - b_n)^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij}x_j - b_i)^2. \end{aligned} \tag{4.28}$$

The quantity S is called the sum of square of residuals. Now, our aim is to find the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ such that S is minimum. The sufficient conditions for which S to be minimum are

$$\frac{\partial S}{\partial x_1} = 0,\ \frac{\partial S}{\partial x_2} = 0,\ \cdots,\ \frac{\partial S}{\partial x_n} = 0 \tag{4.29}$$

Note that the system of equations (4.29) is non-homogeneous and contains n equations with n unknowns $x_1, x_2, \ldots, x_n$. This system of equations can be solved by any method described in previous modules.

Let $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_n = x_n^*$ be the solution of the equations (4.29). Therefore, the least squares solution of the system of equations (4.24) is

$$x^* = (x_1^*, x_2^*, \ldots, x_n^*)^t. \tag{4.30}$$

The sum of square of residuals (i.e. the sum of the squares of the absolute errors) is given by

$$S^* = \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij}x_j^* - b_i)^2. \tag{4.31}$$

Let us consider two examples to illustrate the least squares method which is used to solve inconsistent system of equations.

Example 4.2 *Find g-inverse of the singular matrix* $\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 1 & 2 \end{bmatrix}$ *and hence find a least squares solution of the inconsistent system of equations* $4x + 8y = 2, x + 2y = 1$.

Solution. Let $\alpha_\mathbf{1} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\alpha_\mathbf{2} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$, $\mathbf{A_1} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$.

$\mathbf{A_1^+} = (\alpha_\mathbf{1}^\mathbf{t}\alpha_\mathbf{1})^{-\mathbf{1}}\alpha_\mathbf{1}^\mathbf{t} = \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{17} & \frac{1}{17} \end{bmatrix}$,

$\delta_{\mathbf{2}} = \mathbf{A_1^+}\alpha_{\mathbf{2}} = \begin{bmatrix} \frac{4}{17} & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 8 \\ 2 \end{bmatrix} = 2,$

$\gamma_{\mathbf{2}} = \alpha_{\mathbf{2}} - \mathbf{A_1}\delta_{\mathbf{2}} = \begin{bmatrix} 8 \\ 2 \end{bmatrix} - 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$ (a null vector),

$\beta_{\mathbf{2}} = (1 + \delta_{\mathbf{2}}^{\mathbf{t}}\delta_{\mathbf{2}})^{-\mathbf{1}}\delta_{\mathbf{2}}^{\mathbf{t}}\mathbf{A_1^+} = \frac{1}{5}.2. \begin{bmatrix} \frac{4}{17} & \frac{1}{17} \end{bmatrix} = \begin{bmatrix} \frac{8}{85} & \frac{2}{85} \end{bmatrix}$

$\delta_{\mathbf{2}}\beta_{\mathbf{2}} = \begin{bmatrix} \frac{16}{85} & \frac{4}{85} \end{bmatrix}.$

Therefore,

$$\mathbf{A_2^+} = \begin{bmatrix} \mathbf{A_1^+} - \delta_{\mathbf{2}}\beta_{\mathbf{2}} \\ \beta_{\mathbf{2}} \end{bmatrix} = \begin{bmatrix} \frac{4}{85} & \frac{1}{85} \\ \frac{8}{85} & \frac{2}{85} \end{bmatrix}.$$

This is the g-inverse of $\mathbf{A}$.

Second Part: In matrix notation, the given system of equations is

$\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 1 & 2 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Note that the coefficient matrix is singular. So, it has no conventional solution. But, the least squares solution of this system of equation is

$\mathbf{x} = \mathbf{A^+b}$, i.e.

$$\mathbf{x} = \frac{1}{85} \begin{bmatrix} 4 & 1 \\ 8 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 9/85 \\ 18/85 \end{bmatrix}.$$

Hence, the least squares solution is

$$x = \frac{9}{85}, \ y = \frac{18}{85}.$$

Example 4.3 *Find the least squares solution of the following system of linear equations $x + 2y = 2.0, x - y = 1.0, x + 3y = 2.3,$ and $2x + y = 2.9$. Also, estimate the residual.*

Solution. Let x^*, y^* be the least squares solution of the given system of equations. Then the sum of square of residuals S is

$S = (x^* + 2y^* - 2.0)^2 + (x^* - y^* - 1.0)^2 + (x^* + 3y^* - 2.3)^2 + (2x^* + y^* - 2.9)^2.$

Now, the problem is to find the values of x^* and y^* in such a way that S is minimum. Thus,

$$\frac{\partial S}{\partial x^*} = 0 \text{ and } \frac{\partial S}{\partial y^*} = 0.$$

. . . .

Therefore the normal equations are,

$$2(x^* + 2y^* - 2.0) + 2(x^* - y^* - 1.0) + 2(x^* + 3y^* - 2.3) + 4(2x^* + y^* - 2.9) = 0$$

$$\text{and } 4(x^* + 2y^* - 2.0) - 2(x^* - y^* - 1.0) + 6(x^* + 3y^* - 2.3) + 2(2x^* + y^* - 2.9) = 0.$$

After simplification, these equations reduce to $7x^*+6y^* = 11.1$ and $6x^*+15y^* = 12.8$.

The solution of these equations is $x^* = 1.3$ and $y^* = \dfrac{1}{3} = 0.3333$. This is the least squares solution of the given system of equations.

The sum of the square of residuals is $S = (1.3 + 2 \times 0.3333 - 2)^2 + (1.3 - 0.3333 - 1)^2 + (1.3 + 3 \times 0.3333 - 2.3)^2 + (2 \times 1.3 + 0.3333 - 2.9)^2 = 0.0033$.

4.4 Method to solve ill-conditioned system

It is very difficult to solve a system of ill-conditioned equations. Few methods are available to solve an ill-conditioned system of linear equations. One simple concept to solve an ill-conditioned system is to carry out the calculations with large number of significant digits. But, computation with more significant digits takes much time. One better method is to improve upon the accuracy of the approximate solution by an iterative method. Such an iterative method is consider below.

Let us consider the following ill-conditioned system of equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, 2, \ldots, n. \tag{4.32}$$

Let $\{\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n\}$ be an approximate solution of (4.32). Since this is an approximate solution, therefore $\sum\limits_{j=1}^{n} a_{ij}\widetilde{x}_j$ is not necessarily equal to b_i. For this solution, let the right hand vector be $\widetilde{b}_i$, i.e. $b_i = \widetilde{b}_i$.

Thus, for this solution the equation (4.32) becomes

$$\sum_{j=1}^{n} a_{ij}\widetilde{x}_j = \widetilde{b}_i, \quad i = 1, 2, \ldots, n. \tag{4.33}$$

Subtracting (4.33) from (4.32), we get

$$\sum_{j=1}^{n} a_{ij}(x_j - \widetilde{x}_j) = (b_i - \widetilde{b}_i)$$

$$\text{i.e.,} \quad \sum_{j=1}^{n} a_{ij}\varepsilon_i = d_i \tag{4.34}$$

where $\varepsilon_i = x_i - \widetilde{x}_i, d_i = b_i - \widetilde{b}_i, i = 1, 2, \ldots, n$.

Now, equation (4.34) is again a system of linear equations whose unknowns are $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. By solving these equations we obtained the values of ε_i's. Hence, the new solution is given by $x_i = \varepsilon_i + \widetilde{x}_i$ and this solution is better approximation to $\widetilde{x}_i$'s. This technique may be repeated to get more better solution.

4.5 The relaxation method

The relaxation method, invented by Southwell in 1946, is an iterative method used to solved a system of linear equations.

Let

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \tag{4.35}$$

be the ith, $i = 1, 2, \ldots, n$ equation of a system of linear equations. Let $\mathbf{x^{(k)}} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})^t$ be the kth iterated solution of the system of linear equations.

Then

$$\sum_{j=1}^{n} a_{ij}x_j^{(k)} \simeq b_i, \quad i = 1, 2, \ldots, n.$$

Now, we denote the kth iterated residual for the ith equation by $r_i^{(k)}$. Therefore, the value of $r_i^{(k)}$ is given by

$$r_i^{(k)} = b_i - \sum_{j=1}^{n} a_{ij}x_j^{(k)}, \quad i = 1, 2, \ldots, n. \tag{4.36}$$

If $r_i^{(k)} = 0$ for all $i = 1, 2, \ldots, n$, then $(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})^t$ is the exact solution of the given system of equations. If the residuals are not zero or not small for all equations, then apply the same method to reduce the residuals.

. . . .

In relaxation method, the solution can be improved successively by reducing the largest residual to zero at that iteration. To get the fast convergence, the equations are rearranged in such a way that the largest coefficients in the equations appear on the diagonals, i.e. the coefficient matrix becomes diagonally dominant.

The aim of this method is to reduce the largest residual to zero. Let r_p be the largest residual (in magnitude) occurs at the pth equation for a particular iteration. Then the value of the variable x_p be increased by dx_p where

$$dx_p = -\frac{r_p}{a_{pp}}.$$

That is, x_p is replaced by $x_p + dx_p$ to relax r_p, i.e. to reduce r_p to zero. Then the modified solution after this iteration is

$$\mathbf{x^{(k+1)}} = \left(x_1^{(k)}, x_2^{(k)}, \ldots, x_{p-1}^{(k)}, x_p + dx_p, x_{p+1}^{(k)}, \ldots, x_n^{(k)}\right).$$

The method is repeated until all the residuals become zero or tends to zero.

Example 4.4 *Solve the following system of linear equations by relaxation method taking* $(0, 0, 0)$ *as initial solution*

$$27x + 6y - z = 54, \qquad 6x + 15y + 2z = 72, \qquad x + y + 54z = 110.$$

Solution. The given system of equations is diagonally dominant.

The residuals r_1, r_2, r_3 are given by the following equations

$$r_1 = 54 - 27x - 6y + z$$
$$r_2 = 72 - 6x - 15y - 2z$$
$$r_3 = 110 - x - y - 54z.$$

Here, the initial solution is $(0, 0, 0)$, i.e. $x = y = z = 0$. Therefore, the residuals are $r_1 = 54, r_2 = 72, r_3 = 110$. The largest residual in magnitude is r_3. Thus, the third equation has more error and we have to improve x_3.

Then the increment dx_3 in x_3 is now calculated as $dx_3 = -\dfrac{r_3}{a_{33}} = \dfrac{110}{54} = 2.037$.

Thus the first iterated solution is $(0, 0, 0 + 2.037)$, i.e. $(0, 0, 2.037)$.

In next iteration we determine the new residuals of large magnitudes and relax it to zero. The process is repeated until all the residuals become zero or very small.

All steps of all iterations are shown below:

	residuals			max		increment	solution		
k	r_1	r_2	r_3	(r_1, r_2, r_3)	p	dx_p	x	y	z
0	–	–	–	–	–	–	0	0	0
1	54	72	110	110	3	2.037	0	0	2.037
2	56.037	67.926	0.003	67.926	2	4.528	0	4.528	2.037
3	28.869	0.006	–0.4526	28.869	1	1.069	1.069	4.528	2.037
4	0.006	–6.408	–5.89	–6.408	2	–0.427	1.069	4.101	2.037
5	2.568	–0.003	–5.168	–5.168	3	–0.096	1.069	4.101	1.941
6	2.472	0.189	0.016	2.472	1	0.092	1.161	4.101	1.941
7	–0.012	–0.363	–0.076	–0.363	2	–0.024	1.161	4.077	1.941
8	0.132	–0.003	–0.052	0.132	1	0.005	1.166	4.077	1.941
9	–0.003	–0.033	–0.057	–0.057	3	0.001	1.166	4.077	1.940
10	–0.004	–0.031	–0.003	–0.031	2	–0.002	1.166	4.075	1.940
11	0.008	–0.001	–0.001	0.008	1	0.000	1.166	4.075	1.940
12	–0.008	–0.001	–0.001	0.008	2	0.000	1.166	4.075	1.940

In this case, all residuals are very small. The solution of the given system of equations is $x_1 = 1.166,\ x_2 = 4.075,\ x_3 = 1.940$, correct upto three decimal places.

4.6 Successive overrelaxation (S.O.R.) method

The relaxation method can be modified to achieve fast convergence. For this purpose, a suitable relaxation factor w is introduced. The ith equation of the system of equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, 2, \ldots, n \tag{4.37}$$

is

$$\sum_{j=1}^{n} a_{ij}x_j = b_i.$$

This equation can be written as

. . . .

$$\sum_{j=1}^{i-1} a_{ij}x_j + \sum_{j=i}^{n} a_{ij}x_j = b_i. \tag{4.38}$$

Let $\left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\right)$ be the initial solution and

$$\left(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i^{(k)}, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right),$$

be the solution when ith equation being consider. Then the equation (4.38) becomes

$$\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} + \sum_{j=i}^{n} a_{ij}x_j^{(k)} = b_i. \tag{4.39}$$

Since $\left(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i^{(k)}, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right)$ is an approximate solution of the given system of equations, therefore the residual at the ith equation is determine from the following equation:

$$r_i = b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(k)}, \quad i = 1, 2, \ldots, n. \tag{4.40}$$

We denote the differences of x_i's at two consecutive iterations by $\varepsilon_i^{(k)}$ and it is defined as $\varepsilon_i^{(k)} = x_i^{(k+1)} - x_i^{(k)}$.

In the successive overrelaxation (SOR) method, it is assumed that

$$a_{ii}\, \varepsilon_i^{(k)} = w\, r_i, \quad i = 1, 2, \ldots, n, \tag{4.41}$$

where w is a scalar, called the **relaxation factor**.

Thus, the equation (4.41) becomes

$$a_{ii}x_i^{(k+1)} = a_{ii}x_i^{(k)} - w\left[\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} + \sum_{j=i}^{n} a_{ij}x_j^{(k)} - b_i\right], \tag{4.42}$$
$$i = 1, 2, \ldots, n; \quad k = 0, 1, 2, \ldots$$

The iteration process is repeated until desired accuracy is achieved.

The above iteration method is called the **overrelaxation method** when $1 < w < 2$, and is called the **under relaxation** method when $0 < w < 1$. When $w = 1$, the method becomes well known Gauss-Seidal's iteration method.

The proper choice of w can speed up the convergence of the iteration scheme and it depends on the given system of equations.

Example 4.5 *Solve the following system of linear equations*

$$4x_1 + 2x_2 + x_3 = 5,$$
$$x_1 + 5x_2 + 2x_3 = 6,$$
$$-x_1 + x_2 + 7x_3 = 2$$

by SOR method taken relaxation factor $w = 1.02$.

Solution. The SOR iteration scheme for the given system of equations is

$$\begin{aligned} 4x_1^{(k+1)} &= 4x_1^{(k)} - 1.02\Big[4x_1^{(k)} + 2x_2^{(k)} + x_3^{(k)} - 5\Big] \\ 5x_2^{(k+1)} &= 5x_2^{(k)} - 1.02\Big[x_1^{(k+1)} + 5x_2^{(k)} + 2x_3^{(k)} - 6\Big] \\ 7x_3^{(k+1)} &= 7x_3^{(k)} - 1.02\Big[-x_1^{(k+1)} + x_2^{(k+1)} + 7x_3^{(k)} - 2\Big]. \end{aligned}$$

Let $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$.

The calculations of all iterations are shown below:

k	x_1	x_2	x_3
0	0	0	0
1	1.275	0.9639	0.33676
2	0.67204	0.93023	0.24707
3	0.72414	0.95686	0.25257
4	0.70811	0.95736	0.25006
5	0.70882	0.95823	0.25008
6	0.70835	0.95829	0.25001
7	0.70835	0.95832	0.25000
8	0.70833	0.95833	0.25000
9	0.70833	0.95833	0.25000

The solutions at iterations 8th and 9th are same. Hence, the required solution is

$$x_1 = 0.7083, \quad x_2 = 0.9583, \quad x_3 = 0.2500$$

correct up to four decimal places.

. . . .

Self Assessment (MCQ/Short answer questions)

1. Let w be the relaxation factor. The method is called overrelaxation, if
   (a) $0 < w < 1$ (b) $w = 1$ (c) $1 < w < 2$ (d) $w > 2$

2. Let w be the relaxation factor. The method is called underrelaxation, if
   (a) $0 < w < 1$ (b) $w = 1$ (c) $1 < w < 2$ (d) $w > 2$

3. Let w be the relaxation factor. Then the relaxation method becomes Gauss-Seidal iteration method, if
   (a) $0 < w < 1$ (b) $w = 1$ (c) $1 < w < 2$ (d) $w > 2$

4. A system of equations is called ill-conditioned, if
   (a) all the coefficients are very small
   (b) all the coefficients are very large
   (c) for minor changes of the coefficient(s), change of solution is high
   (d) none of these

5. The matrix $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 0.33 & 1 \end{bmatrix}$ is
   (a) well-conditioned (b) ill-conditioned

6. The matrix $\mathbf{B} = \begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix}$ is
   (a) well-conditioned (b) ill-conditioned

7. For consistent system of linear equations, the least squares solution is same as actual solution.
   (a) true (b) false

8. Least squares method is also applicable, if the number of variables is not equal to the number of equations.
   (a) true (b) false

9. Least squares solution for inconsistent system satisfies all the equations.
   (a) true (b) false

10. The rate of convergence of Gauss-Seidal iteration method is more than the over-relaxation method.
    (a) true (b) false

11. The matrix norms $\|\mathbf{A}\|_1, \|A\|_2$ and $\|\mathbf{A}\|_\infty$ for the matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & -1 & 5 \\ 3 & 2 & 6 \end{bmatrix}$ are $\cdots\cdots$.

Answer to the questions

1. (c)
2. (a)
3. (b)
4. (c)
5. (b)
6. (a)
7. (a)
8. (a)
9. (b)
10. (b)
11. 15, $\sqrt{104}$, 11

Self Assessment (Long answer questions)

1. Test the following system for ill-condition

   $10x + 7y + 8z + 7w = 32$

   $7x + 5y + 6z + 5w = 23$

   $8x + 6y + 10z + 9w = 33$

   $7x + 5y + 9z + 10w = 31.$

2. Solve the following system of equations

   $3x_1 + x_2 + 2x_3 = 6$

   $-x_1 + 4x_2 + 2x_3 = 5$

   $2x_1 + x_2 + 4x_3 = 7$

   by SOR method taken $w = 1.01$.

3. Solve the following system of equations

   $8x_1 + x_2 - x_3 = 8, 2x_1 + x_2 + 9x_3 = 12, x_1 - 7x_2 + 2x_3 = -4$ by relaxation method taking (0, 0, 0) as initial solution.

4. Find the least squares solution of the following equations $x + y = 3.0, 2x - y = 0.03, x + 3y = 7.03$, and $3x + y = 4.97$. Also, estimate the residue.

Learn More

1. Danilina, N.I., Dubrovskaya, S.N., and Kvasha, O.P., and Smirnov, G.L. Computational Mathematics. Moscow: Mir Publishers, 1998.

2. Hildebrand, F.B. Introduction of Numerical Analysis. New York: London: McGraw-Hill, 1956.

3. Householder, A.S. The Theory of Matrices in Numerical Analysis. New York: Blaisdell, 1964.

4. Jain, M.K., Iyengar, S.R.K., and Jain, R.K. Numerical Methods for Scientific and Engineering Computation. New Delhi: New Age International (P) Limited, 1984.

5. Krishnamurthy, E.V., and Sen, S.K. Numerical Algorithms. New Delhi: Affiliated East-West Press Pvt. Ltd., 1986.

6. Mathews, J.H. Numerical Methods for Mathematics, Science, and Engineering, 2nd ed., NJ: Prentice-Hall, Inc., 1992.

7. Pal, M. Numerical Analysis for Scientists and Engineers: Theory and C Programs. New Delhi: Narosa, Oxford:Alpha Sciences, 2007.