Communications in Mathematics and Applications
Vol. 12, No. 4, pp. 853–867, 2021
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v12i4.1766

http://www.rgnpublications.com

Research Article

Study of Numerical Solution of Linear System of Equations by Using SOR Algorithm with $0 < \omega < 2$

Najmuddin Ahmad* and Fauzia Shaheen

Department of Mathematics, Integral University, Lucknow, India

Received: March 22, 2021 **Accepted:** June 13, 2021

Abstract. In this paper, we are studying new approaches in numerical accuracy of the linear system of equations by successive over-relaxation method, analyzing the convergence criteria of iterative methods and comparing the SOR method with other iterative methods. We have shown SOR method converges more rapidly with the others with the help of some typical examples. All the calculations have been performed with the help of MATLAB 2020R.

Keywords. Jacobi method, Gauss-Seidal method, Richardson method, SOR method, Spectral radius

Mathematics Subject Classification (2020). 65F10, 65F08

1. Introduction

Numerical analysis is the area of mathematics and computer science that creates, analyzes, and implements algorithms for solving numerically the problems of continuous mathematics. Numerical linear algebra refers to problems involving the solution of systems of linear equations, possibly with a very large number of variables [6]. Linear systems are usually written using matrix-vector notation,

$$Ax = b,$$

with A the matrix of coefficients for the system, x the column vector of the unknown variable $x^1, \dots, x^n$ and b a given column vector. For larger linear systems there are two types of methods:

*Email: najmuddinahmad33@gmail.com

(1) direct method, and (2) indirect method. Direct methods lead to a theoretically exact solution x in a finite number of steps. Indirect methods are approximate methods which create a sequence of approximating solutions of increasing accuracy. Indirect methods are also called as Iterative methods [4].

The rate of converges of an iterative method depends strongly on the eigen values of the coefficient matrix A. Hence, generally there is a transformation of the coefficient matrix A into another matrix, called a preconditioner, with a more favorable eigen values. A good preconditioner improves the convergence of the iterative methods. Indeed, a preconditioner plays an important role on the convergence of iterative methods [8].

The first iterative method was given by the Jacobi (1824) and later by the Gauss-Seidel (1848) (cf. Curtis *et al.* [1]). After about 100 years the popular *Successive Over Relaxation* (SOR) method was discovered by D. M. Young [10]. He introduced a relaxation factor to the Gauss-Seidal method to increase the rate of convergence. Gauss-Seidal method is a particular case of SOR method. To find the value of relaxation parameter, is a difficult task. The value of relaxation parameter in SOR method is not yet find for all types of system of equations. The value of relaxation parameter, lies in between 0 and 2 such that the radius of convergence of iteration matrix of SOR method, should be less than 1 (see [2, 11]). In Section 1, we have given some basic ideas of a linear system of equations and iterative methods which are used to solve the problems. In Second 2, we have described briefly the iterative methods. In Section 3, we have given convergence criteria of iterative methods. Some numerical examples are given in Section 4. In Section 5, we have concluded some results based on numerical examples.

2. Iterative Methods

2.1 Jacobi Method

Consider the linear system of equations as

$$Ax = b,$$

where $A \in \mathbb{R}^{n\times n}$ is a coefficient matrix and $b \in \mathbb{R}^n$ is a given n-dimensional real vector. The iterative formula in component form according to Jacobi is given as

$$x_i^{(k)} = \frac{1}{a_{ii}}\left(b_i - \sum_{i\neq j} a_{ij}x_j^{(k-1)}\right),$$

where $x_i^{(k)}$ is the solution vector after kth iteration by taking $x_i^{(0)}$ (initial guess) as $(0,0,\dots,0)^t$. Jacobi method takes the value of previous iteration to for the solution vector. If splitting of the coefficient matrix A is given as $A = \mathfrak{D} + L + U$, where $\mathfrak{D}$ is a diagonal matrix, L is strictly lower triangular matrix, and U is strictly upper triangular matrix. Then, the Jacobi method in the matrix form is given as:

$$\begin{aligned} x^{(k)} &= -\mathfrak{D}^{-1}(L+U)x^{(k-1)} + \mathfrak{D}^{-1}b \\ &= \mathfrak{B}_J x^{(k-1)} + \mathfrak{C}, \end{aligned}$$

where $\mathfrak{B}_J = -\mathfrak{D}^{-1}(L+U)$, and is an iteration matrix of Jacobi method and $\mathfrak{C} = \mathfrak{D}^{-1}b$ is an iteration vector.

2.2 Gauss-Seidal Method

This method is also called as the method of successive displacement. In Gauss-Seidal method the updated values of variables are used to find the solution vector. The iterative formula in component form according to Gauss-Seidal is given as:

$$x_i^{(k)} = \frac{1}{a_{ii}}\left(b_i - \sum_{i<J} a_{ij}x_j^{(k-1)} - \sum_{i>J} a_{ij}x_j^{(k)}\right).$$

The Gauss-Seidal method in the matrix form is given as:

$$\begin{aligned} x^{(k)} &= -(\mathfrak{D}+L)^{-1}Ux^{(k-1)} + (\mathfrak{D}+L)^{-1}b \\ &= \mathfrak{B}_{GS}x^{(k-1)} + \mathfrak{C}, \end{aligned}$$

where $\mathfrak{B}_{GS} = -(\mathfrak{D}+L)^{-1}U$ is an iteration matrix, and $\mathfrak{C} = (\mathfrak{D}+L)^{-1}b$ is an iteration vector.

2.3 Richardson's Method

In the matrix form the Richardson's method is given as:

$$\begin{aligned} x_i^{(k)} &= x_i^{(k-1)} + \omega(b-A)x_i^{(k-1)} \\ &= (I-\omega A)x_i^{(k-1)} + \omega b, \\ x^{(k)} &= \mathfrak{B}_{RM}x^{(k-1)} + \mathfrak{C}, \end{aligned}$$

where ω is a scalar parameter that has to be chosen such that the sequence $x^{(k)}$ converges, where $\mathfrak{B}_{RM} = (I-A)$ is an iteration matrix, and $\mathfrak{C} = \omega b$ is an iteration vector.

2.4 SOR Method

For any iterative method, in finding $x^{(k)}$ from $x^{(k-1)}$, we move a certain amount in a particular direction from $x^{(k-1)}$ to $x^{(k)}$. This direction is the vector $x^{(k)} - x^{(k-1)}$ since

$$x^{(k)} = x^{(k-1)} + (x^{(k)} - x^{(k-1)}).$$

If we assume that the direction from $x^{(k-1)}$ to $x^{(k)}$ is taking as closer, but not all the way, to the true solution x, then it would make sense to move in the same direction $x^{(k)} - x^{(k-1)}$, but farther along that direction. Here is how we derive the SOR method from the Gauss-Seidel method. First, notice that we can write the Gauss-Seidel equation as:

$$\mathfrak{D}x^{(k)} = b - Lx^{(k)} - Ux^{(k-1)}$$

so that

$$x^{(k)} = \mathfrak{D}^{-1}[b - Lx^{(k)} - Ux^{(k-1)}].$$

We can subtract $x^{(k-1)}$ from both sides to get

$$x^{(k)} - x^{(k-1)} = \mathfrak{D}^{-1}[b - Lx^{(k)} - \mathfrak{D}x^{(k-1)} - Ux^{(k-1)}].$$

Now, think of this as the Gauss-Seidel correction $(x^{(k)}-x^{(k-1)})_{GS}$. As suggested above, it turns out that convergence $x^{(k)}\to x$ of the sequence of approximate solutions to the true solution is often faster if we go beyond the standard Gauss-Seidel correction. The idea of the SOR method is to iterate,

$$x^{(k)}=x^{(k-1)}+\omega(x^{(k)}-x^{(k-1)})_{GS},$$

where, as we just found

$$(x^{(k)}-x^{(k-1)})_{GS}=\mathfrak{D}^{-1}[b-Lx^{(k)}-\mathfrak{D}x^{(k-1)}-Ux^{(k-1)}]$$

and where generally $1<\omega<2$. Notice that if $\omega=1$ then this is the Gauss-Seidel method. Written out in detail, the SOR method is

$$x^{(k)}=x^{(k-1)}+\omega\mathfrak{D}^{-1}[b-Lx^{(k)}-\mathfrak{D}x^{(k-1)}-Ux^{(k-1)}].$$

We can multiply both sides by matrix $\mathfrak{D}$ and divide both sides by ω to rewrite this as

$$\frac{1}{\omega}\mathfrak{D}x^{(k)}=\frac{1}{\omega}\mathfrak{D}x^{(k-1)}+[b-Lx^{(k)}-\mathfrak{D}x^{(k-1)}-Ux^{(k-1)}]$$

then collect the $x^{(k)}$ terms on the left hand side to get,

$$\begin{aligned}\left(L+\frac{1}{\omega}\mathfrak{D}\right)x^{(k)}&=\frac{1}{\omega}\mathfrak{D}x^{(k-1)}+[b-\mathfrak{D}x^{(k-1)}-Ux^{(k-1)}]\\&=\left(\frac{1}{\omega}\mathfrak{D}-\mathfrak{D}-U\right)x^{(k-1)}+b.\end{aligned}$$

When we solve for $x^{(k)}$ we get

$$x^{(k)}=\left(L+\frac{1}{\omega}\mathfrak{D}\right)^{-1}\left[\left(\frac{1}{\omega}\mathfrak{D}-\mathfrak{D}-U\right)x^{(k-1)}+b\right].$$

Notice that the SOR Method is also of the form

$$x^{(k)}=\mathfrak{B}_{SOR}x^{(k-1)}+\mathfrak{C}.$$

The iteration matrix B_{SOR} that determines convergence of the SOR method is

$$\left(L+\frac{1}{\omega}\mathfrak{D}\right)^{-1}\left(\frac{1}{\omega}\mathfrak{D}-\mathfrak{D}-U\right)$$

so optimal convergence is achieved by choosing a value of ω that minimizes

$$\left\|\left(L+\frac{1}{\omega}\mathfrak{D}\right)^{-1}\left(\frac{1}{\omega}\mathfrak{D}-\mathfrak{D}-U\right)\right\|.$$

Thus finally, this method introduces a parameter ω whose role is to minimize the spectral radius, the largest in modulus eigen value, of their iterative matrix. Successive over relaxation is the acceleration of Gauss-Seidal method for a suitable choice of relaxation parameter ω. So the iterative formula for SOR method in component form according to Young [10] is given as

$$x_i^{(k)}=(1-\omega)x_i^{(k-1)}+\frac{\omega}{a_{ii}}\left(b_i-\sum_{i<J}a_{ij}x_j^{(k-1)}-\sum_{i>J}a_{ij}x_j^{(k)}\right)$$

and SOR method in the matrix form is given as

$$x^{(k)}=(\mathfrak{D}+\omega L)^{-1}[-\omega U+(1-\omega)\mathfrak{D}]x^{(k-1)}+(\mathfrak{D}+\omega L)^{-1}\omega bx^{(k)}=\mathfrak{B}_{SOR}x^{(k-1)}+\mathfrak{C}$$

where

$$\mathfrak{B}_{SOR} = (\mathfrak{D} + \omega L)^{-1}[-\omega U + (1-\omega)\mathfrak{D}]$$

is an iteration matrix, and $\mathfrak{C} = \omega(\mathfrak{D} + \omega L)^{-1} b$ is an iteration vector.

3. Convergence Criteria of Iterative Methods

The basic idea of iterative methods is to construct a sequence of vectors $x^{(k)}$ that enjoy the property of convergence

$$x = \lim_{(k\to\infty)} x^{(k)}, \tag{3.1}$$

where x is the solution of linear system of equations. In practice, the iterative process is stopped at the minimum value of n such that $\|x^{(n)} - x\| < \epsilon$ where ϵ is a fixed tolerance, and $\|\cdot\|$ is any convenient vector norm. However, since the exact solution is obviously not available, it is necessary to introduce suitable stopping criteria to monitor the convergence of the iteration. Iterative Methods for solving Linear Systems to start with, we consider iterative methods of the form given $x^{(0)}$,

$$x^{(k+1)} = \mathfrak{B}x^{(k)} + \mathfrak{C}, \quad k \ge 0,$$

where $\mathfrak{B}$ is an $n \times n$ square matrix called the iteration matrix, and $\mathfrak{C}$ is a vector that is obtained from the right hand side b.

3.1 Convergence of Richardson Method

If $\lambda_j \in \mathbb{C}^1$, $j = 1, \dots, n$, are the eigenvalues of $A \in \mathscr{L}(\mathbb{R}^n)$, then, for constant $\omega > 0$, the eigenvalues of the iteration matrix $\mathfrak{B}_{RM} = I - \omega A$ of Richardson's method are $1 - \omega\lambda_1, \dots, 1 - \omega\lambda_n$. Hence, we obtain

$$\rho(\mathfrak{B}_{RM}) = \max_{i=1,\dots,n} |1 - \omega\lambda_i|.$$

Evidently, we are interested in the optimal value of ω for which $\rho(\mathfrak{B}_{RM})$ is minimal. For this suppose that A is symmetrizable and hence that all eigenvalues $\lambda_1, \dots, \lambda_n$ are real. Then with the algebraically smallest and largest values

$$\lambda_{\min} = \max_{i=1,\dots,n} \lambda_j, \quad \lambda_{\max} = \max_{i=1,\dots,n} \lambda_j.$$

We evidently have

$$\rho(\mathfrak{B}_{RM}) = \max\{|1 - \omega\lambda_{\min}|, |1 - \omega\lambda_{\max}|\}.$$

In particular, for the case of positive eigenvalues we obtain the convergence result.

Theorem 3.1. *Let all eigenvalues of $A \in \mathscr{L}(\mathbb{R}^n)$ be real and positive. Then Richardson's method converges if and only if $0 < \omega < \frac{2}{\lambda_{\max}}$, where $\lambda_{\max}$ is the algebraically largest eigenvalue of A. If there is convergence, then $\rho(\mathfrak{B}_{RM})$ is minimal for*

$$\omega_{opt} = \frac{2}{(\lambda_{\max} + \lambda_{\min})}, \quad \rho(\mathfrak{B}_{opt}) = \frac{(\lambda_{\max} - \lambda_{\max})}{(\lambda_{\max} + \lambda_{\min})}.$$

Theorem 3.2. *Let $A \in \mathscr{L}(\mathbb{R}^n)$ be symmetric and positive-definite. Then Richardson's method converges if and only if,* $0 < \omega < \frac{2}{\|A\|_2}$.

3.2 Convergence of Jacobi Method

Theorem 3.3. *If A is a strictly diagonally dominant matrix by rows, the Jacobi and Gauss-Seidel methods are convergent.*

Proof. Let us prove the part of the theorem concerning the Jacobi method, while for the Gauss-Seidel method we refer to [9]. Since A is strictly diagonally dominant by rows, $|a_{i,i}| > \sum_{j=1}^{n} |a_{i,j}|$ for $j \neq i$ and $i = 1, \dots, n$. As a consequence

$$\|\mathfrak{B}\|_\infty = \max_{i=1,\dots,n} \sum_{j=1, j\neq i}^{n} \frac{|a_{i,j}|}{|a_{i,i}|} < 1$$

so that the Jacobi method is convergent. □

Theorem 3.4. *If A and $2D - A$ are symmetric and positive definite matrices, then the Jacobi method is convergent and $\rho(\mathfrak{B}_J) = \|\mathfrak{B}_J\|_A = \|\mathfrak{B}_J\|_{\mathfrak{D}}$.*

3.3 Convergence of Gauss-Seidal Method

Theorem 3.5. *If A is symmetric positive definite, the Gauss-Seidel method is monotonically convergent with respect to the norm $\|\cdot\|_A$.*

3.4 Convergence of SOR Method

Theorem 3.6 ([3]). *Let $A \in \mathscr{L}(\mathbb{R}^n)$ be symmetric, positive definite. Then the SOR method converges for* $0 < \omega < 2$.

Proof. For a symmetric, positive definite A the decomposition becomes $A = \mathfrak{D} - L - L^t$ where the diagonal elements of $\mathfrak{D}$ are strictly positive. Thus, in particular, $\mathfrak{B}$ is nonsingular and so must be and $\mathfrak{D} - \omega L$. For $0 < \omega < 2$ the matrices

$$\mathscr{B} = \frac{1}{\omega}(\mathfrak{D} - \omega L), \quad \mathscr{C} = \frac{1}{\omega}[(1-\omega)\mathfrak{D} + \omega L^t]$$

are well defined and real, and we can write

$$\begin{aligned}
\mathfrak{B}_{SOR} &= \mathscr{B}^{-1}\mathscr{C} \\
&= \mathscr{B}^{-1}\left[\frac{1}{\omega}\mathfrak{D} - \mathfrak{D} + L^t\right] \\
&= \mathscr{B}^{-1}\left[\frac{1}{\omega}(\mathfrak{D} - \omega L) + L + L^t - \mathfrak{D}\right] \\
&= \mathscr{B}^{-1}(\mathscr{B} - A) \\
&= I - \mathscr{B}^{-1}A\,.
\end{aligned}$$

The following short calculation shows that

$$A - \mathfrak{B}_{SOR}^t A \mathfrak{B}_{SOR} = A - (I - \mathscr{B}^{-1}A)^t A (I - \mathscr{B}^{-1}A)$$

$$
\begin{aligned}
&= (\mathscr{B}^{-1}A)^t A + A\mathscr{B}^{-1}A - (\mathscr{B}^{-1}A)^t A(\mathscr{B}^{-1}A) \\
&= (\mathscr{B}^{-1}A)^t[\mathscr{B}^t + \mathscr{B} - A]\mathscr{B}^{-1}A\,, \qquad (3.2)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathscr{B}^t + \mathscr{B} - A &= \frac{2}{\omega}\mathfrak{D} - L - L^t - (\mathfrak{D} - L - L^t) \\
&= \frac{1}{\omega}(2-\omega)\mathfrak{D}\,.
\end{aligned}
$$

Since $0 < \omega < 2$ this proves that $\mathscr{B}^t + \mathscr{B} - A$ is symmetric positive definite, whence (3.2) implies the same for $A - \mathfrak{B}^t_{SOR} A \mathfrak{B}_{SOR}$. Let now $\lambda \in \mathbb{C}^1$ be any eigenvalue of $\mathfrak{B}_{SOR}$ and $u \in \mathbb{C}^n$, $u \neq 0$ a corresponding eigenvector. Then

$$
u^* A u > u^* \mathfrak{B}^t_{SOR} A \mathfrak{B}_{SOR} u = (\lambda u)^* A(\lambda u) = |\lambda|^2 u^* A u
$$

so that $|\lambda|^2 < 1$ and hence also

$$
\rho(\mathfrak{B}_{SOR}) < 1
$$

as claimed. □

Theorem 3.7 ([5]). *If all diagonal elements of $A \in \mathscr{L}(\mathbb{C}^n)$ are nonzero, then the spectral radius of the SOR iteration matrix satisfies,*

$$
\rho(\mathfrak{B}_{SOR}) \geq |\omega - 1|. \qquad (3.3)
$$

Hence a necessary condition for the SOR method to converge is $|\omega - 1| < 1$ (for $\omega \in \mathbb{R}$ this condition becomes $\omega \in (0,2)$).

Proof. Because L is strictly lower triangular we have

$$
\det \mathfrak{D}^{-1} = \det(\mathfrak{D} + \omega L)^{-1}
$$

and

$$
\begin{aligned}
\det(\mathfrak{B}_{SOR}) &= (\mathfrak{D} + \omega L)^{-1} \det((1-\omega)\mathfrak{D} + \omega U) \\
&= \det((1-\omega)I + \omega \mathfrak{D}^{-1} U) \\
&= \det((1-\omega)I) \\
&= (1-\omega)^n\,,
\end{aligned}
$$

since $\mathfrak{D}^{-1}U$ is strictly upper triangular. Since $\det(\mathfrak{B}_{SOR})$ is the product of the eigenvalues of $\mathfrak{B}_{SOR}$ it follows that

$$
\rho(\mathfrak{B}_{SOR})^n \geq \det(\mathfrak{B}_{SOR}) = (1-\omega)^n
$$

and therefore that (3.3) holds. □

In the positive-definite case the necessary condition is also sufficient for convergence [3], as the following theorem shows.

Theorem 3.8. *If $A \in \mathbb{C}^{n \times n}$ is hermetian positive definite, then the (pointwise or block) forward SOR method converges for all $\omega \in (0,2)$.*

Proof. We have

$$\begin{aligned}\mathfrak{B}_{SOR} &= (\mathfrak{D}+\omega L)^{-1}[-\omega U+(1-\omega)\mathfrak{D}]\\ &= I-\left(\frac{1}{\omega}(\mathfrak{D}-\omega L)\right)^{-1}A\\ &= I-\mathscr{B}^{-1}A,\end{aligned}$$

where

$$\mathscr{B}=\frac{1}{\omega}(\mathfrak{D}-\omega L).$$

Let $\lambda\in\mathbb{C}$ be an eigenvalue of $\mathfrak{B}_{SOR}$ with corresponding eigenvector $v\in\mathbb{C}^n$. Then

$$Av=(1–\lambda)\mathscr{B}v.$$

Since

$$v^*Av>0,\quad \lambda\neq 1.$$

Hence

$$\frac{(v^*\mathscr{B}v)}{(v^*Av)}=\frac{1}{(1-\lambda)}.$$

Consequently,

$$\begin{aligned}2Re\frac{1}{(1-\lambda)} &= \frac{1}{(1-\lambda)}+\frac{1}{(1-\bar{\lambda})})\\ &= \frac{(v^*\mathscr{B}v)}{(v^*Av)}+\frac{\overline{(v^*\mathscr{B}v)}}{\overline{(v^*Av)}}\\ &= \frac{(v^*(\mathscr{B}+\mathscr{B}^*)v)}{(v^*Av)}.\end{aligned}$$

Since $\mathscr{B}=\frac{1}{\omega}(\mathfrak{D}-\omega L)$ and $\mathfrak{D}^*=\mathfrak{D}$ and $L^*=U$ we have

$$\begin{aligned}\mathscr{B}+\mathscr{B}^* &= \omega^{-1}\mathfrak{D}-L+(\omega^{-1}\mathfrak{D}-L)^*\\ &= 2\omega^{-1}\mathfrak{D}–L-U^t\\ &= A+(2\omega^{-1}-1)\mathfrak{D}\end{aligned}$$

and

$$\begin{aligned}2Re\frac{1}{(1-\lambda)} &= \frac{(v^*(\mathscr{B}+\mathscr{B}^*)v)}{(v^*Av)}\\ &= 1+\left(\frac{2}{\omega}–1\right)\frac{(v^*\mathfrak{D}v)}{(v^*Av)}.\end{aligned}$$

If $A\in\mathbb{C}^{n\times n}$ is hermetian positive definite, its (block) diagonal $\mathfrak{D}$ is hermetian positive definite. Hence

$$2Re\frac{1}{(1–\lambda)}=1+\left(\frac{2}{\omega}–1\right)\frac{(v^*\mathfrak{D}v)}{(v^*Av)}>1.$$

If we set $\lambda=\alpha+i\beta$, then

$$1<2Re\frac{1}{(1–\lambda)}=2\frac{(1-\alpha)}{(1-\alpha)^2+\beta^2}$$

which implies

$$|\lambda|^2 = \alpha^2 + \beta^2 < 1.$$

Since λ was an arbitrary eigenvalue of $\mathfrak{B}_{SOR}$, $\rho(\mathfrak{B}_{SOR}) < 1$. □

Definition 3.1. Let the $N \times N$ matrix A be partitioned into the form

$$A = \begin{bmatrix} A_{1,1} & \dots & A_{1,q} \\ \vdots & \ddots & \vdots \\ A_{q,1} & \dots & A_{q,q} \end{bmatrix},$$

where $A_{i,j}$ is an $n_i \times n_j$ submatrix and $n_1 + \ldots + n_q = N$. The $q \times q$ block matrix of A is said to have A-property if there exists two disjoint nonempty subsets S_R and S_B of $\{1,2,\ldots,q\}$ such that $S_R \cup S_B = \{1,2,\ldots,q\}$ and such that if $A_{i,j} \neq 0$ and $i \neq j$ then $i \in S_R$ and $j \in S_B$ or $j \in S_R$ and $i \in S_B$.

Theorem 3.9. *If the matrix A enjoys the A-property and if $\mathfrak{B}_J$ has real eigenvalues, then the SOR method converges for any choice of $x^{(0)}$ iff*

$$\rho(\mathfrak{B}_J) < 1 \quad \text{and} \quad 0 < \omega < 2.$$

Moreover,

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - (\rho(\mathfrak{B}_J))^2}}$$

and the corresponding asymptotic convergence factor is

$$\rho(\mathfrak{B}_{opt}) = \frac{1 - \sqrt{1 - (\rho(\mathfrak{B}_J))^2}}{1 + \sqrt{1 - (\rho(\mathfrak{B}_J))^2}}.$$

Theorem 3.10. *Let $A = \mathfrak{D} + L + U$ be a matrix that satisfies the technical assumption*

$$\det(\mathfrak{D}k - \gamma L - \gamma^{-1}U) = \det(k\mathfrak{D} - L - U), \quad \text{for all } k, \gamma \in \mathbb{R} - \{0\} \tag{3.4}$$

and $\mathfrak{B}_J$ and $\mathfrak{B}_{SOR}$ the iteration matrices of the Jacobi and SOR methods as defined above. If μ is an eigenvalue of $\mathfrak{B}_J$ and $\lambda \neq 0$ satisfies,

$$\mu = \frac{\lambda + \omega - 1}{\omega \lambda^{1/2}} \tag{3.5}$$

for some $\omega \in (0,2)$ then λ is an eigenvalue of $\mathfrak{B}_{SOR}$.

Proof. We have the definitions for $\mathfrak{B}_J$ and $\mathfrak{B}_{SOR}$ that

$$\mathfrak{B}_J = -\mathfrak{D}^{-1}(L+U), \quad \mathfrak{B}_{SOR} = (\mathfrak{D} + \omega L)^{-1}[-\omega U + (1-\omega)\mathfrak{D}].$$

Let λ be one eigenvalue of $\mathfrak{B}_{SOR}$, $\lambda \neq 0$. Because U and L are strictly upper and lower triangular matrices, and their main diagonals are zeros, we have

$$\det(-\mathfrak{D}^{-1})\det(\mathfrak{D} - \omega L) = 1$$

$$\begin{aligned} \det(\mathfrak{B}_{SOR} - \lambda I) &= \det\left[(\mathfrak{D} + \omega L)^{-1}\{(1-\omega)\mathfrak{D} + \omega U\} - \lambda I\right] \\ &= \det(\mathfrak{D}^{-1})\det(\mathfrak{D} - \omega L)\det\left[(\mathfrak{D} + \omega L)^{-1}\{(1-\omega)\mathfrak{D} + \omega U\} - \lambda I\right] \end{aligned}$$

$$
\begin{aligned}
&= \det\left[(1-\omega)I + \omega\mathfrak{D}^{-1}U - \lambda I + \omega\lambda\mathfrak{D}^{-1}L\right] \\
&= \omega^n \det\left[\left(\frac{1}{\omega} - 1\right)I + \mathfrak{D}^{-1}U - \frac{\lambda I}{\omega} + \lambda\mathfrak{D}^{-1}L\right] \\
&= (-1)^n\omega^n\lambda^{1/2}\det(\mathfrak{D}^{-1})\det\left[\frac{(\omega+\lambda-I)}{(\omega\lambda^{1/2}})\mathfrak{D} - \lambda^{-1/2}U - \lambda^{1/2}L\right] \\
&= (-1)^n\omega^n\lambda^{1/2}\det\left[\frac{\omega+\lambda-I}{\omega\lambda^{1/2}}I - \mathfrak{D}^{-1}U - \mathfrak{D}^{-1}L\right] \\
&= \omega^n\lambda^{1/2} det\left[\mathfrak{B}_J - \frac{\omega+\lambda-I}{\omega\lambda^{1/2}}I\right].
\end{aligned}
$$

Because $\lambda \neq 0$ satisfies

$$\mu = \frac{(\omega+\lambda-I)}{(\omega\lambda^{1/2})},$$

where μ is an eigenvalue of $\mathfrak{B}_J$. In turn, when μ is an eigenvalue of $\mathfrak{B}_J$, and λ satisfies that relation, then λ is an eigenvalue of $\mathfrak{B}_{SOR}$. □

Theorem 3.11. *Assume $A = \mathfrak{D} + L + U$ is a matrix that satisfies (3.4), and assume that $\mathfrak{B}_J = I - \mathfrak{D}^{-1}A$ has only real eigenvalues and that $\beta = \rho(\mathfrak{B}_J) < 1$. Then the SOR iteration converges for every $\omega \in (0,2)$, and the spectral radius of the SOR matrix is*

$$\rho(\mathfrak{B}_{SOR}) = f(x) = \begin{cases} \frac{1}{4}[\omega\beta + \rho(\omega\beta)^2 - 4(\omega-1)]^2, & \text{for } 0 < \omega \leq \omega_{opt}, \\ \omega - 1, & \text{for } \omega_{opt} \leq \omega < 2, \end{cases} \tag{3.6}$$

where ω_{opt}, the optimal value of ω is

$$\omega_{opt} = \frac{2}{1+\sqrt{1-\beta^2}}. \tag{3.7}$$

For any other value of ω we have

$$\rho(\mathfrak{B}_{SORopt}) < \rho(\mathfrak{B}_{SOR}), \quad \text{for } \omega \in (0,2) - \{\omega_{opt}\}.$$

Proof. For a given ω, $\rho(\mathfrak{B}_{SOR})$ is the largest eigenvalue of $\mathfrak{B}_{SOR}$ in absolute value. Suppose that μ is an eigenvalue of $\mathfrak{B}_J$. From (3.5) we have

$$\lambda = \frac{1}{4}\left(\omega\mu \pm \sqrt{(\omega\mu)^2 - 4(\omega-1)}\right)^2. \tag{3.8}$$

Equation (3.8) gives two eigenvalues for $\mathfrak{B}_{SOR}$. First, if

$$(\omega\mu)^2 - 4(\omega-1) < 0,$$

then λ is imaginary, and the absolute value of λ is

$$|\lambda| = \left(\frac{1}{4}\omega^2\mu^2 + \omega - 1 - \frac{1}{4}\omega^2\mu^2\right) = \omega - 1$$

when

$$\widetilde{\omega} \equiv \frac{(1-\sqrt{1-\mu^2})}{\mu^2} < \omega < 2.$$

Here, $\rho(\mathfrak{B}_{SOR}) = |\lambda|$ is independent of $\mu = \rho(\mathfrak{B}_J)$. Second, if $(\omega\mu)^2 - 4(\omega-1) \geq 0$, then

$$\rho\mathfrak{B}_{SOR}) = \max_{\mu\in\rho(\mathfrak{B}_J)} \frac{1}{4}\left[\omega|\mu| \pm \sqrt{(\omega|\mu|)^2 - 4(\omega-1)}\right]^2,$$

$\omega \in (0,\widetilde{\omega}]$. On one hand, for a fixed $\omega, \rho(\mathfrak{B}_{SOR})$ is an increasing function with respect to variable $|\mu|$. Hence, to get the spectral radius of $\mathfrak{B}_{SOR}$, let

$$|\mu| = \rho(\mathfrak{B}_J) = \beta,$$

we have

$$\rho(\mathfrak{B}_{SOR}) = \frac{1}{4}\left[\omega\beta \pm \sqrt{(\omega\beta)^2 - 4(\omega-1)}\right]^2.$$

On the other hand, $\rho(\mathfrak{B}_{SOR})$ can be proved to be a decreasing function with respect of ω in $(0,\widetilde{\omega}]$. Details are as follows: We have the first order derivative of $\rho(\mathfrak{B}_{SOR})$

$$\rho'(\mathfrak{B}_{SOR}) = \frac{1}{2}\left(\omega\beta + \sqrt{(\omega\beta)^2 - 4(\omega-1)\beta} + \frac{\beta^2\omega - 2}{\sqrt{(\omega\beta)^2 - 4(\omega-1)}}\right).$$

To determine its sign, we need to examine the sign of

$$\left(\beta + \frac{\beta^2\omega - 2}{\sqrt{(\omega\beta)^2 - 4(\omega-1)}}\right).$$

Because

$$\sqrt{(\omega\beta)^2 - 4(\omega-1)} > 0, \quad \beta < 1, \quad \omega < 2$$

and

$$\begin{aligned}\beta\sqrt{(\omega\beta)^2 - 4(\omega-1)} + (\beta^2\omega - 2) < \sqrt{\omega)^2 - 4\omega + 4} + \omega - 2 &= \sqrt{(\omega-2)^2} + \omega - 2 \\ &= 2 - \omega + \omega - 2 \\ &= 0.\end{aligned}$$

Therefore, $\rho'(\mathfrak{B}_{SOR}) < 0$ in the interval $(0,\widetilde{\omega}]$, which implies that $\rho(\mathfrak{B}_{SOR})$ is a decreasing function of ω. When $\omega = \widetilde{\omega}$, $\rho(\mathfrak{B}_{SOR})$ gets its minimum in the interval $(0,\widetilde{\omega}]$. We have proved above that in the interval $(\widetilde{\omega},2)$, $\rho(\mathfrak{B}_{SOR}) = \omega - 1$ is an increasing function and also gets its minimum when ω approaches to $\widetilde{\omega}$. Moreover, $\rho(\mathfrak{B}_{SOR})$ is continuous at the point $\widetilde{\omega}$. Considering that the optimal parameter ω is the very number that makes $\rho(\mathfrak{B}_{SOR})$ gets its minimum. Therefore,

$$\widetilde{\omega} = \omega_{opt} = \frac{2(1-\sqrt{1-\beta^2})}{\beta^2} = \frac{2}{(1+\sqrt{1-\beta^2})}.$$

According to the statement above, for any value of ω,

$$\rho(\mathfrak{B}_{SORopt}) < \rho(\mathfrak{B}_{SOR}), \quad \omega \in (0,2) - \{\omega_{opt}\}.$$ □

4. Numerical Examples

Example 4.1. Let us consider following system of equations,

$$\left.\begin{aligned}2x + y &= 1 \\ x + 7y &= 5\end{aligned}\right\}. \tag{4.1}$$

Solutions. *Exact solution:*

$$x = 0.153846153846154,$$
$$y = 0.692307692307692.$$

Approximate solution:

Methods	Noumber of iterations	Approximate solution	Error vector
Richardson method	37	0.153846152408944 0.692307685840244	1.0e-08* 0.143721012868525 0.646744535703903
Jacobi method	15	0.153846157129925 0.692307692516186	1.0e-08* −0.328377147695846 −0.020849333370876
Gauss-Seidal method	8	0.153846157129925 0.692307691838582	1.0e-08* −0.328377142144731 0.046911019513374
SOR method	6	0.153846158504998 0.692307691940860	1.0e-08* −0.465884453237919 0.036683256432468

Example 4.2. Let us consider following system of equations:

$$\left.\begin{array}{l} 25x - 15y - 5z = 1 \\ -15x + 18y + 0z = 2 \\ -5x + 0y + 11z = 3 \end{array}\right\}. \tag{4.2}$$

Solutions. *Exact solution:*

$$x = 0.068148148148148,$$
$$y = 0.054320987654321,$$
$$z = 0.303703703703704.$$

Approximate solution:

Methods	Number of iterations	Approximate solution	Error vector
Richardson method	57	0.068148090688256 0.054321117273847 0.303703626032756	1.0e-06* 0.057459892177647 −0.129619525786329 0.077670947618014
Jacobi method	50	0.068148015875257 0.054321064702911 0.303703661677200	1.0e-06* 0.132272891056884 −0.077048590389761 0.042026503832560
Gauss-Seidal method	25	0.068148055689840 0.054321064702911 0.303703661677200	1.0e-07* 0.924583084538355 −0.770485904175167 0.420265038325596
SOR method	10	0.06814807768745 0.054321000057282 0.303703609196386	1.0e-07* 0.704606918366935 −0.124029609013809 0.945073176383815

Example 4.3. Let us consider following system of equations:

$$\left.\begin{array}{l} 4x-1y-2z+2t=1 \\ -1x+4y-1z-2t=-1 \\ -2x-1y+4z-1t=2 \\ 2x-2y-1z+4t=-2 \end{array}\right\}. \tag{4.3}$$

Solutions. *Exact solution*:

$x=1.000000000000000,$

$y=-0.333333333333333,$

$z=0.666666666666667,$

$t=-1.000000000000000.$

Approximate solution:

Methods	Number of iterations	Approximate solution	Error vector
Richardson method	25	0.999896322221324 −0.333379047079772 0.666616332563392 −1.000088417776678	1.0e-03* 0.103677778676259 0.045713746438303 0.050334103274818 0.088417776678118
Jacobi method	30	1.024445888756703 −0.348448414847807 0.651551544317527 −0.975553976375561	−0.024445888756704 0.015115081514474 0.015115122349139 −0.024446023624439
Gauss-Seidal method	15	1.000150581564877 −0.333263485847253 0.666751824240107 −1.000019077646038	1.0e-03* −0.150581564877572 −0.069847486080132 −0.085157573440187 0.019077646038590
SOR method	7	1.000200692595382 −0.334580690589569 0.665736816742371 −1.000698016266749	−0.000200692595382 0.001247357256236 0.000929849924295 0.000698016266749

Example 4.4. Let us consider following system of equations:

$$\left.\begin{array}{l} 4x-1y+0z+0t=1 \\ -1x+4y-1z+0t=0 \\ 0x-1y+4z-1t=0 \\ 0x+0y-1z+4t=0 \end{array}\right\}. \tag{4.4}$$

Solutions. *Exact solution*:

$x=0.267942583732057,$

$y=0.071770334928230,$

$z=0.019138755980861,$

$t=0.004784688995215.$

Approximate solution:

Methods	Number of iterations	Approximate solution	Error vector
Richardson method	20	0.267942582595424 0.071770334184293 0.019138754141750 0.004784688535437	1.0e-08* 0.113663339851300 0.074393620430602 0.183911142911941 0.045977785727985
Jacobi method	20	0.267942582595424 0.071770330381696 0.019138754141750 0.004784686185303	1.0e-08* 0.113663339851300 0.454653355241863 0.183911142911941 0.280991216666110
Gauss-Seidal method	8	0.267942576785572 0.071770330381696 0.019138754141750 0.004784688535437	1.0e-08* 0.694648555343846 0.454653355241863 0.183911142911941 0.045977785727985
SOR method	6	0.267942575355385 0.071770332194892 0.019138755288967 0.004784688904958	1.0e-08* 0.837667218922533 0.273333805589360 0.069189434390160 0.009025772750720

5. Conclusion

We have concluded that the wrong choice of relaxation parameter may lead to poor convergence. Finding the value of the relaxation parameter is a very complex process for a very large system of equations. In particular, if coefficient matrix A is symmetric and positive definite, the SOR method converges more faster for $\omega_{opt} = \frac{2}{1+\sqrt{1-\beta^2}}$. We have also concluded that, with the help of some typical examples, the SOR method converges more rapidly as compared to Richardson, Jacobi and Gauss-Seidel methods by taking a carefully suitable choice of relaxation parameter $\omega \in (0,2)$.

Acknowledgement

Manuscript communication number (MCN): IU/R&D/2020-MCN000998 office of research and development Integral University, Lucknow.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] F. Curtis, G. Patrick and O. Wheatley, *Applied Numerical Analysis*, 5th edition, Addison-Wesley, New York (1998), URL: `http://www.cse.iitm.ac.in/~vplab/downloads/opt/Applied%20Numerical%20Analysis.pdf`.

[2] T. K. Enyew, G. Awgichew, E. Haile and G. D. Abie, Second-refinement of Gauss-Seidel iterative method for solving linear system of equations, *Ethiopian Journal of Science and Technology* **13**(1) (2020), 1 – 15, DOI: 10.4314/ejst.v13i1.1.

[3] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer-Verlag, Heidelberg — New York (1994), DOI: 10.1007/978-1-4612-4288-8.

[4] L. A. Hageman and D. M. Young, *Applied Iterative Methods*, Academic Press, New York (1981), URL: `https://www.math.hkust.edu.hk/~mamu/courses/531/Applied_Iterative_Methods.pdf`.

[5] W. Kahan, *Gauss-Seidel Methods for Solving Large Systems of Linear Equations*, Ph.D. Thesis, University of Toronto, Toronto, Ontario, Canada (1958).

[6] S. Karunanithi, N. Gajalakshmi, M. Malarvizhi and M. Saileshwari, A study on comparison of Jacobi, Gauss-Seidel and SOR methods for the solution in system of linear equations, *International Journal of Mathematics Trends and Technology* **56**(4) (2018), 214 – 222, DOI: 10.14445/22315373/IJMTT-V56P531.

[7] M. Louka, *Iterative Methods for the Numerical Solution of Linear Systems*, PhD Thesis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece, URL: `https://www.di.uoa.gr/sites/default/files/documents/grad/phdbook/Louka.pdf`.

[8] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd edition, SIAM – Society for Industrial and Applied Mathematics (2003), DOI: 10.1137/1.9780898718003.

[9] J. Xu, A new class of iterative methods for nonselfadjoint or indefinite problems, *SIAM Journal on Numerical Analysis* **29**(2) (1992), 303 – 319, DOI: 10.1137/0729020.

[10] D. M. Young, *Iterative Solution of Large Linear Systems*, Dover Publications (1971), URL: `http://gen.lib.rus.ec/book/index.php?md5=1b9a661e5563daba113b7ffc30d26c18`.

[11] D. M. Young, Iterative methods for solving large systems of linear equations, *Acta Universitatis Carolinae. Mathematica et Physica* **15**(1-2) (1974), 179 – 188, URL: `http://dml.cz/dmlcz/142353`.