Convergence Analysis Of The Modified Gauss-Seidel Iterative Method For $H$-Matrices*

Shu-Xin Miao†, Bing Zheng‡

Received 17 September 2011

Abstract

In 2002, Kotakemori et al. [H. Kotakemori, K. Harada, M. Morimoto and H. Niki, A comparison theorem for the iterative method with the preconditioner $I + S_{\max}$, J. Comput. Appl. Math., 145(2002), 373–378] considered the modified Gauss-Seidel method for irreducibly diagonally dominant $Z$-matrices with the preconditioner $P = I + S_{\max}$. In this paper, we consider a modified Gauss-Seidel method for solving the linear systems, which is a generalization of the method considered by Kotakemori et al., and prove its convergence when the coefficient matrix is an $H$-matrix. Numerical examples are given to illustrate our theoretical analysis.

1 Introduction

Consider the following linear system

$$Ax = b,$$

where $A = (a_{i,j})$ is an $n \times n$ nonsingular matrix, $x$ and $b$ are $n$-dimensional vectors. If $A$ has a splitting of the form $A = M - N$, where $M$ is nonsingular, then the splitting iterative method for solving (1) can be expressed as

$$x_{i+1} = M^{-1}Nx_i + M^{-1}b, \quad i = 0, 1, 2, \ldots$$

It is well known that the above iterative scheme is convergent if and only if $\rho(M^{-1}N) < 1$, where $\rho(M^{-1}N)$ denotes the spectral radius of the iterative matrix $M^{-1}N$. The smaller is $\rho(M^{-1}N)$, the faster is the convergence. For improving the convergent rate of corresponding iterative method, preconditioning techniques are used [2]. In particular, we consider the following equivalent left preconditioned linear system of (1)

$$PAx = Pb,$$
where $P$, called the preconditioner, is nonsingular. The corresponding iterative method for solving (2) is given by

$$x_{i+1} = M_p^{-1}N_px_i + M_p^{-1}Pb, \quad i = 0, 1, 2, \ldots,$$

based on the splitting $PA = M_p - N_p$, where $M_p$ is nonsingular.

Many left preconditioner $P$ were proposed, see [5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19] and the references therein. In 2002, Kotakemori et al. [11] considered the preconditioner $P_{\text{max}} = I + S_{\text{max}}$, where

$$S_{\text{max}} = (s_{i,j}^m) = \begin{cases} -a_{i,k_i}, & i = 1, \ldots, n-1, j > i; \\ 0, & \text{otherwise}, \end{cases}$$

with $k_i = \min\{j | \max_j |a_{i,j}|, i < n\}$. As for the discussion of the preconditioner $P_{\text{max}} = I + S_{\text{max}}$, we refer to [9, 11, 12, 15, 19]. It is reported that the modified Gauss-Seidel method with the preconditioner $P_{\text{max}}$ is superior to the classical Gauss-Seidel method under some conditions when $A$ is an irreducibly diagonally dominant $Z$-matrices.

In this paper, we consider the generalized preconditioner $P_{\text{max}}(\alpha) = I + S_{\text{max}}(\alpha)$, where

$$S_{\text{max}}(\alpha) = (s_{i,j}^m) = \begin{cases} -\alpha_i a_{i,k_i}, & i = 1, \ldots, n-1, j > i; \\ 0, & \text{otherwise}, \end{cases}$$

with $k_i = \min\{j | \max_j |a_{i,j}|, i < n\}$, $\alpha_i(i = 1, \ldots, n-1)$ are positive real numbers. When $\alpha_i = 1(i = 1, 2, \ldots, n-1)$, the preconditioner $P_{\text{max}}(\alpha)$ reduces to the one considered in [11].

The basic purpose of the present paper is to prove the convergence of the modified Gauss-Seidel method with the preconditioner $P_{\text{max}}(\alpha)$ for solving (1) for the case that the coefficient matrix is an $H$-matrix.

Without loss of generality, we always assume that $A$ has a splitting of the form $A = I - L - U$, where $I$ is the identity matrix, $-L$ and $-U$ are strictly lower-triangular and strictly upper-triangular parts of $A$, respectively.

The remainder of the present paper is organized as follows. Next section is the preliminaries. The convergence of the modified Gauss-Seidel method are studied for $H$-matrix in Section 3. In Section 4, numerical examples are given to illustrate our theoretical analysis.

## 2 Preliminaries

In this section, we give some of the notations, definitions and lemmas which will be used in what follows.

A vector $x = (x_1, x_2, \ldots, x_n)^T$ is called nonnegative (positive) and denoted by $x \geq 0$ ($x > 0$), if $x_i \geq 0$ ($x_i > 0$) for all $i$. Similarly, a matrix $A = (a_{i,j})$ is called nonnegative (positive) and denoted by $A \geq 0$ ($A > 0$), if $a_{i,j} \geq 0$ ($a_{i,j} > 0$) for all $i, j$. The absolute value of a matrix $A$ is denoted by $|A| = (|a_{i,j}|)$. The comparison matrix of $A$ is defined as $(A) = (\tilde{a}_{i,j})$, where $\tilde{a}_{i,j}$ satisfies

$$\tilde{a}_{i,j} = \begin{cases} |a_{i,j}|, & i = j; \\ -|a_{i,j}|, & i \neq j. \end{cases}$$
Therefore, we have

\[ = (u_1, u_2, ..., u_n)^T > 0 \]

Thus, we get that

\[ u_i - \sum_{j=1, j \neq i}^n |a_{i,j}| u_j > 0 \quad \text{for} \quad i = 1, 2, ..., n - 1. \]

Therefore, we have

\[
\begin{align*}
&u_i - \sum_{j=1}^{i-1} |a_{i,j}| u_j - \sum_{j=i+1, j \neq k_i}^n |a_{i,j}| u_j + |a_{i,k_i}| u_{k_i} - |a_{i,k_i}| \sum_{j=1}^n |a_{k_i,j}| u_j \\
&= u_i - \sum_{j=1, j \neq i}^n |a_{i,j}| u_j + |a_{i,k_i}| (u_{k_i} - \sum_{j=1, j \neq k_i}^n |a_{k_i,j}| u_j)
\end{align*}
\]
Hence, the inequality $u_i - \sum_{j=1, j \neq i}^{n} |a_{i,j}|u_j > 0$. Noting that $k_i < n$, again from (4), the inequality $u_{k_i} - \sum_{j=1, j \neq k_i}^{n} |a_{k_i,j}|u_j > 0$ holds. Hence, for $i = 1, 2, ..., n - 1$,

$$u_i - \sum_{j=1}^{i-1} |a_{i,j}|u_j - \sum_{j=i+1, j \neq k_i}^{n} |a_{i,j}|u_j + |a_{i,k_i}|u_{k_i} - |a_{i,k_i}| \sum_{j=1}^{n} |a_{k_i,j}|u_j > 0$$

Under the assumptions, we further obtain that

$$u_i - \sum_{j=1}^{n} |a_{i,j}|u_j - \sum_{j=i+1, j \neq k_i}^{n} |a_{i,j}|u_j + |a_{i,k_i}|u_{k_i} > |a_{i,k_i}| \sum_{j=1}^{n} |a_{k_i,j}|u_j > 0 \text{ for } i = 1, 2, ..., n-1.$$

Hence,

$$\beta_i = \frac{u_i - \sum_{j=1}^{i-1} |a_{i,j}|u_j - \sum_{j=i+1, j \neq k_i}^{n} |a_{i,j}|u_j + |a_{i,k_i}|u_{k_i}}{|a_{i,k_i}| \sum_{j=1}^{n} |a_{k_i,j}|u_j}$$

are well defined and $\beta_i > 1$ for $i = 1, 2, ..., n - 1$.

REMARK: It should be remarked that $\beta_i$ ($i = 1, 2, ..., n - 1$) in Theorem 1 depends on the positive vector $u$. There are many such vectors $u$ satisfying $u > 0$, how to choose applicable $u$ is very important for practical computation. In general, we can let $u = (1,1,...,1)^T$ when $A$ is the strictly diagonally dominant $H$-matrix, while when $A$ is not strictly diagonally dominant, it follows from [7] that the elements $m_{i,j}$ of $(A)^{-1}$ satisfies

$$\sum_{j=1}^{n} m_{i,j} \geq 1, \quad i = 1, 2, ..., n,$$

hence we can let $u_i = \sum_{j=1}^{n} m_{i,j}$ for $i = 1, 2, ..., n$ and $u = (u_1, u_2, ..., u_n)^T$. However, finding out $\beta_i$ ($i = 1, 2, ..., n - 1$) which are independent of the vector $u$ is still an open problem need further study.

Now we are in the position to establish the convergence of the modified Gauss-Seidel method with the preconditioner $P_{\text{max}}(\alpha) = I + S_{\text{max}}(\alpha)$ for $H$-matrices.

THEOREM 2. Let $A$ be an $H$-matrix with unit diagonal elements. If $\beta_i$, $M_\alpha$ and $N_\alpha$ are defined as in Theorem 1, then for $0 \leq \alpha_i < \beta_i$, $i = 1, 2, ..., n - 1$, the splitting $A_\alpha = M_\alpha - N_\alpha$ is an $H$-splitting and $\rho(M_\alpha^{-1}N_\alpha) < 1$.

PROOF. In order to prove the splitting $A_\alpha = M_\alpha - N_\alpha$ is an $H$-splitting, we only need to show that $(M_\alpha) - |N_\alpha|$ is an $M$-matrix.

Let $[(M_\alpha) - |N_\alpha|]u_i$ be the $i$-th element in the vector $(M_\alpha) - |N_\alpha|)u$ for $i =$
1, 2, ..., $n - 1$, where $u = (u_1, u_2, ..., u_n)^T$ is a positive vector. Then we have

$$[(\langle M \rangle - |N|)u]_i = |1 - \alpha_i a_{i,k_i} a_{k_i,i}| u_i - \sum_{j=1, j \neq i}^{n} |a_{i,j} - \alpha_i a_{i,k_i} a_{k_i,j}| u_j$$

$$\geq u_i - \alpha_i |a_{i,k_i} a_{k_i,i}| u_i - \sum_{j=1}^{i-1} |a_{i,j}| u_j - \alpha_i \sum_{j=1}^{i-1} |a_{i,k_i} a_{k_i,j}| u_j$$

$$- \alpha_i \sum_{j=i+1, j \neq k_i}^{n} |a_{i,k_i} a_{k_i,j}| u_j$$

$$- \alpha_i \sum_{j=i+1, j \neq k_i}^{n} |a_{i,k_i} a_{k_i,j}| u_j,$$

and

$$[(\langle M \rangle - |N|)u]_n = u_n - \sum_{j=1, j \neq i}^{n} |a_{n,j}| u_j > 0.$$

If $0 \leq \alpha_i \leq 1$ ($i = 1, 2, ..., n - 1$), then we have

$$[(\langle M \rangle - |N|)u]_i \geq u_i - \alpha_i |a_{i,k_i} a_{k_i,i}| u_i - \sum_{j=1}^{i-1} |a_{i,j}| u_j - \alpha_i \sum_{j=1}^{i-1} |a_{i,k_i} a_{k_i,j}| u_j$$

$$- \alpha_i \sum_{j=i+1, j \neq k_i}^{n} |a_{i,k_i} a_{k_i,j}| u_j$$

$$- \alpha_i \sum_{j=i+1, j \neq k_i}^{n} |a_{i,k_i} a_{k_i,j}| u_j$$

$$= u_i - \sum_{j=1, j \neq i}^{n} |a_{i,j}| u_j + \alpha_i |a_{i,k_i} a_{k_i,i}| u_j - \alpha_i \sum_{j=1, j \neq k_i}^{n} |a_{k_i,j}| u_j$$

$$= (u_i - \sum_{j=1, j \neq i}^{n} |a_{i,j}| u_j) + \alpha_i |a_{i,k_i} a_{k_i,i}| (u_k_i - \sum_{j=1, j \neq k_i}^{n} |a_{k_i,j}| u_j).$$

Since $u_i - \sum_{j=1, j \neq i}^{n} |a_{i,j}| u_j > 0$ and $u_k_i - \sum_{j=1, j \neq k_i}^{n} |a_{k_i,j}| u_j > 0$, one get that

$$[(\langle M \rangle - |N|)u]_i > 0 \text{ for } i = 1, 2, ..., n - 1.$$
If \( 1 < \alpha_i < \beta_i \) (\( i = 1, 2, ..., n - 1 \)), from (5) and the definition of \( \beta_i \), we have

\[
[(\mathcal{M}_\alpha) - |N_\alpha|]u_i \geq \ u_i - \alpha_i |a_{i,k_i}a_{k_i,i}|u_i - \sum_{j=1}^{i-1} |a_{i,j}|u_j - \alpha_i \sum_{j=1}^{i-1} |a_{i,k_i}a_{k_i,j}|u_j
- \sum_{j=i+1, j \neq k_i}^{n} |a_{i,j}|u_j - (\alpha_i - 1)|a_{i,k_i}|u_{k_i}
- \alpha_i \sum_{j=i+1, j \neq k_i}^{n} |a_{i,k_i}a_{k_i,j}|u_j
= u_i - \sum_{j=1}^{i-1} |a_{i,j}|u_j - \sum_{j=i+1, j \neq k_i}^{n} |a_{i,j}|u_j
+ |a_{i,k_i}|u_{k_i} - \alpha_i |a_{i,k_i}| \sum_{j=1}^{n} |a_{k_i,j}|u_j
> 0.
\] (8)

Therefore, it follows from (5)–(8) that

\[
(\mathcal{M}_\alpha) - |N_\alpha| u > 0 \quad \text{for} \quad 0 \leq \alpha_i < \beta_i.
\]

By Lemma 2, we know that \( \langle \mathcal{M}_\alpha \rangle - |N_\alpha| \) is an M-matrix for \( 0 \leq \alpha_i < \beta_i \) (\( i = 1, 2, ..., n - 1 \)). From Definition 3, \( A_{\alpha} = \mathcal{M}_\alpha - N_\alpha \) is an H-splitting for \( 0 \leq \alpha_i < \beta_i \) (\( i = 1, 2, ..., n - 1 \)). Hence, Lemma 1 yields \( \rho(M_\alpha^{-1}N_\alpha) < 1 \) for \( 0 \leq \alpha_i < \beta_i \) (\( i = 1, 2, ..., n - 1 \)), the proof is completed.

REMARK: From Theorem 2, we can see that the modified Gauss-Seidel method is convergent for all \( 0 \leq \alpha_i < \beta_i \), \( i = 1, 2, ..., n - 1 \) with the preconditioner \( P_{S_{\max}}(\alpha) \) when the coefficient matrix \( A \) of (1) is an H-matrix. The convergence condition when \( A \) is an H-matrix is much weaker than the one, studied in [11, 12, 19], when \( A \) is an M-matrix.

4 Examples

In this section, we use two examples to verify our theoretical analysis in Section 3.

It is well known that the Toeplitz matrices arise in many applications, such as solutions to differential and integral equations, spline functions, and problems and methods in physics, mathematics, statistics, and signal processing [6]. Therefore, in our first example, we consider the case that the coefficient matrix \( A \) of (1) is a Toeplitz matrix.

EXAMPLE 1. Let the coefficient matrix of (1) be a symmetric Toeplitz matrix as

\[
A = \begin{bmatrix}
    a & b & c & \cdots & b \\
    b & a & b & \cdots & c \\
    c & b & a & \cdots & b \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    b & c & b & \cdots & a
\end{bmatrix}_{n \times n},
\]
where \( a = 1, b = 1/n \) and \( c = 1/(n-2) \). It is clear that \( A \) is an \( H \)-matrix.

The spectral radii of modified Gauss-Seidel iteration matrix with various values of \( \alpha_i \) for \( i = 1, \ldots, n-1 \) and \( n \) are listed in Table 1

<table>
<thead>
<tr>
<th>( \alpha_i )</th>
<th>( n = 90 )</th>
<th>( n = 120 )</th>
<th>( n = 180 )</th>
<th>( n = 210 )</th>
<th>( n = 300 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.2175</td>
<td>0.2171</td>
<td>0.2168</td>
<td>0.2167</td>
<td>0.2165</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2123</td>
<td>0.2133</td>
<td>0.2142</td>
<td>0.2145</td>
<td>0.2150</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2097</td>
<td>0.2114</td>
<td>0.2130</td>
<td>0.2134</td>
<td>0.2142</td>
</tr>
<tr>
<td>1.0</td>
<td>0.2081</td>
<td>0.2101</td>
<td>0.2121</td>
<td>0.2127</td>
<td>0.2137</td>
</tr>
<tr>
<td>1.2</td>
<td>0.2064</td>
<td>0.2089</td>
<td>0.2113</td>
<td>0.2120</td>
<td>0.2132</td>
</tr>
<tr>
<td>1.5</td>
<td>0.2039</td>
<td>0.2070</td>
<td>0.2101</td>
<td>0.2109</td>
<td>0.2125</td>
</tr>
</tbody>
</table>

EXAMPLE 2. When the central difference scheme on a uniform grid with \( N \times N \) interior nodes \( (N^2 = n) \) is applied to the discretization of the two-dimension convection-diffusion equation

\[-\Delta u + \frac{\partial u}{\partial x} + 2 \frac{\partial u}{\partial y} = f\]

in the unit square \( \Omega \) with Dirichlet boundary conditions, we obtain a system of linear equations (1) with the coefficient matrix

\[ A = I \otimes P + Q \otimes I, \]

where \( \otimes \) denotes the Kronecker product,

\[ P = \text{tridiag} \left( -\frac{2+h}{8}, 1, -\frac{2-h}{8} \right) \text{ and } Q = \text{tridiag} \left( -\frac{1+h}{4}, 0, -\frac{1-h}{8} \right) \]

are \( N \times N \) tridiagonal matrices, and the step size is \( h = 1/N \).

It is clear that the matrix \( A \) is an \( M \)-matrix, see for example [19], so it is an \( H \)-matrix. We list the spectral radii of modified Gauss-Seidel iteration matrix with various values of \( \alpha_i \) for \( i = 1, \ldots, n-1 \) and \( n \) in Table 2

<table>
<thead>
<tr>
<th>( \alpha_i )</th>
<th>( n = 16 )</th>
<th>( n = 64 )</th>
<th>( n = 81 )</th>
<th>( n = 100 )</th>
<th>( n = 256 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.6159</td>
<td>0.8687</td>
<td>0.8927</td>
<td>0.9108</td>
<td>0.9621</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5020</td>
<td>0.8182</td>
<td>0.8507</td>
<td>0.8754</td>
<td>0.9464</td>
</tr>
<tr>
<td>1.0</td>
<td>0.4582</td>
<td>0.7993</td>
<td>0.8350</td>
<td>0.8621</td>
<td>0.9405</td>
</tr>
<tr>
<td>1.5</td>
<td>0.2980</td>
<td>0.7372</td>
<td>0.7836</td>
<td>0.8190</td>
<td>0.9217</td>
</tr>
<tr>
<td>1.8</td>
<td>0.2270</td>
<td>0.6833</td>
<td>0.7396</td>
<td>0.7824</td>
<td>0.9061</td>
</tr>
<tr>
<td>2.0</td>
<td>0.2827</td>
<td>0.6343</td>
<td>0.7006</td>
<td>0.7505</td>
<td>0.8928</td>
</tr>
</tbody>
</table>

From Table 1 and 2, it can be seen that the modified Gauss-Seidel method is convergent for Example 1 and 2 when \( \alpha_i \in [0, \beta_i] \), i.e., \( \rho(M_\alpha^{-1}N_\alpha) < 1 \). This confirm
the result of Theorem 2 in Section 3. In particular, if we take \( \alpha_i = 1 \) for \( i = 1, \ldots, n-1 \), then the preconditioner \( P_{S_{\text{max}}} (\alpha) \) reduces to the one considered in [11].

**Acknowledgment.** The work was supported by the National Natural Science Foundation of China under grant number 11171371.

**References**


