A Generalization of the SOR Method for Solving Linear System of Equations

D. Khojasteh Salkuyeh

Department of Mathematics, University of Mohaghegh Ardabili, P.O. Box179, Ardabil, Iran

Abstract

The successive over relaxation (SOR) algorithm is a stationary iterative method for solving linear system of equations. It is also a promising preconditioner for popular iterative solvers such as GMRES method. In this paper, a generalization of the SOR method say GESOR is proposed and its convergence properties are discussed. Some numerical experiments are also given to show the efficiency of the proposed method. Numerical results presented here show that the GESOR method is often more effective than the SOR method.

Keywords: SOR, Generalization, Symmetric Positive Definite, Diagonally Dominant, Convergence.

1 Introduction

Consider the linear system of equations

\[ Ax = b, \]  

where the matrix \( A \in R^{n \times n} \) and \( x, b \in R^n \). Let \( A \) be a nonsingular matrix with nonzero diagonal entries and

\[ A = D - E - F, \]

where \( D \) is the diagonal of \( A \), \( -E \) its strict lower part, and \( -F \) its strict upper part. Then the Jacobi and the Gauss-Seidel methods for solving Eq. (1) are defined as

\[ x_{k+1} = D^{-1}(E + F)x_k + D^{-1}b, \]
\[ x_{k+1} = (D - E)^{-1}F x_k + (D - E)^{-1}b, \]

1 Corresponding author
E-mail address: khojaste@uma.ac.ir
respectively. In the successive over relaxation (SOR) method the system $Ax = b$ is written as $\omega Ax = \omega b$, where $\omega$ is a parameter. Then the coefficient matrix $\omega A$ is decomposed in the form

$$\omega A = (D - \omega E) - ((1 - \omega)D + \omega F).$$

Then the system $\omega Ax = \omega b$ is written as follows

$$x = (D - \omega E)^{-1}((1 - \omega)D + \omega F)x + \omega (D - \omega E)^{-1}b,$$

and the SOR method is defined as

$$x_{k+1} = (D - \omega E)^{-1}((1 - \omega)D + \omega F)x_k + \omega (D - \omega E)^{-1}b.$$

Although there are many iterative methods such as GMRES [7] and Bi-CGSTAB[8] algorithms for solving Eq. (1) which are more effective than these three stationary iterative methods, they have been used as preconditioner for common iterative solvers (see for example [9,10,11]). In [2], we have proposed the generalized Jacobi (GJ) and Gauss-Seidel (GGS) methods and studied their convergence properties. In this paper, we propose a generalization of the SOR (say GESOR) method and verify its convergence properties.

This paper is organized as follows. In section 2, we review the GJ and GGS methods and propose the GESOR method and study its convergence properties. Some numerical experiments are given in section 3. Concluding remarks and future works are given in section 4.

2 The GJ, GGS and The Gesor Methods

Let $A = (a_{ij})$ be an $n \times n$ matrix and $T_m = (t_{ij})$ be a banded matrix of bandwidth $2m + 1$ defined as

$$t_{ij} = \begin{cases} a_{ij}, & |i - j| \leq m, \\ 0, & \text{otherwise}. \end{cases}$$

We consider the decomposition $A = T_m - E_m - F_m$, where $- E_m$ and $- F_m$ are the strict lower and upper part of the matrix $A - T_m$, respectively. In other words matrices $T_m$, $E_m$ and $F_m$ are defined as following

$$T_m = \begin{pmatrix} a_{11} & \cdots & a_{1,m+1} \\ \vdots & \ddots & \vdots \\ a_{m+1,1} & \cdots & a_{m+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,m} & \cdots & a_{n,n} \end{pmatrix}.$$
Then the GJ and GGS methods are defined as
\[ x_{k+1} = T_m^{-1}(E_m + F_m)x_k + T_m^{-1}b, \]
\[ x_{k+1} = (T_m - E_m)^{-1}F_m x_k + (T_m - E_m)^{-1}b, \]
respectively. In [2] the convergence properties of these methods have been studied. It has been proved that for any natural number \( m \leq n \), both of the GJ and GGS algorithms are convergent to the solution of Eq. (1) for any initial guess \( x_0 \) if \( A \) is symmetric positive definite (SPD) or M-matrix. In general, it can not be deduced the larger \( m \) results in smaller spectral radius of the iteration matrix for the GJ and GGS algorithms. However this is true for some special cases (see [2]).

Here similar to the GJ and GGS methods, we define the GESOR method as follows. First, let \( \omega \) be a fix parameter such that the matrix \( T_m - \omega E_m \) be nonsingular. As the SOR method the GESOR iteration is defined as
\[ x_{k+1} = (T_m - \omega E_m)^{-1}((1 - \omega)T_m + \omega F_m)x_k + \omega(T_m - \omega E_m)^{-1}b. \]

Let
\[ B^{(m)}_{GESOR} = (T_m - \omega E_m)^{-1}((1 - \omega)T_m + \omega F_m), \]
be the iteration matrix of the GESOR method. For the convergence of the method it is enough to show that
\[ \rho(B^{(m)}_{GESOR}) = \max_{\lambda \in \sigma(B^{(m)}_{GESOR})} |\lambda| < 1, \]
where \( \sigma(X) \) is the set of all eigenvalues of the matrix \( X \). As we know well the SOR method is convergent for the symmetric positive definite (SPD) matrices if \( 0 < \omega < 2 \) (see [1, 5, 6]). We show that with an additional condition a similar theorem can be stated.

**Theorem 1** Let \( A \) and \( T_m \) be SPD matrices. Then for every \( 0 < \omega < 2 \), the GESOR method converges with any initial guess \( x_0 \).
Proof. The proof is similar to the Ostrowski and Reich theorem [5] for the SOR method. Let \(0 < \omega < 2\). Obviously \(F_m = E_m^T\), since \(A\) is symmetric. Assume that

\[
B = \frac{1}{\omega} T_m - E_m.
\]

First of all, we show that the matrix \(B\) is nonsingular. By contradiction, let \(B\) be singular. Then, there would exist a nonzero vector \(x\) such that \(Bx = 0\). In this case \(x^T B x = 0\). Matrix \(A\) is an SPD matrix. Hence

\[
0 < x^T A x = x^T (T_m - E_m - E_m^T) x = x^T T_m x - 2 x^T E_m x \implies x^T E_m x < \frac{1}{2} x^T T_m x. \tag{2}
\]

Now by using this inequality we have

\[
x^T B x = \frac{1}{\omega} x^T (T_m - \omega E_m) x = \frac{1}{\omega} (x^T T_m x - \omega x^T E_m x) > \frac{1}{\omega} (1 - \frac{\omega}{2}) x^T T_m x.
\]

The matrix \(T_m\) is SPD. Hence \(x^T T_m x > 0\). On the other hand \(1 - \frac{\omega}{2} > 0\).

Therefore \(x^T B x > 0\), which is a contradiction. Now we have

\[
B + B^T - A = \frac{1}{\omega} T_m - E_m + \frac{1}{\omega} T_m - E_m^T - (T_m - E_m - E_m^T) = (\frac{2}{\omega} - 1) T_m.
\]

Obviously \(\frac{2}{\omega} - 1 > 0\) and \(T_m\) is SPD then as a result we see that the matrix \(B + B^T - A\) is SPD. Let \(Q = A^{-1} (2B - A)\). We show that if \(\lambda\) is an eigenvalue of \(Q\), then the real part of \(\lambda\) is positive, i.e., \(\text{Re}(\lambda) > 0\). Let \((\lambda, x)\) be an eigenpair of \(Q\). In this case

\[
A^{-1} (2B - A) x = \lambda x \implies (2B - A) x = \lambda A x \implies x^T (2B - A) x = \lambda x^T A x. \tag{3}
\]

By taking transpose of the latter equation gives, because \(A = A^T\),

\[
x^T (2B^T - A) x = \overline{\lambda} x^T A x. \tag{4}
\]

By adding two sides of Eqs. (3) and (4), it follows that

\[
x^T (B + B^T - A) x = \text{Re}(\lambda) x^T A x.
\]

Both of the matrices \(A\) and \(B + B^T - A\) are SPD, hence we conclude that \(\text{Re}(\lambda) > 0\). It can be easily seen that \(Q + I\) is nonsingular. Therefore

\[
(Q - I)(Q + I)^{-1} = 2 (A^{-1} B - I) \times \frac{1}{2} (A^{-1} B)^{-1}
\]

\[
= I - B^{-1} A
\]

\[
= I - \left(\frac{1}{\omega} T_m - E_m\right)^{-1} (T_m - E_m - F_m)
\]

\[
= I - \omega (T_m - \omega E_m)^{-1} (T_m - E_m - F_m)
\]

\[
= (T_m - \omega E_m)^{-1} (T_m - \omega E_m - \omega T_m + \omega E_m + \omega F_m)
\]

\[
= (T_m - \omega E_m)^{-1} ((1 - \omega) T_m + \omega F_m)
\]

\[
= B(m)_{GSOB}.
\]

Let \((\mu, x)\) be an eigenpair of the matrix \(B(m)_{GSOB}\). Then
\[(Q - I)(Q + I)^{-1}x = \mu x. \quad (5)\]

By setting \(y = (Q + I)^{-1}x\), we see that \(y \neq 0\). Hence \(x = (Q + I)y\) and from Eq. (5) we have

\[(Q - I)y = \mu (Q + I)y.\]

Therefore

\[(1 - \mu)Qy = (1 + \mu)y.\]

We have \(\mu \neq 1\), since \(y \neq 0\). Hence

\[Qy = \frac{1 + \mu}{1 - \mu} y.\]

This relation shows that \(\lambda = \frac{1 + \mu}{1 - \mu}\) is an eigenvalue of \(Q\). As a result we have \(\mu = \frac{\lambda - 1}{\lambda + 1}\). So

\[|\mu|^2 = |\mu|^2 = \frac{|\lambda|^2 + 1 - 2 \text{Re}(\lambda)}{|\lambda|^2 + 1 + 2 \text{Re}(\lambda)}.\]

Having in mind that \(\text{Re}(\lambda) > 0\) we conclude that

\[|\mu| < 1 \Rightarrow \rho(B_{\text{GESOR}}^{(m)}) < 1.\]

This relation completes the proof. \(\square\)

**Definition 1** A matrix \(A\) is said to be irreducible if there exists a permutation matrix \(P\) such that \(PAP^T\) be a block upper triangular matrix.

**Definition 2** A matrix \(A = (a_{ij})\) is said to be strictly diagonally dominant (SDD) if

\[|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \cdots, n.\]

Matrix \(A\) is said to be weakly diagonally dominant (WDD) if

\[|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \cdots, n,\]

and is said to be irreducibly weakly diagonally dominant (IWDD) if \(A\) is WDD dominant and irreducible.

**Theorem 2** Any IWDD matrix is nonsingular.

**Proof.** See [4, 6]. \(\square\)

**Theorem 3** Let \(0 < \omega \leq 1\). Let also \(A\) be an IWDD matrix and \(T_m\) be irreducible. Then the associated GESOR method is convergent for every initial guess \(x_0\).
Proof. The proof is similar to the proof of theorem for the SOR method [6]. Obviously $T_m - \omega E_m$ is and IWDD matrix and therefore is nonsingular. Let $\lambda$ be an eigenvalue of $B_{GSOR}^{(m)}$ such that $1 \geq \lambda \geq 1$. We show that this is impossible. Then we have
\[
\det(B_{GSOR}^{(m)} - \lambda I) = 0 \Rightarrow \det((T_m - \omega E_m)^{-1} \det((1 - \omega)T_m + \omega F_m) - \lambda (T_m - \omega E_m) = 0
\]
\[
\Rightarrow \det((1 - \omega - \lambda)T_m + \omega F_m + \lambda \omega E_m) = 0
\]
It can be easily verified that $1 - \omega - \lambda \neq 0$. Hence $\det(Q) = 0$, where
\[
Q = T_m - \frac{\omega}{\omega + \lambda - 1} F_m - \frac{\lambda \omega}{\omega + \lambda - 1} E_m.
\]
It can be verified that
\[
\left| -\frac{\omega}{\omega + \lambda - 1} \right| \leq 1, \quad \text{and} \quad \left| -\frac{\lambda \omega}{\omega + \lambda - 1} \right| \leq 1.
\]
Therefore we conclude that the matrix $Q$ is an IWWD matrix. But we have $\det(Q) = 0$, which is a contradiction. Hence $1 \geq \lambda > 1$, and the desired result is obtained. □

In [2], in the special case, it has been proved that the GJ and GGS methods are faster than the Jacobi and Gauss-Seidel methods. Here, an open problem for the GESOR method is “For what kind of matrices is the GESOR method faster than the SOR method?”

In the next section, we give some numerical experiments. Although, the convergence of the GESOR method for the SPD and IWDD matrices have been proved but it can be applied for the general system of equations, as the SOR method.

3 Numerical Examples

All the numerical experiments presented in this section were computed in double precision with some MATLAB codes on a personal computer Pentium 4 - 256 MHz. For the numerical experiments we consider the equation
\[
- \Delta u + g(x, y)u = f(x, y), \quad (x, y) \in \Omega = (0,1) \times (0,1).
\]

Discretizing Eq. (6) on an $nx \times nx$ grid, by using the second order centered differences for the Laplacian gives a linear system of equations of order $n = nx \times nx$ with $n$ unknowns $u_{ij} = u(ih, jh)(1 \leq i, j \leq n)$:
\[
- u_{i-1,j} - u_{i,j-1} + (4 + h^2 g(ih, jh))u_{ij} - u_{i+1,j} - u_{i,j+1} = h^2 f(ih, jh),
\]
where $h = 1/nx$. The boundary conditions are taken so that the exact solution of the system is $x = (1, \cdots, 1)^T$. We consider the linear systems arisen from this kind of discretization for three functions $g(x, y) = \exp(xy)$, $g(x, y) = x + y$, and $g(x, y) = 0$. It can be easily verified that the coefficient matrices of these systems are SPD (see for example [3]). For each function we give the numerical results of the GESOR method presented in section 2 for $nx = 20, 30$ and $40$ and for two values of $\omega$, i.e., $\omega = 0.8$
and 1.2. The stopping criterion $\|x_{k+1} - x_k\| < 10^{-7}$ was used and the initial guess was taken to be zero vector. We let $m = 1$. In this case $T_m$ is tridiagonal matrix and then the matrix $T_m - \omega E_m$ an upper Hessenberg matrix. In the implementation of the GESOR methods we used the LU factorization of $T_m - \omega E_m$. Note that for computing the LU factorization of $T_m - \omega E_m$ it is enough annihilate its only lower off diagonal. Numerical results are given in Tables 1, 2, and 3. In each table the number of iterations of the method and the CPU time (in parenthesis) for convergence are given (timings are in seconds). We also give the numerical results related to the function $g(x, y) = -\exp(4xy)$ in Table 4 for $nx = 80, 90, 100$ for $\omega = 0.8, 1.2$. In all of the tables a dagger (†) shows that the method does not converge after 10000 iterations. As we see the GESOR method is more effective than the SOR method, both iterations and CPU times.

### Table 1
Numerical results for $g(x, y) = \exp(xy)$

<table>
<thead>
<tr>
<th>$nx$</th>
<th>SOR (0.8)</th>
<th>GESOR (0.8)</th>
<th>SOR (1.2)</th>
<th>GESOR (1.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>757 (0.16)</td>
<td>55 (0.00)</td>
<td>516 (0.11)</td>
<td>56 (0.00)</td>
</tr>
<tr>
<td>30</td>
<td>1625 (0.72)</td>
<td>56 (0.00)</td>
<td>1111 (0.61)</td>
<td>57 (0.00)</td>
</tr>
<tr>
<td>40</td>
<td>2804 (2.09)</td>
<td>56 (0.06)</td>
<td>1920 (1.37)</td>
<td>57 (0.06)</td>
</tr>
</tbody>
</table>

### Table 2
Numerical results for $g(x, y) = x + y$

<table>
<thead>
<tr>
<th>$nx$</th>
<th>SOR (0.8)</th>
<th>GESOR (0.8)</th>
<th>SOR (1.2)</th>
<th>GESOR (1.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>767 (0.17)</td>
<td>55 (0.00)</td>
<td>523 (0.11)</td>
<td>56 (0.00)</td>
</tr>
<tr>
<td>30</td>
<td>1646 (0.72)</td>
<td>56 (0.00)</td>
<td>1126 (0.44)</td>
<td>57 (0.00)</td>
</tr>
<tr>
<td>40</td>
<td>2841 (2.14)</td>
<td>56 (0.06)</td>
<td>1946 (1.42)</td>
<td>57 (0.05)</td>
</tr>
</tbody>
</table>

### Table 3
Numerical results for $g(x, y) = 0$

<table>
<thead>
<tr>
<th>$nx$</th>
<th>SOR (0.8)</th>
<th>GESOR (0.8)</th>
<th>SOR (1.2)</th>
<th>GESOR (1.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>800 (0.17)</td>
<td>55 (0.00)</td>
<td>546 (0.11)</td>
<td>56 (0.00)</td>
</tr>
<tr>
<td>30</td>
<td>1720 (0.71)</td>
<td>56 (0.00)</td>
<td>1176 (0.49)</td>
<td>57 (0.00)</td>
</tr>
<tr>
<td>40</td>
<td>2970 (2.09)</td>
<td>56 (0.06)</td>
<td>2034 (1.43)</td>
<td>57 (0.05)</td>
</tr>
</tbody>
</table>

### Table 4
Numerical results for $g(x, y) = -\exp(4xy)$

<table>
<thead>
<tr>
<th>$nx$</th>
<th>SOR(0.8)</th>
<th>GESOR(0.8)</th>
<th>SOR(1.2)</th>
<th>GESOR(1.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>†</td>
<td>56 (0.33)</td>
<td>9036 (35.76)</td>
<td>57 (0.33)</td>
</tr>
<tr>
<td>90</td>
<td>†</td>
<td>56 (0.38)</td>
<td>†</td>
<td>57 (0.38)</td>
</tr>
<tr>
<td>100</td>
<td>†</td>
<td>56 (0.50)</td>
<td>†</td>
<td>57 (0.55)</td>
</tr>
</tbody>
</table>
4 Conclusion and Future Works

In this paper, we have proposed a generalization of the SOR method say GESOR method and studied its convergence properties for the SPD matrices. In the decomposition of the coefficient matrix a banded matrix $T_m$ of bandwidth $2m + 1$ is chosen. Matrix $T_m$ is chosen such that the computation of $w = (T_m - \omega E_m)^{-1} y$ can be easily done for any vector $y$. To do so one may use the LU factorization of $T_m - \omega E_m$. In practice $m$ is chosen very small, e.g., $m = 1, 2$. For $m = 1$, $T_m$ is a tridiagonal matrix and the LU factorization of $T_m - \omega E_m$ can be easily obtained. The new method is suitable for sparse matrices such as matrices arisen from discretization of the PDEs. These kinds of matrices are usually pentadiagonal. In this case for $m = 1$, $T_m$ is tridiagonal and each of the matrices $E_m$ and $F_m$ contains only one nonzero diagonal and a few additional computations are needed in comparing with the SOR method (as we did in this paper). Numerical results show that the new methods are more effective than the SOR method.

References

Surf and download all data from SID.ir: www.SID.ir

Translate via STRS.ir: www.STRS.ir

Follow our scientific posts via our Blog: www.sid.ir/blog

Use our educational service (Courses, Workshops, Videos and etc.) via Workshop: www.sid.ir/workshop