STAIR MATRICES AND THEIR GENERALIZATIONS WITH APPLICATIONS TO ITERATIVE METHODS I: A GENERALIZATION OF THE SOR METHOD *

HAO LU †

Abstract. Stair matrices and their generalizations are introduced. Some properties of the matrices are presented. Like triangular matrices this class of matrices provides bases of matrix splittings for iterative methods. A remarkable feature of iterative methods based on the new class of matrices is that the methods are easily implemented for parallel computation. In particular, a generalization of the SOR method is introduced. The SOR theory on determination of the optimum parameter is extended to the generalized method to include a wide class of matrices. The asymptotic rate of convergence of the new method is derived for Hermitian positive definite matrices using bounds of the eigenvalues of Jacobi matrices and numerical radius. Finally, numerical tests are presented to corroborate the obtained results.

Key words. stair matrices and their generalization, iterative method, parallel computation, generalization of the SOR method, optimum parameter, numerical radius, convergence rate

AMS subject classifications. 65F10, 65F15, 65F50

1. Introduction. Let A be a nonsingular matrix and split $A = M - N$ with a nonsingular matrix M. A basic iterative method to solve the linear system $A\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x}^{(k+1)} = M^{-1}(N\mathbf{x}^{(k)} + \mathbf{b}). \tag{1.1}$$

The matrix $M^{-1}N$ is called an iteration matrix. It is well known that the iterative method (1.1) converges if and only if the spectral radius $\rho(M^{-1}N) < 1$. Different matrix splittings yield different iterative methods. Split $A = D - L - U$, where and throughout the paper D is the (block) diagonal of A, $-L$ and $-U$ are the strictly (block) lower and the strictly (block) upper triangular matrices of A, respectively, and assume D is nonsingular. We have

- the Jacobi method if $M = D$,
- the Gauss-Seidel method if $M = D - L$ and
- the SOR method if $M = \frac{D}{\omega} - L$, where ω is a real parameter.

Detail discussions of basic iterative methods are found in [7], [18] and [22].

The Jacobi method is easily implemented on parallel computing platforms, but it is neither robust nor as fast as the Gauss-Seidel method and the SOR method in sequential case. With a proper over-relaxation parameter the SOR method substantially improves the Jacobi method and the Gauss-Seidel method in term of order improvement when applied to elliptic equations [7], [18] and [22]. However, the SOR method and the Gauss-Seidel method are not easily implemented for parallel computation because we have to solve triangular systems at each iteration.

The aim of the present paper is to search new matrix splittings to construct new iterative methods to have all advantages of the Jacobi method and the SOR method. The paper is organized as follows. Firstly, stair matrices and their generalizations are introduced. Some properties of the matrices are presented. Like triangular matrices this class of matrices provides bases of matrix splittings for iterative methods. An

*Version July 10, 1998

†Institute for Scientific Computation, Texas A&M University, College Station, Texas 77843-3404, USA (na.hlu@na-net.ornl.gov)

iterative method based on the new class of matrices is easily implemented for parallel computation. Secondly, a generalization of the SOR method is introduced. The SOR theory on determination of the optimum parameter is extended to the new method to include a wide class of matrices. Thirdly, the asymptotic rate of convergence of the new method is estimated for Hermitian positive definite matrices using bounds of the eigenvalues of Jacobi matrices and numerical radius. Finally, numerical examples are presented to corroborate the analysis.

2. Stair matrices and their generalizations. In this section we first introduce stair matrices and then generalize stair matrices for application to iterative methods. Some properties of the matrices are also presented.

We denote $A = (a_{ij})_{n\times n}$ an $n \times n$ matrix. The entries a_{ij} can be $n_i \times n_j$ blocks. In the case a_{ij} are blocks we still treat them as basic entries. If we emphasize that entries of a matrix are blocks, notation A_{ij} is used to represent the (i, j)th entry instead of a_{ij}. $\det(A)$ denotes the determinant of A. For a tridiagonal matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & & & & \\ a_{21} & a_{22} & a_{23} & & & \\ & \ddots & \ddots & \ddots & & \\ & & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ & & & a_{n,n-1} & a_{nn} \end{pmatrix}, \tag{2.1}$$

we briefly denote $A = \mathrm{tridiag}(a_{i,i-1}, a_{ii}, a_{i,i+1})$. A stair matrix is a special tridiagonal matrix defined as follows.

Definition 2.1. *A tridiagonal matrix* $A = \mathrm{tridiag}(a_{i,i-1}, a_{ii}, a_{i,i+1})$ *is called a stair matrix if one of the following conditions is satisfied*

I. $a_{i,i-1} = 0,\ a_{i,i+1} = 0,\ i = 1, 3, \ldots, 2\lfloor \frac{n-1}{2} \rfloor + 1$*;*

II. $a_{i,i-1} = 0,\ a_{i,i+1} = 0,\ i = 2, 4, \ldots, 2\lfloor \frac{n}{2} \rfloor$.

A stair matrix is of the type I if the condition I is satisfied and is of the type II if the condition II holds.

For example, a 6×6 stair matrix is of the form

$$A = \begin{pmatrix} \times & & & & & \\ \times & \times & \times & & & \\ & & \times & & & \\ & & \times & \times & \times & \\ & & & & \times & \\ & & & & \times & \times \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} \times & \times & & & & \\ & \times & & & & \\ & \times & \times & \times & & \\ & & & \times & & \\ & & & \times & \times & \times \\ & & & & & \times \end{pmatrix}.$$

For convience, A stair matrix is denoted by $A = \mathrm{stair}(a_{i,i-1}, a_{ii}, a_{i,i+1})$. In particular, $A = \mathrm{stair1}(a_{i,i-1}, a_{ii}, a_{i,i+1})$ and $A = \mathrm{stair2}(a_{i,i-1}, a_{ii}, a_{i,i+1})$ represent a stair matrix of the type I and a stair matrix of the type II, repectively.

Lemma 2.2. *An* $n \times n$ *stair matrix* $A = \mathrm{stair}(a_{i,i-1}, a_{ii}, a_{i,i+1})$ *is nonsingular if and only if* a_{ii}, $i = 1, 2, \ldots, n$ *are nonsingular. Furthermore, if A is nonsigular then*

$$A^{-1} = D^{-1}(2D - A)D^{-1}, \tag{2.2}$$

where $D = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.

Proof. It is straightforward to show that

$$\det(A) = \prod_{i=1}^{n} \det(a_{ii}), \tag{2.3}$$

which implies the first part of the lemma. The second part follows from a matrix multiplication. □

Applying Lemma 2.2 we immediately obtain the solution of a stair linear system

$$A\mathbf{x} = \mathbf{b} \tag{2.4}$$

by computing $A^{-1}\mathbf{b}$, where A is a stair matrix. To further reduce computational cost, however, we solve (2.4) based on the structure of A as follows.

ALGORITHM I. This algorithm solves the stair linear system (2.4). The solution overwrites $\mathbf{b}$.

if (A is of the type I)
 for $i = 1 : 2 : 2\lfloor \frac{n-1}{2} \rfloor + 1$
  $b_i = a_{ii}^{-1} b_i$
 endfor i
 for $i = 2 : 2 : 2\lfloor \frac{n}{2} \rfloor$
  $b_i = a_{ii}^{-1}(b_i - a_{i,i-1}b_{i-1} - a_{i,i+1}b_{i+1})$
 endfor i
endif
if (A is of the type II)
 for $i = 2 : 2 : 2\lfloor \frac{n}{2} \rfloor$
  $b_i = a_{ii}^{-1} b_i$
 endfor i
 for $i = 1 : 2 : 2\lfloor \frac{n-1}{2} \rfloor + 1$
  $b_i = a_{ii}^{-1}(b_i - a_{i,i-1}b_{i-1} - a_{i,i+1}b_{i+1})$
 endfor i
endif,

where $b_i = 0$ if $i < 1$ or $i > n$. It is readily seen that in the scale case the algorithm needs at most $3n$ arithmetic operations, i.e., n additions, n multiplications and n divisions. In block case, the algorithm needs n matrix-vector products of the form $a_{ij}b_j$ $j = i-1, i+1$, n vector additions to compute $b_i - (a_{i,i-1}b_{i-1}) - (a_{i,i+1}b_{i+1})$ and solving n linear systems of the form $a_{ii}^{-1}d$. A remarkable feature of the algorithm is its high parallelism. For example, if A is a stair matrix of the type I, first, for all odd i the computations of $a_{ii}^{-1}b_i$ can be fulfilled by different processors at same time. Then $b_i = a_{ii}^{-1}(b_i - a_{i,i-1}b_{i-1} - a_{i,i+1}b_{i+1})$ are easily computed in parallel for even i.

The high parallelism of algorithm I is achieved if a_{ij} are complex numbers or small blocks. To obtain a good matrix splitting such that the iterative method (1.1) is almost fully parallelized at each iteration for a wide class of matrices we now generalize stair matrices by defining

- $\mathcal{L}_n^1 = \{A : A \text{ is an } n \times n \text{ matrix and } A = \text{stair}(a_{i,i-1}, a_{ii}, a_{i,i+1})\}$,
- $\mathcal{L}_n^k = \{A : A \text{ is an } n \times n \text{ matrix and } A = \text{stair}(A_{i,i-1}, A_{ii}, A_{i,i+1})$, where each diagonal block A_{ii} is an $n_i \times n_i$ matrix and $A_{ii} \in \mathcal{L}_{n_i}^r$ with $r < k\}$.

LEMMA 2.3. *$\mathcal{L}_n^1 \subset \mathcal{L}_n^2 \subset \cdots \subset \mathcal{L}_n^n \subset \cdots$ and $\mathcal{L}_n^k = \mathcal{L}_n^n$ if $k \geq n$.*

Proof. The first part of the lemma follows straightforwardly from the definition of $\mathcal{L}_n^k$. If $n = 1$ the equation $\mathcal{L}_n^k = \mathcal{L}_n^n$ is trivial for $k \geq n$. Assume that the conclusion of the second part is true for $n \leq m-1$. Following the definition of $\mathcal{L}_n^k$ we find that it is true for $n = m$ too. □

Because of Lemma 2.3 we introduce the following notation.

DEFINITION 2.4. $\mathcal{L}_n \equiv \mathcal{L}_n^n$.

According to the definition of $\mathcal{L}_n$ we find that all $n \times n$ triangular matrices are elements of $\mathcal{L}_n$ and all $n \times n$ matrices of the form

$$A = \begin{pmatrix} \times & & & & & & \\ \times & \times & \times & \times & \times & \times & \cdots \\ \times & & \times & & & & \\ \times & & \times & \times & \times & \times & \cdots \\ \times & & \times & & \times & & \\ \times & & \times & & \times & \times & \cdots \\ \vdots & & \vdots & & \vdots & & \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} \times & \times & \times & \times & \times & \times & \cdots \\ & \times & & & & & \\ & \times & \times & \times & \times & \times & \cdots \\ & \times & & \times & & & \\ & \times & & \times & \times & \times & \cdots \\ & \times & & \times & & \times & \\ & \vdots & & \vdots & & \vdots & \end{pmatrix}$$

belong to $\mathcal{L}_n$ too. We call the latter zebra matrices.

THEOREM 2.5. *Let $A = (a_{ij})_{n\times n} \in \mathcal{L}_n$. Then $A^* \in \mathcal{L}_n$, where A^* is the conjugate transpose of A, and*

$$\det(A) = \prod_{i=1}^{n} \det(a_{ii}). \tag{2.5}$$

If A is nonsingular then $A^{-1} \in \mathcal{L}_n$.

Proof. Assume $A = \text{stair}(A_{i,i-1}, A_{ii}, A_{i,i+1})$. Then $A^* \in \mathcal{L}_n$ follows immediately from induction. Let m be the number of the diagonal blocks of A and denote $D = \text{blockdiag}(A_{11}, A_{22}, \ldots, A_{mm})$. We find

$$\det(A) = \prod_{i=1}^{m} \det(A_{ii}).$$

Using this equation and induction with the equation (2.3) we easily show (2.5).

If A is nonsingular it follows from (2.2) that

$$A^{-1} = D^{-1}(2D - A)D^{-1} = \text{stair}(B_{i,i-1}, B_{ii}, B_{i,i+1}),$$

where the blocks B_{ij} are given by

$$B_{ij} = \begin{cases} -A_{ii}^{-1}A_{ij}A_{jj}^{-1}, & \text{if } j = i-1, i+1 \\ A_{ii}^{-1}, & \text{if } j = i. \end{cases}$$

Again using induction we find $A^{-1} \in \mathcal{L}_n$. $\square$

If $A = \text{stair}(A_{i,i-1}, A_{ii}, A_{i,i+1}) \in \mathcal{L}_n$ we can repeatedly apply Algorithm I to solve the linear system $A\mathbf{x} = \mathbf{b}$. Assume $A = (a_{ij})_{n\times n}$ with complex entries a_{ij} and A_{ii} are $n_i \times n_i$ blocks. Let $C(n)$ denote the number of arithmetic operations for solving the linear system $A\mathbf{x} = \mathbf{b}$. Because a matrix-vector product $A_{ik}\mathbf{b}_k$ $(k = i-1, i+1)$ needs at most $n_i n_k$ multiplications and $(n_i - 1)n_k$ additions, the computation of $\mathbf{b}_i - A_{i,i-1}\mathbf{b}_{i-1} - A_{i,i+1}\mathbf{b}_{i+1}$ needs at most $2(n_i n_{i-1} + n_i n_{i+1})$ arithmetic operations. We have the following bound of $C(n)$:

$$C(n) \leq \sum_{i=1}^{m} C(n_i) + 2\sum_{i=1}^{m-1} n_i n_{i+1}. \tag{2.6}$$

Based on this inequality and induction we obtain $C(n) \leq n^2$. Therefore, solving a linear system $A\mathbf{x} = \mathbf{b}$ with $A \in \mathcal{L}_n$ is at most as expensive as solving a triangular linear system. However, the system $A\mathbf{x} = \mathbf{b}$ is easily solved in parallel by repeatedly performing Algorithm I.

3. A generalization of the SOR method. As we have seen in the previous section the iterative method (1.1) is easily implemented for parallel computation if $M \in \mathcal{L}_n$. In this section, we generalize the SOR method based on a splitting $A = M - N$, where $M \in \mathcal{L}_n$.

Let A be an $n \times n$ matrix with a nonsingular diagonal D. We denote $B = I - D^{-1}A$ the Jacobi matrix of A throughout the paper. Split $A = D - P - Q$ such that $P \in \mathcal{L}_n$, where the diagonals of P and Q are zero. A generalization of the SOR method for the linear system $A\mathbf{x} = \mathbf{b}$ is defined by

$$\mathbf{x}^{(k+1)} = \mathcal{S}_\omega \mathbf{x}^{(k)} - (D - \omega P)^{-1}\omega \mathbf{b}, \tag{3.1}$$

where and throughout the paper $\mathcal{S}_\omega$ stands for

$$\mathcal{S}_\omega = (D - \omega P)^{-1}((1-\omega)D + \omega Q) \tag{3.2}$$

with a real parameter ω.

Applying the results on matrices in $\mathcal{L}_n$ in the previous section, we have the following result similarly to Kahan's result [10] on the SOR method.

THEOREM 3.1. *Let $A = D - P - Q$ such that $P, Q \in \mathcal{L}_n$. If $\rho(\mathcal{S}_\omega) < 1$, then $0 < \omega < 2$.*

Proof. Represent $\mathcal{S}_\omega = (I - \omega D^{-1}P)^{-1}((1-\omega)I + \omega D^{-1}Q)$. Under the condition of the theorem we find $\omega D^{-1}P, \omega D^{-1}Q \in \mathcal{L}_n$ with zero diagonals. It follows from Theorem 2.5 that

$$\det((I - \omega D^{-1}P)^{-1}) = 1, \qquad \det((1-\omega)I + \omega D^{-1}Q) = (1-\omega)^n. \tag{3.3}$$

Let λ_i, $i = 1, \ldots, n$ be the eigenvalues of $\mathcal{S}_\omega$. Then the determinant of $\mathcal{S}_\omega$ is given by

$$\det(\mathcal{S}_\omega) = \prod_{k=1}^{n} \lambda_k = (1-w)^n. \tag{3.4}$$

Therefore,

$$\rho(\mathcal{S}_\omega) = \max_i |\lambda_i| \geq |\det(\mathcal{S}_\omega)|^{1/n} \geq |w - 1|, \tag{3.5}$$

which implies the conclusion of the theorem. ∎

A real square matrix $A = (a_{ij})_{n \times n}$ is a Z-matrix if $a_{ii} > 0$ for $i = 1, \ldots n$ and $a_{ij} \leq 0$ for $i \neq j$, $i, j = 1, \ldots, n$. In particular, a nonsingular Z-matrix is also called an M-matrix. We now show that the important result on the SOR method for Z-matrices given by Young [22] (Theorem 5.1, pages 120–122), which is an extension of the result of Stein and Rosenberg [15] on the Gauss-Seidel method, is still true for the new method (3.1).

THEOREM 3.2. *Let A be a Z-matrix with nonsingular diagonal D. Split $A = D - P - Q$ such that $P \in \mathcal{L}_n$ and P, Q are nonnegative matrices. If $\mathcal{S}_\omega$ is defined by (3.2) with $0 < \omega \leq 1$, then*

(a) $\rho(B) < 1$ if and only if $\rho(\mathcal{S}_\omega) < 1$;
(b) $\rho(B) < 1$ if and only if A is an M-matrix;
(c) if $\rho(B) < 1$ then $\rho(\mathcal{S}_\omega) \leq 1 - \omega + \omega\rho(B)$;
(d) if $\rho(B) \geq 1$ then $\rho(\mathcal{S}_\omega) \geq 1 - \omega + \omega\rho(B)$.

Proof. Under the conditions of the theorem $D^{-1}P$ is nonnegative and $D^{-1}P \in \mathcal{L}_n$. It follows from the proof of Theorem 2.5 and induction we find that $(I - \omega D^{-1}P)^{-1}$ is a nonnegative matrix. The rest of the proof is essentially the same as that of Theorem 5.1 in [22]. We delete further details. ∎

Most results on the SOR method can be extended to (3.1). In the present paper, however, we don't intend to cover all of them.

4. Determination of the optimum parameter. In this section we extend some elegant results on determination of the optimum parameter of the SOR method to the method (3.1). Some examples are presented to illustrate that our generalization improves the SOR theory to include a wide class of matrices.

4.1. Theoretical results. The proofs of the results in this subsection are essential the same as those in the SOR case in [18] and [22]. We present some of them for readers' convenience. First, we extend the concept of p-consistently ordered matrices.

DEFINITION 4.1. *Let $A = D - P - Q$, where D is the diagonal of A. If there is a positive constant p such that*

$$\det(\beta D - \alpha P - \alpha^{-(p-1)} Q) = \det(\beta D - P - Q) \tag{4.1}$$

for any constants $\alpha \neq 0$ and β. Then A is called a p-consistently ordered matrix with respect to (P, Q).

If A is a p-consistently ordered matrix. The following lemma shows the characteristic polynomial of the Jacobi matrix of A. If A is a p-constantly ordered matrix with respect to (L, U) the result was first observed by Young [20] for $p = 2$ and was extended by Varga [17] for any positive integer p.

LEMMA 4.2. *Let A be an $n \times n$ matrix with a nonsingular diagonal D. If there exist matrices P and Q such that A is a p-consistently ordered matrix with respect to (P, Q), then*

$$\det(\lambda I - B) = \lambda^k \prod_{i=1}^{m} (\lambda^p - \mu_i^p), \tag{4.2}$$

where $\mu_i \neq 0$ for $i = 1, \ldots, m$, m and k are nonnegative integers.

Proof. Note that the equality (4.1) implies that

$$\det(\beta I - \alpha D^{-1} P - \alpha^{-(p-1)} D^{-1} Q) = \det(\beta I - B). \tag{4.3}$$

Let λ be a nonzero eigenvalue of B. We deduce to prove the result provided we can prove that $\lambda e^{2\pi i r/p}$ are eigenvalues of B for $r = 1, \ldots, p-1$, where $i = \sqrt{-1}$. With an application of the equality (4.3) a simple computation shows that

$$\begin{aligned}
\det(\lambda e^{2\pi i r/p} I - B) &= e^{2\pi i r n/p} \det(\lambda I - e^{-2\pi i r/p} D^{-1}(P + Q)) \\
&= e^{2\pi i r n/p} \det(\lambda I - e^{-2\pi i r/p} D^{-1} P - e^{2\pi i r(p-1)/p} D^{-1} Q) \\
&= e^{2\pi i r n/p} \det(\lambda I - B) = 0,
\end{aligned}$$

which shows the desired result. ∎

Applying Lemma 4.2 we now proceed to show the relation between the eigenvalues of $\mathcal{S}_\omega$ and the eigenvalues of the Jacobi matrix.

THEOREM 4.3. *Let $A = D - P - Q$ be a p-consistently ordered matrix with respect to (P, Q), where D is diagonal and nonsingular, and $P \in \mathcal{L}_n$. Assume that the parameters $\lambda \neq 0$ and μ satisfy the following relation*

$$(\lambda + \omega - 1)^p = \lambda^{p-1} \omega^p \mu^p, \tag{4.4}$$

where $\omega \neq 0$. If λ is an eigenvalue of $\mathcal{S}_\omega$ then μ is an eigenvalue of B. Conversely, if μ is an eigenvalue of B then λ is an eigenvalue of $\mathcal{S}_\omega$.

Proof. It is readily seen that $I-\omega D^{-1}P \in \mathcal{L}_n$. According to Theorem 2.5 we find $\det(I-\omega D^{-1}P)=1$, which implies that $\det(I-\omega D^{-1}P)^{-1}=1$. Therefore,

$$\begin{aligned}
\det(\lambda I-\mathcal{S}_\omega) &= \det(I-\omega D^{-1}P)^{-1}\det((\lambda+\omega-1)I-\lambda\omega D^{-1}P-\omega D^{-1}Q)\\
&= \det((\lambda+\omega-1)I-\lambda\omega D^{-1}P-\omega D^{-1}Q)\\
&= \det((\lambda+\omega-1)I-\lambda^{1-1/p}\omega(\lambda^{1/p}D^{-1}P+(\lambda^{1/p})^{-(p-1)}D^{-1}Q)\\
&= \lambda^{(1-1/p)n}\omega^n\det\Big(\frac{\lambda+\omega-1}{\lambda^{1-1/p}\omega}I-(\lambda^{1/p}D^{-1}P+(\lambda^{1/p})^{-(p-1)}D^{-1}Q)\Big)\\
&= \lambda^{(1-1/p)n}\omega^n\det(\frac{\lambda+\omega-1}{\lambda^{1-1/p}\omega}I-B).
\end{aligned}$$

Applying Lemma 4.2 we find

$$\det(\lambda I-\mathcal{S}_\omega)=\lambda^{(1-1/p)(n-k)}\omega^{n-k}(\lambda-\omega-1)^k\prod_{i=1}^{m}\Big(\frac{(\lambda-\omega-1)^p}{\lambda^{p-1}\omega^p}-\mu_i^p\Big). \tag{4.5}$$

If λ is an eigenvalue of $\mathcal{S}_\omega$ then $\det(\lambda I-\mathcal{S}_\omega)=0$, which implies that either (a) $\lambda-\omega-1=0$ with $k\geq 1$ or (b)

$$\frac{(\lambda-\omega-1)^p}{\lambda^{p-1}\omega^p}-\mu_i^p=0 \tag{4.6}$$

for some i. In the case (a) we find $\mu=0$ is the unique solution of (4.4). Since $k\geq 1$ Lemma 4.2 shows that $\mu=0$ is an eigenvalue of B. In case (b) it follows from Lemma 4.2 that any μ satisfying $\mu^p=\mu_i^p$ is an eigenvalue of B. Therefore, the equation (4.6) implies that any μ satisfying (4.4) is an eigenvalue of B. Conversely, assume that μ is an eigenvalue of B. If $\mu=0$ it follows from Lemma 4.2 that $k\geq 1$. The equation (4.4) shows $\lambda=1-\omega$. Applying (4.6) shows that $\lambda=1-\omega$ is an eigenvalue of $\mathcal{S}_\omega$. If $\mu\neq 0$ Lemma 4.2 shows that there is some i such that $\mu^p=\mu_i^p$. Again applying (4.6) shows that λ satisfying (4.4) is an eigenvalue of $\mathcal{S}_\omega$. ◻

Let ω_ρ is the unique positive real root of the equation

$$(\rho(B)\omega_\rho)^p=p^p(p-1)^{1-p}(\omega_\rho-1), \tag{4.7}$$

where $\rho(B)$ is the spectral radius of the associated Jacobi matrix. For $p=2$, ω_ρ can be expressed equivalently as

$$\omega_\rho=\frac{2}{1+\sqrt{1-\rho^2(B)}}=1+\left(\frac{\rho(B)}{1+\sqrt{1-\rho^2(B)}}\right)^2. \tag{4.8}$$

Following Varga's approach [16], we immediately obtain the following result on the optimum parameter.

THEOREM 4.4. *Let $A=D-P-Q$ be a p-consistently ordered matrix with respect to (P,Q), where D is a nonsingular matrix and $P\in\mathcal{L}_n$. If all the pth power of the Jacobi matrix B are nonnegative and $\rho(B)<1$ then with ω_ρ defined by (4.7)*

(a) $\rho(\mathcal{S}_{\omega_\rho})=(p-1)(\omega_\rho-1)$;

(b) $\rho(\mathcal{S}_\omega)\geq\rho(\mathcal{S}_{\omega_\rho})$ for all $\omega\neq\omega_\rho$.

Moreover, the iteration (3.1) converges for all ω with $0<\omega<p/(p-1)$.

For the case of $p=2$ we can also show that Young's fundamental result on the SOR method [20] and [21] is valid for the new method, which gives the spectral radius of $\mathcal{S}_\omega$ for $0<\omega<2$.

THEOREM 4.5. *Let $A = D - P - Q$ be a 2-consistently ordered matrices with respect to (P, Q), where D is a nonsingular matrix and $P \in \mathcal{L}_n$. If all the eigenvalues of the Jacobi matrix B are real then $\rho(\mathcal{S}_\omega) < 1$ if and only if $\rho(B) < 1$ and $0 < \omega < 2$. Moreover, if $\rho(B) < 1$ then with ω_ρ given by (4.8)*

(a) $\rho(\mathcal{S}_{\omega_\rho}) = \omega_\rho - 1$;

(b) if $\omega \neq \omega_\rho$ *then* $\rho(\mathcal{S}_\omega) > \rho(\mathcal{S}_{\omega_\rho})$;

(c) $\rho(\mathcal{S}_\omega) = \begin{cases} \left(\dfrac{\omega\rho(B) + (\omega^2\rho(B)^2 - 4(\omega - 1))^{1/2}}{2}\right)^2, & \text{if } \omega < \omega_\rho \\ w - 1, & \text{if } \omega_\rho \leq \omega < 2. \end{cases}$

4.2. Examples. In this subsection we present some examples to show that the new method not only is easily parallelized but also yields faster convergence than the SOR method for a wide class of matrices.

Example 1. Let $A = \text{tridiag}(a_{i,i-1}, a_{ii}, a_{i,i+1})$. Then A is a 2-consistently ordered matrix with repect to (L, U) [18] and [22]. Splitt $A = D - P - Q$, where

$$P = -\text{stair1}(a_{i,i-1}, 0, a_{i,i+1}), \quad Q = -\text{stair2}(a_{i,i-1}, 0, a_{i,i+1}).$$

Then $\det(\beta I - \alpha P - \alpha^{-1}Q)$ is independent of α for all $\alpha \neq 0$ because

$$\widetilde{D}^{-1}(\beta I - \alpha P - \alpha^{-1}Q)\widetilde{D} = (\beta I - P - Q),$$

where $\widetilde{D} = \text{diag}(\alpha_i I_i)$ and α_i is given by

$$\alpha_i = \begin{cases} 1, & \text{if } i \text{ is odd}, \\ \alpha, & \text{if } i \text{ is even}. \end{cases} \tag{4.9}$$

Therefore, for the linear system $A\mathbf{x} = \mathbf{b}$ according to Theorem 4.5 the SOR method

$$\mathbf{x}^{(k+1)} = (D - \omega L)^{-1}(((1 - \omega)D + \omega U)\mathbf{x}^{(k)} + \omega\mathbf{b}) \tag{4.10}$$

and the method

$$\mathbf{x}^{(k+1)} = (D - \omega P)^{-1}(((1 - \omega)D + \omega Q)\mathbf{x}^{(k)} + \omega\mathbf{b}) \tag{4.11}$$

share the same optimal asymptotic rate of convergence which is taken for ω_ρ given by (4.8), but the method (4.11) is much more easily implemented for parallel computation as explained in section 2.

In the second example, we show that the conditions on determination of the optimum parameter of the new method are still satisfied for a wide class of matrices for which the corresponding conditions required by the SOR method fail.

Example 2. Consider a $2p \times 2p$ matrix of the form

$$A = \text{tridiag}(a_{i,i-1}, a_{ii}, a_{i,i+1}) + \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1,2p} \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \\ a_{2p,1} & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{4.12}$$

As it is shown in [18] and [22] A is not a p-constantly ordered matrix with respect to (L, U). The theory on determination of optimum over-relaxation parameter of the

SOR method is not applicable to the matrices of the form (4.12). However, define a $2p \times 2p$ zebra matrix P by

$$(P)_{ij} = \begin{cases} -a_{ij}, & j = i-1, i+1,\ i = 2, 4, \ldots, 2p-2, \\ -a_{ij}, & j = 1, 2p-1,\ i = 2p, \\ 0, & \text{otherwise} \end{cases} \tag{4.13}$$

and split $A = D - P - Q$. Then $\det(\beta I - \alpha P - \alpha^{-1} Q)$ is independent of α for $\alpha \neq 0$ because

$$\widetilde{D}^{-1}(\beta I - \alpha P - \alpha^{-1} Q)\widetilde{D} = (\beta I - P - Q),$$

where $\widetilde{D} = \operatorname{diag}(I_1, \alpha I_2, I_3, \alpha I_4, \ldots, I_{2p-1}, \alpha I_{2p})$. Therefore, A is a 2-consistently ordered matrix with respect to (P, Q). Theorem 4.5 is still applicable for the iteration (4.11) if the eigenvalues of the Jacobi matrix are real.

For the particular case where $a_{2p,1} = 0$ and $a_{i,i+1} = 0$ for $i = 1, 2, \ldots, 2p-1$. It follows from [18] that A is a $2p$-consistently ordered matrix with respect to (L, U). However, A is still a 2-consistently ordered matrix with respect to (P, Q).

Finally, Example 3 shows that the advantages of high parallelism at each iteration and fast convergence rate of the new method are inherited when the method is applied to matrices arising for discretization of partial differential equations in a high dimensional space.

Example 3. Define a class of matrices by

- $T_1 = \{A : A = \operatorname{tridiag}(a_{i,i-1}, a_{ii}, a_{i,i+1})\}$,
- $T_k = \{A : A = \operatorname{tridiag}(A_{i,i-1}, A_{ii}, A_{i,i+1})$, where $A_{i,i-1}$, $A_{i,i+1}$ are diagonal matrices and $A_{ii} \in T_{k-1}\}$.

A number of matrices arising from discretization of partial differential equations belong to this class of matrices. For example, the difference matrices of self-adjusted elliptic equations in k-dimensions are in T_k. Similarly we denote

- $T_1^{\mathrm{s}} = \{A : A = \operatorname{stair}(a_{i,i-1}, a_{ii}, a_{i,i+1})\}$,
- $T_k^{\mathrm{s}} = \{A : A = \operatorname{stair}(A_{i,i-1}, A_{ii}, A_{i,i+1})$, where $A_{i,i-1}$, $A_{i,i+1}$ are diagonal matrices and $A_{ii} \in T_{k-1}^{\mathrm{s}}\}$.

Let $A = \operatorname{tridiag}(A_{i,i-1}, A_{ii}, A_{i,i+1}) \in T_k$ and $D = \operatorname{diag}(D_i)$ be the diagonal of A, where D_i is the diagonal of A_{ii}. Now we show that there exist a splitting $A = D - P - Q$, where $P, Q \in T_k^{\mathrm{s}}$, and a diagonal matrix $\widehat{D}$ such that

$$\widehat{D}^{-1}(\beta D - \alpha P - \alpha^{-1} Q)\widehat{D} = \beta D - P - Q \tag{4.14}$$

for any β and nonzero α. This is true for $k = 1$ as we have seen in Example 1. Assume that the equation (4.14) holds for any matrix in T_{k-1}. In particular, for each diagonal block A_{ii}, there are $P_i, Q_i \in T_{k-1}^{\mathrm{s}}$ and a diagonal matrix $\widehat{D}_i$ such that $A_i = D_i - P_i - Q_i$ and

$$\widehat{D}_i^{-1}(\beta D_i - \alpha P_i - \alpha^{-1} Q_i)\widehat{D}_i = \beta D_i - P_i - Q_i. \tag{4.15}$$

for any β and nonzero α. Let

$$P = \operatorname{stair1}(-A_{i,i-1}, P_i, -A_{i,i+1}), \qquad Q = \operatorname{stair2}(-A_{i,i-1}, Q_i, -A_{i,i+1}).$$

It is straightforward to show that $A = D - P - Q$. Define $\widehat{D} = \operatorname{diag}(\alpha_i \widehat{D}_i)$, where α_i is defined by (4.9). Applying (4.15) we show (4.14).

The equation (4.14) implies that A is a 2-consistently ordered matrix with respect to (P, Q). Therefore, the result on determination of optimum parameter is applicable to matrices in T_k. Another advantage of applying splitting $A = D - P - Q$ in the iteration (4.11) is the high parallelism of the scheme at each iteration.

5. Convergence for Hermitian positive definite matrices. Let A and D be Hermitian positive definite matrices satisfying $A = D - E - E^*$. Assume $D - \omega E$ is nonsingular for a given parameter ω. In this section we consider the asymptotic rate of convergence of the iterative scheme (3.1) for Hermitian positive definite matrices. We address our problem in a more general framework by choosing $P = E$ and $Q = E^*$. The iteration matrix $\mathcal{S}_\omega$ becomes

$$\mathcal{S}_\omega = (D - \omega E)^{-1}((1-\omega)D + \omega E^*) \tag{5.1}$$

5.1. Notation and preliminaries. Let A be an $n \times n$ complex matrix. We use $H(A) = \frac{1}{2}(A + A^*)$ and $S(A) = \frac{1}{2}(A - A^*)$ to represent the Hermitian and the anti-Hermitian parts of A, respectively.

For any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ we denote $(\mathbf{u}, \mathbf{v}) = \mathbf{u}^*\mathbf{v}$ the inner product of $\mathbf{u}$ and $\mathbf{v}$. It is well-known that

$$(\mathbf{u}, \mathbf{v}) = \overline{(\mathbf{v}, \mathbf{u})}, \qquad (\mathbf{u}, A\mathbf{v}) = (A^*\mathbf{u}, \mathbf{v}). \tag{5.2}$$

For an $n \times n$ complex matrix A the Rayleigh quotient of A for a nonzero vector $\mathbf{x} \in \mathbb{C}^n$ is $q(A, \mathbf{x}) = (\mathbf{x}, A\mathbf{x})/(\mathbf{x}, \mathbf{x})$ and the numerical radius of A is defined by

$$r(A) = \sup\{|(\mathbf{x}, A\mathbf{x})| : \ \mathbf{x} \in \mathbb{C}^n, \ (\mathbf{x}, \mathbf{x}) = 1\}. \tag{5.3}$$

$V(A) = \{(\mathbf{x}, A\mathbf{x}) : \ \mathbf{x} \in \mathbb{C}^n, \ (\mathbf{x}, \mathbf{x}) = 1\}$ is called the field of values or the numerical range of A. Numerical radius is considered to be an efficient norm to measure convergence of basic iterative methods. See [2], [4], [11] and [14] for some examples.

For a nonnegative matrix A, in 1975, Goldberg, Tadmor and Zwas [6] showed that the numerical radius of A is equal to the spectral radius of its symmetric part.

LEMMA 5.1. *If A is an $n \times n$ nonnegative matrix, then $r(A) = \rho(H(A))$.*

In general, we have the following bounds for numerical radius [2], [11].

LEMMA 5.2. *Let A be an $n \times n$ matrix. Then*

$$\max(\rho(H(A)), \rho(S(A))) \le r(A) \le \sqrt{\rho^2(H(A)) + \rho^2(S(A))} \tag{5.4}$$

LEMMA 5.3. *Let A, R and S be $n \times n$ matrices. If $R^*R = S^*S$ and R is nonsingular then $r(RAR^*) = r(SAS^*)$.*

Proof. Under the conditions of the lemma S is nonsingular. For any $\mathbf{x} \neq 0$ an elementary calculation shows

$$\frac{(\mathbf{x}, RAR^*\mathbf{x})}{(\mathbf{x}, \mathbf{x})} = \frac{(\mathbf{y}, A\mathbf{y})}{((R^*)^{-1}\mathbf{y}, (R^*)^{-1}\mathbf{y})} = \frac{(\mathbf{y}, A\mathbf{y})}{((S^*)^{-1}\mathbf{y}, (S^*)^{-1}\mathbf{y})} = \frac{(\mathbf{z}, SAS^*\mathbf{z})}{(\mathbf{z}, \mathbf{z})},$$

which implies the desired result, where $\mathbf{y} = R^*\mathbf{x}$ and $\mathbf{z} = (S^*)^{-1}\mathbf{y}$. ◻

Other properties of numerical radius can be found in [5], [8], [9] and [13].

A Stieltjes matrix is a symmetric M-matrix. For a Stieltjes matrix we have the following property on its Cholesky factorization.

LEMMA 5.4. *Let A be a Stieltjes matrix and $A = \widetilde{L}\widetilde{L}^T$ be the Cholesky factorization of A. Then $\widetilde{L}$ is an M-matrix.*

Proof. The proof is straightforward. ◻

5.2. Main results. In this subsection, we derive a bound for the spectral radius of the iteration matrix $\mathcal{S}_\omega$ for a Hermitian positive definite matrix with $0 < \omega < 2$.

Let $A = D - E - E^*$ with a Hermitian positive definite matrix D and assume that $D - \omega E$ is nonsingular for $0 \leq \omega \leq 2$. In 1954, Ostrowski showed that $\rho(\mathcal{S}_\omega) < 1$ if and only if A is positive definite and $0 < \omega < 2$ [12]. In 1973, Varga [19] pointed out that $D - \omega E$ is nonsingular for $0 \leq \omega \leq 2$ if A and D are positive definite. Combining Varga's observation with Ostrowski's Theorem [12] (see also [18]) we immediately obtain the following result.

THEOREM 5.5. *Let $A = D - E - E^*$ be an $n \times n$ Hermitian matrix, where D is Hermitian positive definite. Then $D - \omega E$ is nonsingular and $\rho(\mathcal{S}_\omega) < 1$ if and only if A is positive definite and $0 < \omega < 2$.*

For a Hermitian positive definite matrix $A = D - E - E^*$ with a Hermitian positive definite matrix D the scheme (3.1) converges, i.e., $\rho(\mathcal{S}_\omega) < 1$, if and only if $0 < \omega < 2$ according to Theorem 5.5. The following lemma shows an upper bound of the eigenvalues of the Jacobi matrix.

LEMMA 5.6. *Let A and D be $n \times n$ Hermitian positive definite matrices and $A = D - E - E^*$. Then all eigenvalues of the Jacobi matrix $D^{-1}(E + E^*)$ are real and strictly less than 1.*

Proof. $\lambda(D^{-1}(E + E^*)) = \lambda(D^{-1/2}(E + E^*)D^{-1/2})$ shows that all eigenvalues of $D^{-1}(E + E^*)$ are real. Since A is Hermitian positive definite, for any $\mathbf{x} \neq 0$ we find

$$(\mathbf{x}, D^{-1/2}(E + E^*)D^{-1/2}\mathbf{x}) = 1 - (\mathbf{x}, D^{-1/2}AD^{-1/2}\mathbf{x}) < 1,$$

which implies $\lambda(D^{-1}(E + E^*)) < 1$. $\square$

The main result on estimate of the spectral radius $\rho(\mathcal{S}_\omega)$ is presented in the following Theorem. A similar approach can be found in Varga's work in 1973 [19] where he split the matrix E into the Hermitian and the anti-Hermitian parts. Here we treat E globally and use numerical radius.

THEOREM 5.7. *Let A and D be $n \times n$ Hermitian positive definite matrices and $A = D - E - E^*$. Assume that the eigenvalues of the Jacobi matrix $D^{-1}(E + E^*)$ lie on $[\widetilde{\alpha}, \widetilde{\beta}]$ with $\widetilde{\beta} < 1$ and $r \geq r(D^{-1/2}ED^{-1/2})$. Then for $0 < \omega < 2$*

$$\rho(\mathcal{S}_\omega) \leq \begin{cases} \sqrt{1 - \dfrac{\omega(2-\omega)(1-\beta)}{1 - \beta\omega + r^2\omega^2}}, & \text{if } r^2\omega^2 \geq \omega - 1, \\ \sqrt{1 - \dfrac{\omega(2-\omega)(1-\alpha)}{1 - \alpha\omega + r^2\omega^2}}, & \text{if } r^2\omega^2 < \omega - 1, \end{cases} \tag{5.5}$$

where $\alpha = \max(\widetilde{\alpha}, -2r)$ and $\beta = \min(\widetilde{\beta}, 2r)$. Furthermore,

$$\min_{0<\omega<2} \rho(\mathcal{S}_\omega) \leq \begin{cases} \dfrac{\sqrt{4r^2 - \beta^2}}{\sqrt{1 - 2(\beta - 2r^2)} + 1 - \beta}, & \text{if } 4r^2 > \beta, \\ \dfrac{2r}{1 + \sqrt{1 - 4r^2}}, & \text{if } \alpha \leq 4r^2 \leq \beta, \\ \dfrac{\sqrt{4r^2 - \alpha^2}}{\sqrt{1 - 2(\alpha - 2r^2)} + 1 - \alpha}, & \text{if } 4r^2 < \alpha. \end{cases} \tag{5.6}$$

Proof. It follows from Lemma 5.2 that $\alpha \leq \lambda(D^{-1}(E + E^*)) \leq \beta$. Denote $\widetilde{E} = D^{-1/2}ED^{-1/2}$. We find $\mathcal{S}_\omega = D^{-1/2}\widetilde{\mathcal{S}}_\omega D^{1/2}$, where

$$\widetilde{\mathcal{S}}_\omega = (I - \omega\widetilde{E})^{-1}((1-\omega)I + \omega\widetilde{E}^*). \tag{5.7}$$

Therefore, $\mathcal{S}_\omega$ and $\widetilde{\mathcal{S}}_\omega$ have same eigenvalues. Let λ be an eigenvalue of $\widetilde{\mathcal{S}}_\omega$ and $\mathbf{x} \neq 0$ be the corresponding eigenvector. Then $\widetilde{\mathcal{S}}_\omega \mathbf{x} = \lambda \mathbf{x}$, i.e.,

$$\lambda(I - \omega \widetilde{E})\mathbf{x} = ((1-\omega)I + \omega \widetilde{E}^*)\mathbf{x}. \tag{5.8}$$

Multiplying both sides on left by $\mathbf{x}^*$ we obtain

$$\lambda((\mathbf{x},\mathbf{x}) - \omega(\mathbf{x}, \widetilde{E}\mathbf{x})) = (1-\omega)(\mathbf{x},\mathbf{x}) + \omega(\mathbf{x}, \widetilde{E}^*\mathbf{x}). \tag{5.9}$$

First we prove that $(\mathbf{x},\mathbf{x}) - \omega(\mathbf{x}, \widetilde{E}\mathbf{x}) \neq 0$ by reduction to absurdity. Assume that $\omega(\mathbf{x}, \widetilde{E}\mathbf{x}) = (\mathbf{x},\mathbf{x})$. Applying (5.2) shows that

$$\omega(\mathbf{x}, \widetilde{E}^*\mathbf{x}) = \omega(\widetilde{E}\mathbf{x}, \mathbf{x}) = \omega\overline{(\mathbf{x}, \widetilde{E}\mathbf{x})} = (\mathbf{x},\mathbf{x}). \tag{5.10}$$

Furthermore, it follows from (5.9) that $(\omega(\mathbf{x}, E^*\mathbf{x}) + (1-\omega)(\mathbf{x},\mathbf{x})) = (2-\omega)(\mathbf{x},\mathbf{x}) = 0$. Therefore, $\omega = 2$ due to $\mathbf{x} \neq 0$, which contradicts to the assumption $0 < \omega < 2$.

Let $(\mathbf{x}, \widetilde{E}\mathbf{x})/(\mathbf{x},\mathbf{x}) = \kappa e^{i\theta}$, where $\kappa \geq 0$, $0 \leq \theta < 2\pi$ and $i = \sqrt{-1}$. Using (5.9) with a simple computation shows that $\lambda = (1 - \omega + \omega\kappa e^{-\theta i})/(1 - \omega\kappa e^{\theta i})$ and

$$|\lambda|^2 = 1 - \omega(2-\omega)(1 - 2\kappa\cos\theta)/(1 - 2\omega\kappa\cos\theta + \omega^2\kappa^2). \tag{5.11}$$

Since $\lambda(\widetilde{E} + \widetilde{E}^*) = \lambda(D^{-1}(E + E^*))$ and $2\kappa\cos\theta = (\mathbf{x}, (\widetilde{E} + \widetilde{E}^*)\mathbf{x})/(\mathbf{x},\mathbf{x})$, we find $\alpha \leq 2\kappa\cos\theta < \beta$ according to Lemma 5.2. Due to $0 < \omega < 2$ and $\kappa \leq r$ applying (5.11) yields

$$|\lambda|^2 \leq f(2\kappa\cos\theta) < 1, \tag{5.12}$$

where $f(x)$ is a function $f(x)$ defined on $[\alpha, \beta]$ by

$$f(x) = 1 - \omega(2-\omega)(1-x)/(1 - \omega x + \omega^2 r^2). \tag{5.13}$$

The rest of the proof is to compute the maximum value of $f(x)$. By representing

$$f(x) = w - 1 + (2-\omega)\frac{r^2\omega^2 - \omega + 1}{1 - \omega x + \omega^2 r^2},$$

it is clearly seen that if $r^2\omega^2 \geq \omega - 1$ then $f(x)$ is increase for x and if $r^2\omega^2 < \omega - 1$ then $f(x)$ is decrease for x. Hence,

$$\max_{\alpha \leq x \leq \beta} f(x) = \begin{cases} 1 - \dfrac{\omega(2-\omega)(1-\beta)}{1 - \omega\beta + r^2\omega^2}, & \text{if } r^2\omega^2 \geq \omega - 1, \\ 1 - \dfrac{\omega(2-\omega)(1-\alpha)}{1 - \omega\alpha + r^2\omega^2}, & \text{if } r^2\omega^2 < \omega - 1, \end{cases} \tag{5.14}$$

which with (5.12) implies (5.5).

If $4r^2 < 1$ we find that $r^2\omega^2 = \omega - 1$ has two real zeros and one of them lies on $(0, 2)$, which is given by

$$\omega_r = \frac{2}{1 + \sqrt{1 - 4r^2}}. \tag{5.15}$$

We now proceed to show (5.6). For parameter ω satisfying $r^2\omega^2 \geq \omega - 1$, denote $g(\omega) = \omega(2-\omega)/(1 - \beta\omega + r^2\omega^2)$. A simple calculation shows that

$$g'(\omega) = \frac{(\beta - 2r^2)\omega^2 - 2\omega + 2}{(1 - \beta\omega + r^2\omega^2)^2}. \tag{5.16}$$

We find that $g'(\omega)$ has one zero $\omega = 1$ if $\beta - 2r^2 = 0$ and two zeros

$$\omega_1 = \frac{2}{1+\sqrt{1-2(\beta-2r^2)}}, \qquad \omega_2 = \frac{1+\sqrt{1-2(\beta-2r^2)}}{\beta-2r^2}$$

if $\beta - 2r^2 \neq 0$. On the other hand, under the conditions of the theorem $1-2(\beta-2r^2) \geq 1-2\beta+\beta^2 = (1-\beta)^2$, which implies that ω_1 and ω_2 are real. It is straightforward to check that $\omega_2 \geq 2$ if $\beta - 2r^2 > 0$ and $\omega_2 < 0$ if $\beta - 2r^2 < 0$. Therefore, $g'(\omega)$ has a unique zero

$$\omega_{\beta,r} = \frac{2}{1+\sqrt{1-2(\beta-2r^2)}} \tag{5.17}$$

on $(0, 2)$. Furthermore, $g'(\omega) \geq 0$ for $\omega \in (0, \omega_{\beta,r})$ and $g'(\omega) \leq 0$ for $\omega \in (\omega_{\beta,r}, 2)$ regardless of the sign of $\beta - 2r^2$.

If $4r^2 > \beta$, we find $r^2\omega_{\beta,r}^2 > \omega_{\beta,r} - 1$. It turns out that for $r^2\omega^2 \geq \omega - 1$ and $\omega \in (0, 2)$ we have the maximum value

$$\max g(\omega) = g(\omega_{\beta,r}) = \frac{2}{\sqrt{1-2(\beta-2r^2)}+1-\beta}. \tag{5.18}$$

Note that even if $4r^2 < 1$, which implies that ω_r given by (5.15) lies on $(0, 2)$, the inequality $g(\omega_{\beta,r}) \geq g(\omega_r)$ holds.

If $4r^2 \leq \beta$ then $4r^2 < 1$ due to $\beta < 1$. In this case, it is readily seen that $r^2\omega_{\beta,r}^2 \leq \omega_{\beta,r} - 1$ and $\omega_{\beta,r} \geq \omega_r$. Therefore, for $r^2\omega^2 \geq \omega - 1$ and $\omega \in (0, 2)$

$$\max g(\omega) = g(\omega_r) = \frac{2\sqrt{1-4r^2}}{(1-\beta)(1+\sqrt{1-4r^2})} \tag{5.19}$$

If there are some real ω such that $r^2\omega^2 < \omega - 1$, it is straightforward to show that $4r^2 < 1$. Denote $q(\omega) = \omega(2-\omega)/(1-\alpha\omega+r^2\omega^2)$. Similarly, we can prove

$$\sup q(\omega) = q(\omega_r) = \frac{2\sqrt{1-4r^2}}{(1-\alpha)(1+\sqrt{1-4r^2})} \tag{5.20}$$

if $4r^2 \geq \alpha$ and

$$q(\omega_r) \leq \sup q(\omega) = q(\omega_{\alpha,r}) = \frac{2}{\sqrt{1-2(\alpha-2r^2)}+1-\alpha} \tag{5.21}$$

if $4r^2 < \alpha$. Applying (5.5) we find that

$$\min_{0<\omega<2}(\rho(\mathcal{S}_\omega)) \leq \sqrt{1-\max_{0<\omega<2}\Big(\max_{r^2\omega^2\geq\omega-1}(1-\beta)g(\omega), \sup_{r^2\omega^2<\omega-1}(1-\alpha)q(\omega)\Big)},$$

which together with (5.18), (5.19), (5.20) and (5.21) shows (5.6). ∎

If $\widetilde{\alpha}$ and $\widetilde{\beta}$ are the minimum and the maximum eigenvalues of $D^{-1}(E+E^*)$, respectively, it follows from Lemma 5.2 that $r \geq \frac{1}{2}\max(|\widetilde{\alpha}|, |\widetilde{\beta}|)$. Applying Theorem 5.7 yields the following result immediately.

Corollary 5.8. *Let A, D and E satisfy the conditions of Theorem 5.7. Assume α and β are the minimum and the maximum eigenvalues of $D^{-1}(E+E^*)$, respectively, and $r \geq r(D^{-1/2}ED^{-1/2})$. Then for $0 < \omega < 2$ the inequalities (5.5) and (5.6) hold.*

For the method presented in this paper and the SOR method because the diagonal of $E+E^*$ is zero, the matrix $\widetilde{E}+\widetilde{E}^*$ is neither positive definite nor negative definite. Thus, $\alpha \le 0$ and the case $4r^2 < \alpha$ never occurs. In general, however, one can easily find Hermitian positive definite matrices A and D satisfying the conditions of Theorem 5.7 such that $4r^2(D^{-1/2}ED^{-1/2}) < \min \lambda(D^{-1}(E+E^*))$. Applying Theorem 5.7 to Stieltjes matrices we obtain the following result, which generalizes some elegant results on the SOR method for Stieltjes matrices in [16], [18] and [22].

COROLLARY 5.9. *Let A and D be Stieltjes matrices and $A = D - E - E^T$, where E is a nonnegative matrix. Then*

$$\rho(\mathcal{S}_\omega) \le \begin{cases} \dfrac{\rho\omega + 2 - 2\omega}{2 - \rho\omega}, & \text{if } \rho^2\omega^2 \ge 4(\omega-1), \\ \dfrac{\rho\omega + 2\omega - 2}{2 + \rho\omega}, & \text{if } \rho^2\omega^2 < 4(\omega-1), \end{cases} \tag{5.22}$$

$$\min_{0<\omega<2} \rho(\mathcal{S}_\omega) \le \rho(\mathcal{S}_{\omega_\rho}) \le \sqrt{\omega_\rho - 1}, \tag{5.23}$$

where $\omega_\rho = 2/(1+\sqrt{1-\rho^2})$ and $\rho = \rho(D^{-1}(E+E^T))$.

Proof. Because $D^{-1}(E+E^T)$ is a nonnegative matrix we find that ρ is an eigenvalue of $D^{-1}(E+E^T)$ according to Theorem 2.1.1 in [3]. Furthermore, Lemma 5.6 shows $\rho < 1$. Let $D = \widetilde{L}\widetilde{L}^T$ be the Cholesky factorization of D. Then $\widetilde{L}$ is an M-matrix according to Lemma 5.4. Let $r = r(D^{-1/2}ED^{-1/2})$. It follows from Lemma 5.3 that $r = r(\widetilde{L}^{-1}E(\widetilde{L}^T)^{-1})$ due to $(\widetilde{L}^{-1})^T\widetilde{L}^{-1} = D^{-1/2}D^{-1/2}$. Since $\widetilde{L}^{-1}E(\widetilde{L}^T)^{-1}$ is a nonnegative matrix, Lemma 5.1 shows that

$$r = \rho\left(\frac{\widetilde{L}^{-1}E(\widetilde{L}^T)^{-1} + \widetilde{L}^{-1}E^T(\widetilde{L}^T)^{-1}}{2}\right) = \frac{1}{2}\rho((\widetilde{L}^T)^{-1}\widetilde{L}^{-1}(E+E^T)) = \frac{\rho}{2}.$$

Choosing $\widetilde{\alpha} = -\rho$, $\widetilde{\beta} = \rho$ and applying (5.5) of Theorem 5.7 show (5.22). Since $r = \rho/2$ and $\rho < 1$, we find $-\rho \le 4r^2 \le \rho$. Then (5.23) follows from (5.6) of Theorem 5.7. ◻

Note that for a Stieltjes matrix D there is a Stieltjes matrix G such that $G^2 = D$ according to Theorem 6.15 in [1]. We can use G instead of $D^{1/2}$ in the proof of Corollary 5.9. Then $G^{-1}EG^{-1}$ is a nonnegative matrix and $r = r(G^{-1}EG^{-1}) = \rho/2$. This provides another way to prove the corollary.

If we know there exists a real eigenvalue λ of $\mathcal{S}_\omega$ such that $|\lambda| = \rho(\mathcal{S}_\omega)$, the estimate of $\rho(\mathcal{S}_\omega)$ becomes rather simple. Because λ is real it follows from (5.9) that

$$\lambda((\mathbf{x},\mathbf{x}) - \omega(\mathbf{x},\widetilde{E}^*\mathbf{x})) = (1-\omega)(\mathbf{x},\mathbf{x}) + \omega(\mathbf{x},\widetilde{E}\mathbf{x}). \tag{5.24}$$

Adding (5.9) and (5.24) shows that

$$\lambda = \frac{2(1-\omega)+\omega y}{2-\omega y}, \tag{5.25}$$

where $y = (\mathbf{x},(E+E^*)\mathbf{x})/(\mathbf{x},\mathbf{x}) \in [\widetilde{\alpha},\widetilde{\beta}]$ with $\widetilde{\beta} < 1$. Since the right hand side of the equation (5.25) is a increasing function of y, we have

$$\rho(\mathcal{S}_\omega) = |\lambda| \le \max\Big(\frac{2(1-\omega)+\omega\widetilde{\beta}}{2-\omega\widetilde{\beta}}, \frac{2(\omega-1)-\omega\widetilde{\alpha}}{2-\omega\widetilde{\alpha}}\Big). \tag{5.26}$$

6. Numerical examples. In this section we present some numerical results to corroborate our observation. The problem considered is the Poisson equation

$$\begin{cases} \Delta u = 1, \quad (x,y) \in \Omega = (0,1) \times (0,1), \\ u|_{\partial\Omega} = 0 \end{cases} \tag{6.1}$$

on two dimensions. The equation is discretized by a central difference scheme with a uniform mesh size h. The mesh points are numbered in the lexicographic order. This yields the following linear system

$$A\mathbf{x} = \mathbf{b}, \tag{6.2}$$

where A is a block tridiagonal matrix of the form

$$A = \text{tridiag}(A_{i,i-1}, A_{ii}, A_{i,i+1})$$

with $A_{i,i-1} = A_{i,i+1} = -I$ and $A_{ii} = \text{tridiag}(-1, 4, -1)$. The number of the unknowns of the system is $N = (h^{-1} - 1)^2$.

The linear system (6.2) is solved by the SOR method and the new method. It follows from [18] and [22] that A is a 2-consistently ordered matrix with respect to (L, U). Because of $A \in T_2$, for the new method the matrix splitting $A = D - P - Q$ is obtained in the way described in Example 3. Therefore, A is also a 2-consistently ordered matrix with respect to (P, Q). The stopping criterion is

$$\|\mathbf{r}_i\|_2 / \|\mathbf{r}_0\|_2 < 10^{-5}, \tag{6.3}$$

where $\mathbf{r}_i = \mathbf{b} - A\mathbf{x}^{(i)}$ is the ith residual and the initial guess is $\mathbf{x}^{(0)} = (1, 1, \ldots, 1)^T$. It is well known that the spectral radius of the Jacobi matrix of A is given by $\rho(B) = \cos(h\pi)$. Hence, the optimum parameter is given by $\omega_{\text{opt}}(h) = 2/(1 + \sin(h\pi))$ for a given h. Table 6.1 shows the iteration numbers of the SOR method and the new method for different parameter ω and different mesh size. To save the space O stands for the SOR method and N stands for the new method in the table.

TABLE 6.1
The iteration numbers of SOR and the new method

	$\omega_{\text{opt}}(8)$		$\omega_{\text{opt}}(16)$		$\omega_{\text{opt}}(32)$		$\omega_{\text{opt}}(64)$		$\omega_{\text{opt}}(128)$		$\omega_{\text{opt}}(256)$	
h^{-1}	O	N	O	N	O	N	O	N	O	N	O	N
8	19	17	33	31	64	61	127	124	251	245	496	484
16	80	84	36	34	64	62	128	124	254	248	499	489
32	294	320	154	174	69	69	128	124	256	246	507	487
64	1034	1149	554	653	291	358	132	129	256	248	512	490
128	3553	4022	1908	2323	1016	1325	495	679	259	258	512	492
256	11857	13742	6371	8049	3397	4676	1675	2506	841	1315	515	515

As we see for each optimum parameter the iterations of the new method and the SOR method are approximately the same though there are some differences if $\omega \neq \omega_{\text{opt}}$.

7. Conclusions. Using the matrices introduced in the present paper we can also construct some other new iterative schemes based on other basic iterative methods. However, these issues will not be addressed in this series. In part 2 of the series stair matrices and their generalizations will be applied to construction of preconditioners and preconditioned conjugate gradient methods.

REFERENCES

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, New York, 1994.

[2] O. Axelssoni, H. Lu, and B. Polman, *On the numerical radius of matrices and its application to iteration methods*, Linear and multilinear Algebra, 37 (1994), pp. 225–238.

[3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in Mathematical Sciences*, Academic Press, New York, 1979.

[4] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.

[5] M. Goldberg and E. Tadmor, *On numerical radius and its applications*, Linear Algebra Appl., 42 (1982), pp. 263–284.

[6] M. Goldberg, E. Tadmor, and G. Zwas, *Numerical radius of positive matrices*, Linear Algebra Appl., 12 (1975), pp. 209–214.

[7] W. Hackbusch, *Iterative Solution of Large Sparse Linear Systems of Equations*, Applied Math. Sci. Series, No. 95, Springer Verlag, 1994.

[8] R. A. Horn and C. R. Johnson, *Matrix Anlysis*, Cambridge University Press, New York, 1985.

[9] C. R. Johnson, *Numerical determination of the field of values of a general complex matrix*, SIAM J. Numer. Anal., 15 (1978), pp. 595–602.

[10] W. Kahan, *Gauss-Seidel methods of solving large systems of linear equations*, PhD thesis, University of Toronto, Canada, 1958.

[11] H. Lu, *Forward-backward heat equations and analysis of iterative methods*, PhD thesis, Department of Mathematics, University of Nijmegen, The Netherlands, 1995.

[12] A. M. Ostrowski, *On the linear iteration procedures for symmetric matrices*, Rend. Mat. e Appl., 14 (1954), pp. 140–163.

[13] C. Pearcy, *An elementary proof of the power inequality for the numerical radius*, Michigan Math. J., 13 (1966), pp. 289–291.

[14] G. Starke, *Fields of values and the ADI method for nonnormal matrices*, Linear Algebra Appl., 180 (1993), pp. 199–218.

[15] P. Stein and R. L. Rosenberg, *On the solution of linear simultaneous equations by iteration*, J. London Math. Soc., 23 (1948), pp. 111–118.

[16] R. S. Varga, *Orderings of the successive overlaxation scheme*, Pacific J. Math, 9 (1959), pp. 925–939.

[17] ———, *p-cyclic matrices: a generalization of the Young-Frankel successive overrelation scheme*, Pacific J. Math, 9 (1959), pp. 617–628.

[18] ———, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[19] ———, *Extensions of the successive overrelation theory with applications of finite element approximations*, in Topics in Numerical Analysis, J. J. H. Miller, ed., Academic Predd, New York, 1973, pp. 329–343.

[20] D. M. Young, *Iterative methods for solving partial difference equations of elliptic type*, PhD thesis, Harvard University, Cambridge, Massachusetts, 1950.

[21] ———, *Iterative methods for solving partial difference equations of elliptic type*, Trans. Amer. Math. Soc, 76 (1954), pp. 92–111.

[22] ———, *Iterative Solution for Large Systems*, Academic Press, New York, 1971.